Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

J.L. van den Berg, O.J. Boxma

Sojourn times in feedback and processor sharing queues

Department of Operations Research and System Theory Report OS-R8801 January

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

Sojourn Times in Feedback and Processor Sharing Queues

J.L. van den Berg, O.J. Boxma
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

This paper considers an M/M/1 queue with a general probabilistic feedback mechanism. When a customer completes his i-th service, he departs from the system with probability $1-p(i)$ and he cycles back with probability $p(i)$. The mean service time of the customer is the same for each cycle. If this mean service time shrinks to zero, while the feedback probabilities approach one in such a way that the mean total required service time remains a positive constant, then the behaviour of the feedback queue approaches that of an M/G/1 processor sharing queue. Different choices of the feedback probabilities lead to different service time distributions in the processor sharing model.
In a recent paper we have derived the joint distribution of the successive sojourn times of a customer in the feedback queue. We now exploit those results to analyse sojourn times in the M/G/1 queue with processor sharing. In particular, new results will be presented for the sojourn time variance of a customer with given total service request.

1980 Mathematics Subject Classification: 60K25, 68M20.
Key Words & Phrases: M/M/1 feedback queue, M/G/1 processor sharing queue, sojourn times.
Note: This report will be submitted for publication elsewhere.

1. Introduction

This paper considers an M/M/1 queue with a general probabilistic feedback mechanism. When a newly arriving customer, to be called a type-1 customer, has received his service, he departs from the system with probability $1-p(1)$ and is fed back to the end of the queue with probability $p(1)$; in the latter case he becomes a type-2 customer. When he has received his i-th service, he leaves with probability $1-p(i)$ and he cycles back with probability $p(i)$, in the latter case becoming a type-$i+1$ customer. The service times of each customer at all visits are identically, negative exponentially, distributed stochastic variables. The resulting queueing model has the property that the finite-dimensional joint queue-length distribution of type-i customers, $i=1,2,...$, is of product-form type. This property has been exploited in [2,3] to derive the Laplace-Stieltjes Transform (LST) of the joint stationary distribution of the successive sojourn times of a customer.

Feedback queues are useful for modelling many phenomena in computer-communication systems. An important example is timeshare scheduling in computer systems. Kleinrock [6, Vol. 2] contains several early references; Nelson [7] considers queues with a feedback mechanism as described above, but with varying service times, to study the effect that assigning increasing numbers of time quanta to jobs has on mean sojourn times. Other examples of feedback systems are found in telecommunications. For instance, a telephone call may generate several tasks for processing. Such tasks can sometimes be considered as feedbacks. From a customer's viewpoint, a very important performance measure is the response time or sojourn time, here defined as the time spent by a particular sequence of tasks from its arrival to its service completion [4]; and not just its mean is of interest, but also (the tail of) its distribution.

If the mean service time in each loop shrinks to zero, while the feedback probabilities approach one in such a way that the mean total required service time remains a positive constant, then the behaviour of the feedback queue approaches that of an M/G/1 processor sharing queue (M/G/1 PS). Different choices of the feedback probabilities lead to different service time distributions in the processor sharing model.

The queue length process in a round-robin type of queue is usually less amenable to mathematical

Report OS-R8801
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

analysis than the queue length process in its limiting case, a PS queue. This has been the main reason for the queueing analysis of processor sharing. The determination of the *sojourn time* distribution in a PS queue has turned out to be a much harder problem. Only recently the sojourn time distribution in the M/G/1 PS queue has been derived, cf. [5, 8, 9, 10] and the survey of Yashkov [11]. In our case the situation is reversed, as far as complexity is concerned. The feedback model under consideration can be completely analysed [3], and we exploit this to present a new approach to the sojourn time analysis in the M/G/1 PS queue.

The organization of the rest of the paper is as follows. Section 2 presents a detailed model description and a summary of those results of [3] that are essential for the analysis in Section 3. In the latter section we study sojourn times in the M/G/1 PS queue, by taking appropriate limits in the M/M/1 queue with feedback. We restrict ourselves in Section 3 mainly to the sojourn time variance. In a future paper [1], a more extensive analysis of the sojourn time distribution in the M/G/1 PS queue will be presented.

2. Model description and preliminary results

We consider a single server queueing system with infinite waiting room. Customers arrive at the system according to a Poisson process with intensity $\lambda>0$. After having received a service, a customer may either leave the system or be fed back. When a customer has completed his i-th service, he departs from the system with probability $1 - p(i)$ and is fed back with probability $p(i)$. Fed back customers return instantaneously, joining the end of the queue. A customer who is visiting the queue for the i-th time will be called a type-i customer.

The successive service times of a customer are independent, negative exponentially distributed, random variables, with mean β. These service times are also independent of the service times of other customers.

Introduce

$$q(1) := 1, \tag{2.1}$$
$$q(i) := \prod_{j=1}^{i-1} p(j), \quad i=2,3,\ldots .$$

Note that $q(i)$ is the relative arrival rate of type-i customers, $i=1,2,\ldots$. The offered load to the queue per unit of time is $\rho:=\lambda\beta\sum_{i=1}^{\infty}q(i)$. For stability it is required that $\rho<1$. For future reference we introduce the generating function of the probabilities of visiting the queue exactly i times, $i=1,2,\ldots$:

$$Q(z) := \sum_{i=1}^{\infty} q(i)(1-p(i))z^i, \quad |z|\leqslant 1. \tag{2.2}$$

We now summarize those results from [3] which will be needed in Section 3. Let S_j be the time required for the j-th pass through the system (j-th sojourn time), $j=1,2,\ldots$.

Theorem 2.1 *The LST of the joint distribution of the first k successive sojourn times of a customer who is fed back at least $k-1$ times, is given by*

$$E\{e^{-(\omega_1 S_1+\ldots+\omega_k S_k)}\} = \frac{1-\rho}{(1+\beta\omega_1)M_k(k-1,\omega) - \lambda\beta\sum_{j=1}^{k-2}q(k-j-1)M_k(j,\omega) - (\rho-\lambda\beta\sum_{i=1}^{k-2}q(i))}, \tag{2.3}$$

with $\mathrm{Re}\ \omega_i\geqslant 0$, $i=1,\ldots,k$, $\omega := (\omega_1,\ldots,\omega_k)$, *and*

$$M_k(j,\omega) = (1+\beta\omega_{k-j+1})M_k(j-1,\omega) + \lambda\beta\Big[M_k(j-1,\omega) - q(j-1) - \sum_{l=2}^{j-1} q(j-l)(1-p(j-l))\,M_k(l-1,\omega)\Big], \quad j=1,...,k-1, \tag{2.4}$$

$M_k(0,\omega) \equiv 1, \;\; q(0) := 1.$

This result can be derived from (3.13), (3.14) and (3.15) of [3]. The present form is somewhat simpler, and more suitable for analysing sojourn time distributions in the M/G/1 PS queue. Some of the implications of the theorem are:
(i) the sojourn times $\mathbf{S}_j$, $j=1,...,k$, are negative exponentially distributed with mean $\beta/(1-\rho)$.
(ii) The LST of the joint distribution of $\mathbf{S}_i$ and $\mathbf{S}_j$, $1 \leqslant i < j \leqslant k$, is given by

$$E\{e^{-(\omega_i \mathbf{S}_i + \omega_j \mathbf{S}_j)}\} = \frac{1-\rho}{1-\rho+\beta\omega_i+\beta\omega_j+\beta^2\omega_i\omega_j C_{j-i}}, \quad 1 \leqslant i < j \leqslant k, \tag{2.5}$$

where C_{j-i} is determined by

$$C_1 = 1, \tag{2.6}$$
$$C_n = 1+\lambda\beta\sum_{j=1}^{n-1} q(n-j)C_j, \quad n=2,...,k-1.$$

Note that $E\{e^{-(\omega_i \mathbf{S}_i + \omega_j \mathbf{S}_j)}\}$ only depends on i and j through the difference $j-i$.
From (2.5) the correlation coefficient, $corr(\mathbf{S}_i,\mathbf{S}_j)$, can easily be obtained:

$$corr(\mathbf{S}_i,\mathbf{S}_j) = 1-C_{j-i}(1-\rho), \quad 1 \leqslant i < j \leqslant k. \tag{2.7}$$

It follows from (2.6) and (2.7) that $corr(\mathbf{S}_i,\mathbf{S}_j)$ as a function of i and j only depends on $j-i$, and that it decreases if $j-i$ grows.

(iii) The Laplace-Stieltjes transform of the distribution of a customer's total time spent in the system until the end of his k-th pass, $\mathbf{S}^{(k)} := \mathbf{S}_1+...+\mathbf{S}_k$, can be obtained from (2.3) by substituting $\omega_j=\omega_0$, $j=1,...,k$. To derive an expression for the variance of this sojourn time, $var(\mathbf{S}^{(k)})$, it is convenient to use the formula

$$var(\mathbf{S}^{(k)}) = k\; var(\mathbf{S}_1) + 2\sum_{i=1}^{k}\sum_{j=i+1}^{k} cov(\mathbf{S}_i,\mathbf{S}_j).$$

From this formula, (i) and (ii),

$$var(\mathbf{S}^{(k)}) = (\frac{\beta}{1-\rho})^2[k^2-2(1-\rho)\sum_{j=1}^{k-1} jC_{k-j}], \tag{2.8}$$

with $C_1, \ldots, C_{k-1}$ given by (2.6). In fact [3] does not give (2.6), but a somewhat more complicated recurrence relation for the C_n:

$$C_1 = 1, \tag{2.9}$$
$$C_n = (1+\lambda\beta)C_{n-1}-\lambda\beta\sum_{l=2}^{n-1} q(n-l)(1-p(n-l))C_{l-1}, \quad n=2,...,k-1.$$

Noting that $q(n-l)(1-p(n-l))=q(n-l)-q(n-l+1)$, and splitting the sum in (2.9), we obtain

$$C_n-\lambda\beta\sum_{l=2}^{n}q(n-l+1)C_{l-1} = C_{n-1}-\lambda\beta\sum_{l=2}^{n-1}q(n-l)C_{l-1},$$

from which (2.6) follows. Taking generating functions and using (2.2) leads to

$$C(z):=\sum_{n=1}^{\infty}C_nz^n = \frac{z}{(1-z)(1-\lambda\beta\sum_{i=1}^{\infty}q(i)z^i)} = \frac{z}{(1-z)(1-\lambda\beta\frac{z}{1-z}(1-Q(z)))}, \quad |z|<1. \tag{2.10}$$

The sequence $C_1,C_2,...$ is non-decreasing and, cf. (2.7), limited from above. Hence $\lim_{n\to\infty} C_n$ exists; an Abelian theorem now implies that

$$\lim_{n\to\infty} C_n = \lim_{z\to 1}(1-z)C(z) = \frac{1}{1-\rho}. \tag{2.11}$$

3. The M/G/1 processor sharing queue

3.1 Introduction

In this section we show how the results given in Section 2 for the M/M/1 queue with feedback can be used to analyse the sojourn time in the M/G/1 PS queue. We apply a limiting procedure, in which $\beta\to 0$ while the feedback probabilities approach one in such a way that the mean total required service time, $\hat{\beta}$, remains a positive constant. We restrict ourselves to those service times, τ^{PS}, in the PS queue which are composed of negative exponentially distributed stages:

$$E\{\exp(-\omega_0\tau^{PS})\} = \sum_{j=1}^{m}\alpha_j\prod_{i=1}^{r_j}\frac{1}{1+\hat{\beta}_{ij}\omega_0}, \tag{3.1}$$

with $\alpha_1,\ldots,\alpha_m>0$, $\sum_{j=1}^{m}\alpha_j=1$, $r_1,...,r_m$ positive integers (cf. Kleinrock [6, Vol. 1, p. 145]); note that this class of distributions contains the Erlang, hyperexponential and Coxian distributions, and that arbitrary probability distributions of nonnegative random variables can be arbitrarily closely approximated by distributions from this class. This choice of service time distribution enables us to choose the feedback probabilities (hence $Q(z)$) such that τ^{PS} and the total required service time τ^{FB} in the FB (feedback) queue have exactly the same distribution - not just in the limit $\beta\to 0$, but for a wide range of values of β. Observe that, cf. (2.2),

$$E\{\exp(-\omega_0\tau^{FB})\} = \sum_{i=1}^{\infty}q(i)(1-p(i))(\frac{1}{1+\beta\omega_0})^i = Q(\frac{1}{1+\beta\omega_0}), \quad \text{Re } \omega_0\geq 0. \tag{3.2}$$

Now choose

$$Q(z) = \sum_{j=1}^{m}\alpha_j\prod_{i=1}^{r_j}\frac{(1-b_{ij})z}{1-b_{ij}z}, \tag{3.3}$$

with

$$b_{ij} = 1-\beta/\hat{\beta}_{ij} > 0, \quad i=1,...,r_j,\ j=1,...,m. \tag{3.4}$$

Then

$$E\{\exp(-\omega_0\tau^{FB})\} = \sum_{j=1}^{m} \alpha_j \prod_{i=1}^{r_j} \frac{\beta/\hat{\beta}_{ij}}{1+\beta\omega_0-(1-\beta/\hat{\beta}_{ij})} = \sum_{j=1}^{m} \alpha_j \prod_{i=1}^{r_j} \frac{1}{1+\hat{\beta}_{ij}\omega_0} = E\{\exp(-\omega_0\tau^{PS})\}. \quad (3.5)$$

As an example, consider the case of Bernoulli feedback: $Q(z) = (1-p)z/(1-pz)$. In this case,

$$E\{\exp(-\omega_0\tau^{PS})\} = E\{\exp(-\omega_0\tau^{FB})\} = \frac{1}{1+(\beta/(1-p))\omega_0} = \frac{1}{1+\hat{\beta}\omega_0}. \quad (3.6)$$

Hence the total required service times in both the FB queue and the PS queue are negative exponentially distributed with mean $\hat{\beta}=\beta/(1-p)$. When $\beta\to 0$, $p\to 1$, performance measures in the FB queue clearly approach corresponding performance measures in the PS queue (although it should be noted that we have not formally proved this).

In principle we could obtain the LST of the distribution of the sojourn time, $E\{\exp(-\omega_0 \mathbf{S}^{PS})\}$, in the M/G/1 PS queue from (2.3) and (2.4). For the M/D/1 case this has been done in [2], and for the M/M/1 case the procedure has been outlined in [3]. The general case will be studied in [1]; here we restrict ourselves to the (conditional) mean and variance of the sojourn time.

3.2 The mean sojourn time

In the M/G/1 PS queue, the mean sojourn time, $E\{\mathbf{S}^{PS}\}$, of a customer with service demand $\tau^{PS}=x$ is linear in x (cf. Kleinrock [6, Vol. 2]):

$$E\{\mathbf{S}^{PS}|\tau^{PS}=x\} = \frac{x}{1-\rho}. \quad (3.7)$$

In this subsection we show how this result immediately follows from the feedback results of Section 2. Consider a newly arriving customer, say C, who will obtain exactly k services. Statement (i) below Theorem 2.1 implies that the mean sojourn time of C is given by

$$E\{\mathbf{S}^{(k)}\} = \frac{k\beta}{1-\rho}. \quad (3.8)$$

Choose $Q(z)$ as in (3.3) and apply the limiting procedure of Subsection 3.1, taking $\beta=x/k$ and letting $k\to\infty$. The total required service time of C approaches the constant x (indeed, $(1+\beta\omega_0)^{-k} = (1+x\omega_0/k)^{-k} \to e^{-x\omega_0}$). Hence, for $k\to\infty$, C can be viewed as a customer with service request x in the M/G/1 PS queue with service time distribution characterized by (3.1). Formula (3.7) now immediately follows from (3.8).

3.3 The variance of the sojourn time

The sojourn time variance for a customer with service request x in the M/G/1 PS queue, $var(\mathbf{S}^{PS}|\tau^{PS}=x)$, can be obtained from (2.8) by taking $\beta=x/k$ and letting $k\to\infty$ in the way described in Subsection 3.1. Below we first derive $var(\mathbf{S}^{PS}|\tau^{PS}=x)$ for the M/M/1 PS queue, and then for the M/G/1 PS queue. This leads to a simple explicit expression for the asymptotic behaviour of $var(\mathbf{S}^{PS}|\tau^{PS}=x)$ for very large ($x\to\infty$) and very small ($x\to 0$) service requests. Details of the analysis, and numerical results, are given in [1].

The M/M/1 PS queue

As observed in (3.6), the choice $Q(z) = (1-p)z/(1-pz)$ leads, in the FB queue as well as the PS queue, to a negative exponentially distributed total service time with mean $\beta/(1-p) = \hat{\beta}$. To obtain an explicit expression for $var(\mathbf{S}^{(k)})$, see (2.8), we derive C_n, $n=1,2,...$, from (2.10). Substituting

$Q(z) = (1-p)z/(1-pz)$ into (2.10) yields

$$C(z) = z\frac{1-pz}{(1-z)(1-(\lambda\beta+p)z)}. \tag{3.9}$$

Rewriting the right-hand side of (3.9) as

$$z[U_1\frac{1}{1-z} + U_2\frac{1}{1-(\lambda\beta+p)z}],$$

it follows that

$$C_n = U_1 + U_2 x_2^{n-1}, \quad n=1,2,..., \tag{3.10}$$

with $U_1=1/(1-\rho)$, $U_2=-\rho/(1-\rho)$, $x_2=\lambda\beta+p$. Substituting (3.10) into (2.8) yields

$$var(\mathrm{S}^{(k)}) = (\frac{\beta}{1-p})^2[k - 2(1-\rho)U_2\frac{x_2^k+k(1-x_2)-1}{(1-x_2)^2}] = \tag{3.11}$$
$$(\frac{\beta}{1-\rho})^2[k + \frac{2\rho}{1-p}(\frac{k}{1-\rho} - \frac{1-(\lambda\beta+p)^k}{(1-p)(1-\rho)^2})].$$

Let $\hat{\beta}$ be the mean service time for the M/M/1 PS queue and let x be the service time of a tagged customer (cf. Subsection 3.2). Substitute $\beta=x/k$ and $p=1-x/k\hat{\beta}$ into (3.11) (cf. (3.4)). Letting $k\to\infty$ leads to $var(\mathrm{S}^{PS}|\tau^{PS}=x)$:

$$var(\mathrm{S}^{PS}|\tau^{PS}=x) = \lim_{k\to\infty} var(\mathrm{S}^{(k)}) = \frac{2\rho\hat{\beta}x}{(1-\rho)^3} - \frac{2\rho\hat{\beta}^2}{(1-\rho)^4}[1-e^{-x(1-\rho)/\hat{\beta}}], \tag{3.12}$$

a result previously obtained by Ott [8]. Note that the sojourn time variance depends linearly on x for $x\to\infty$:

$$var(\mathrm{S}^{PS}|\tau^{PS}=x) \sim \frac{2\rho\hat{\beta}}{(1-\rho)^3}x - \frac{2\rho\hat{\beta}^2}{(1-\rho)^4}, \quad x\to\infty, \tag{3.13}$$

(see also Kleinrock [6, Vol.2, p. 170]), whereas it depends quadratically on x for $x\to 0$:

$$var(\mathrm{S}^{PS}|\tau^{PS}=x) \sim \frac{\rho}{(1-\rho)^2}x^2 - \frac{\lambda}{3(1-\rho)}x^3, \quad x\to 0. \tag{3.14}$$

The M/G/1 PS queue

We now derive an expression for $var(\mathrm{S}^{PS}|\tau^{PS}=x)$ for the M/G/1 PS queue, in particular showing that the above asymptotic properties hold for general service time distributions. We only give a brief outline of the analysis. Details are given in [1]. We consider service time distributions with LST as in (3.1), by choosing $Q(z)$ as in (3.3), (3.4):

$$Q(z) = \sum_{j=1}^{m}\alpha_j\prod_{i=1}^{r_j}\frac{(1-b_{ij})z}{1-b_{ij}z} = \sum_{j=1}^{m}\alpha_j\prod_{i=1}^{r_j}\frac{\beta z/\hat{\beta}_{ij}}{1-(1-\beta/\hat{\beta}_{ij})z}. \tag{3.15}$$

Analogously to the M/M/1 case analysed above, (2.10) and (3.15) lead to:

$$C_n = U_1 + U_2x_2^{n-1} + \cdots + U_Lx_L^{n-1}, \quad n=1,2,..., \quad \text{with } L \text{ at most } \sum_{j=1}^{m}r_j + 1, \tag{3.16}$$

where $1/x_2, \dots, 1/x_L$ are the roots of

$$1-\lambda\beta\frac{z}{1-z}(1-Q(z)) = 0. \tag{3.17}$$

Note. We assume that there are no multiple roots. This can be easily shown for the Erlang and hyperexponential cases, and seems to be true generally; we'll go into this in detail in [1].

We now prove some properties of x_i and U_i that will be used in the sequel.

LEMMA 3.1
(i) $|x_i| < 1, \quad i=2,...,L$;
(ii) x_i *can be written as*

$$x_i = 1-\beta a_i, \tag{3.18}$$

with a_i *independent of* β, *and* Re $a_i > 0, \quad i=2,...,L$;
(iii) U_i *is independent of* β, $i=1,...,L$, *and* $U_1 = 1/(1-\rho)$.

PROOF
Noting that (see (2.10)),

$$1-\lambda\beta\frac{z}{1-z}(1-Q(z)) = 1-\lambda\beta\sum_{i=1}^{\infty} q(i)z^i,$$

and $\lambda\beta\sum_{i=1}^{\infty} q(i) = \rho < 1$, it follows immediately that $|x_i|<1$, $i=2,...,L$. To prove (ii), substitute (3.15) into (3.17) and replace z by $1/(1-\beta\tilde{z})$. Then (3.17) reduces to

$$1 + \frac{\lambda}{\tilde{z}} - \frac{\lambda}{\tilde{z}}\sum_{j=1}^{m} \alpha_j \prod_{i=1}^{r_j} \frac{1}{1-\hat{\beta}_{ij}\tilde{z}} = 0. \tag{3.19}$$

Since $1/x_i$ is a root of (3.17), $(1-x_i)/\beta$ is a root of (3.19). The fact that β does not occur in the left-hand side of (3.19) implies that $1-x_i$ depends linearly on β. The statement concerning Re $a_i > 0$ now follows from (i).
Because $1/x_i$ is assumed to be a single root of (3.17) it follows from (3.16) and (2.10) that

$$U_i\frac{1}{x_i} = \lim_{z\to 1/x_i} (1-zx_i)C(z) = \lim_{\tilde{z}\to a_i} (1-\frac{x_i}{1-\beta\tilde{z}})C(\frac{1}{1-\beta\tilde{z}}). \tag{3.20}$$

Observing that $\beta C(\frac{1}{1-\beta\tilde{z}})$ is independent of β, it is found that

$$U_i = \lim_{\tilde{z}\to a_i} x_i(1-\frac{x_i}{1-\beta\tilde{z}})C(\frac{1}{1-\beta\tilde{z}})$$

is independent of β.
Finally, $U_1 = 1/(1-\rho)$ follows from (2.11), (3.16) and (i).

Substituting (3.16) into (2.8) yields (cf. (3.11))

$$var(\mathbf{S}^{(k)}) = (\frac{\beta}{1-\rho})^2[k - 2(1-\rho)\sum_{j=2}^{L} U_j \frac{x_j^k + k(1-x_j)-1}{(1-x_j)^2}]. \tag{3.21}$$

Now, let x be the service time of a tagged customer, and take $\beta = x/k$. For $k \to \infty$, $var(\mathbf{S}^{PS} \mid \boldsymbol{\tau}^{PS} = x)$ follows from (3.21) and (i) of Lemma 3.1; integrating over x yields the unconditional sojourn time variance. We collect these results in

THEOREM 3.1 *In the M/G/1 PS queue with service time LST given by (3.1),*

$$var(\mathbf{S}^{PS} \mid \boldsymbol{\tau}^{PS} = x) = \frac{2}{1-\rho} \sum_{j=2}^{L} (1/a_j)^2 U_j [1 - x a_j - e^{-x a_j}], \tag{3.22}$$

$$var(\mathbf{S}^{PS}) = \frac{2}{1-\rho} \sum_{j=2}^{L} (1/a_j)^2 U_j [1 - \hat{\beta} a_j - E\{e^{-a_j \boldsymbol{\tau}^{PS}}\}] + \frac{E\{(\boldsymbol{\tau}^{PS})^2\} - \hat{\beta}^2}{(1-\rho)^2}. \tag{3.23}$$

From (ii) and (iii) of Lemma 3.1 it follows that

$$\lim_{x \to \infty} \frac{2}{1-\rho} \sum_{j=2}^{L} (1/a_j)^2 U_j e^{-x a_j} = 0 .$$

Hence, the sojourn time variance is asymptotically linear in x:

$$var(\mathbf{S}^{PS} \mid \boldsymbol{\tau}^{PS} = x) \sim \frac{2}{1-\rho} \sum_{j=2}^{L} (1/a_j)^2 U_j (1 - x a_j), \qquad x \to \infty . \tag{3.24}$$

From (3.16) and (3.18),

$$\sum_{j=2}^{L} (1/a_j) U_j = \beta \sum_{j=2}^{L} U_j \frac{1}{1-x_j} = \beta \sum_{n=1}^{\infty} (C_n - \frac{1}{1-\rho}) ,$$

and

$$\sum_{j=2}^{L} (1/a_j)^2 U_j = \beta^2 \sum_{j=2}^{L} U_j \frac{1}{(1-x_j)^2} = \beta^2 \sum_{n=1}^{\infty} n (C_n - \frac{1}{1-\rho}) .$$

It can be derived from (2.10) that

$$\beta \sum_{n=1}^{\infty} (C_n - \frac{1}{1-\rho}) = - \frac{\lambda \hat{\beta}_2}{2(1-\rho)^2} ,$$

and

$$\beta^2 \sum_{n=1}^{\infty} n (C_n - \frac{1}{1-\rho}) = - \frac{1}{2(1-\rho)^3} [\frac{\lambda}{3} \hat{\beta}_3 + \lambda^2 (\frac{1}{2} \hat{\beta}_2^2 - \frac{1}{3} \hat{\beta} \hat{\beta}_3)] ,$$

with $\hat{\beta}_i := E\{(\boldsymbol{\tau}^{PS})^i\}$, $i = 2,3$.
Hence, from (3.24),

$$var(\mathbf{S}^{PS} \mid \boldsymbol{\tau}^{PS} = x) \sim \frac{\lambda \hat{\beta}_2}{(1-\rho)^3} x - \frac{1}{(1-\rho)^4} [\frac{\lambda}{3} \hat{\beta}_3 + \lambda^2 (\frac{1}{2} \hat{\beta}_2^2 - \frac{1}{3} \hat{\beta} \hat{\beta}_3)] , \quad x \to \infty. \tag{3.25}$$

Noting that, in (3.22),

$$1-xa_j-e^{-xa_j} = -\sum_{i=2}^{\infty}\frac{(-xa_j)^i}{i!}, \quad j=2,...,L,$$

and using

$$\sum_{j=2}^{L} U_j = C_1-\frac{1}{1-\rho} = -\frac{\rho}{1-\rho}, \qquad \sum_{j=2}^{L} U_j x_j = C_2-\frac{1}{1-\rho} = \lambda\beta-\frac{\rho}{1-\rho},$$

it is found that

$$var(\mathbf{S}^{PS}\,|\,\boldsymbol{\tau}^{PS}=x) \sim \frac{\rho}{(1-\rho)^2}x^2-\frac{\lambda}{3}\frac{1}{1-\rho}x^3, \quad x\to 0. \tag{3.26}$$

This expression appears to be independent of the service time distribution, apart from its first moment (cf. also (3.14)). The quadratic behaviour of $var(\mathbf{S}^{PS}\,|\,\boldsymbol{\tau}^{PS}=x)$ for small service requests x should be contrasted with the linear behaviour for large x.

Note added in proof. Formula (3.26) slightly generalizes Theorem 1 of Yashkov [12]. Formula (3.25) is contained in Theorem 2 of [12]; but for the service time distributions defined in (3.1), Yashkov's theorem follows immediately from (3.22).

REFERENCES

1 VAN DEN BERG, J.L. (1988). The M/G/1 queue with processor sharing and its relation to a feedback queue, *to appear.*

2 VAN DEN BERG, J.L., BOXMA, O.J., GROENENDIJK, W.P. (1987). Sojourn times in the M/G/1 queue with deterministic feedback, *to appear in Stochastic Models.*

3 VAN DEN BERG, J.L., BOXMA, O.J. (1987). Sojourn times in feedback queues, Report OS-R8710, Centre for Mathematics and Computer Science, Amsterdam.

4 FONTANA, B., DIAZ BERZOSA, C. (1985). M/G/1 queue with N-priorities and feedback: joint queue-length distributions and response time distribution for any particular sequence. In: *Teletraffic Issues in an Advanced Information Society,* ITC-11, ed. M. Akiyama. North-Holland Publ. Cy., Amsterdam, 452-458.

5 KITAYEV, M.YU., YASHKOV, S.F. (1978). Distribution of the conditional sojourn time in a system with division of time of servicing, *Engrg. Cybernetics* **16**, 162-167 (Engl. translation of Russian Journal *Izv. Akad. Nauk SSR. Tekh. Kibern.* (1978) No. 4, 211-215).

6 KLEINROCK, L. (1975,1976). Queueing Systems, Vol. I,II (Wiley, New York).

7 NELSON, R.D. (1987). Expected response time for a FCFS feedback queue with multiple classes, Research Report RC 13221, IBM Research Division, T.J. Watson Research Center, Yorktown Heights, N.Y.

8 OTT, T.J. (1984). The sojourn-time distribution in the M/G/1 queue with processor sharing. *J. Appl. Prob.* **21**, 360-378.

9 SCHASSBERGER, R. (1984). A new approach to the M/G/1 processor-sharing queue, *Adv. Appl. Prob.* **16**, 202-213.

10 YASHKOV, S.F. (1983). A derivation of response time distribution for an M/G/1 processor-sharing queue, *Problems Contr. & Info. Theory* **12**, *133-148.*

11 YASHKOV, S.F. (1987). Processor-sharing queues: some progress in analysis, *Queueing Systems* **2**, 1-17.

12 YASHKOV, S.F. (1986). A note on asymptotic estimates of the sojourn time variance in the M/G/1 queue with processor-sharing, *Syst. Anal. Model. Simul.* **3**, 267-269.