Queueing Systems 9 (1991) 365–402

Invited paper

The $M/G/1$ queue with processor sharing and its relation to a feedback queue

J.L. van den Berg [1]

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

O.J. Boxma

Centre for Mathematics and Computer Science, P.O. Box 4079, 1009 AB Amsterdam, The Netherlands
Faculty of Economics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Received 29 March 1990; revised 11 February 1991

The central model of this paper is an $M/M/1$ queue with a general probabilistic feedback mechanism. When a customer completes his ith service, he departs from the system with probability $1-p(i)$ and he cycles back with probability $p(i)$. The mean service time of each customer is the same for each cycle. We determine the joint distribution of the successive sojourn times of a tagged customer at his loops through the system. Subsequently we let the mean service time at each loop shrink to zero and the feedback probabilities approach one in such a way that the mean total required service time remains constant. The behaviour of the feedback queue then approaches that of an $M/G/1$ processor sharing queue, different choices of the feedback probabilities leading to different service time distributions in the processor sharing model. This is exploited to analyse the sojourn time distribution in the $M/G/1$ queue with processor sharing.

Some variants are also considered, viz., an $M/M/1$ feedback queue with additional customers who are always present, and an $M/G/1$ processor sharing queue with feedback.

Keywords: $M/M/1$ feedback queue; $M/G/1$ processor sharing queue; sojourn times.

1. Introduction

In the past decade considerable progress has been made in the analysis of sojourn time processes in networks of queues; cf. the survey paper [8]. However, results are still very scarce when the possibility exists of customers overtaking

[1] Present address: PTT Research, P.O. Box 421, 2260 AK Leidschendam, The Netherlands.

each other. One of the simplest forms of overtaking occurs in the single server queue with the processor sharing discipline. Under this discipline, when j customers are present all receive service simultaneously at a service rate $1/j$. Hence the sojourn time of a tagged customer is not only determined by the number of customers (and their remaining work) found upon his arrival, but also by later arrivals. Such later arrivals may in fact overtake the tagged customer; the resulting dependencies severely complicate the analysis of the sojourn times.

In 1970 Coffman et al. [10] have obtained the Laplace–Stieltjes transform (LST) of the sojourn time distribution in the $M/M/1$ queue with processor sharing ($M/M/1$ PS). More than ten years later Yashkov [31], Schassberger [25] and Ott [23] independently derived the LST of the sojourn time distribution in the $M/G/1$ PS queue. The analysis in their studies is quite intricate. In this paper we present a different approach to the sojourn time problem in the $M/G/1$ PS queue. The main advantage of our approach is its intrinsic simplicity. We start from an $M/M/1$ FCFS (First-Come-First-Served) queue with feedback. When a customer has completed his ith service, he departs from the system with probability $1-p(i)$ and he is fed back to the end of the queue with probability $p(i)$. The joint distribution of the numbers of customers being in their first, second,... loop has a product form. We exploit this product form to give a straightforward (although rather elaborate) derivation of the sojourn time distribution of a tagged customer in the $M/M/1$ feedback queue. The $M/G/1$ PS queue is obtained from the $M/M/1$ feedback queue via a limiting procedure. We let the feedback probabilities approach one and the mean service time at each loop approach zero, such that a customer's total required mean service time remains constant. Different choices of the feedback probabilities lead to different service time distributions in the PS queue. Application of this limiting procedure to the sojourn time results obtained for the $M/M/1$ feedback queue leads to results for the corresponding quantities in the PS queue. Our method gives much insight into some basic sojourn time properties of the $M/G/1$ PS queue, like the fact that the mean conditional sojourn time of a customer with service request x is linear in x. Another advantage of our approach is that it opens up possibilities for obtaining sojourn time approximations for the $M/G/1$ PS queue. The idea of using a feedback queue to study sojourn times in a PS queue has also been employed by Schassberger [25], but in his feedback model the service times at each loop are deterministic and there is no product form.

Although the analysis of a PS queue via a product-form feedback queue occupies a central place in this study, the feedback queue is also of interest in itself. We present a detailed analysis of the feedback queue, and of some of its variants. The paper is organized as follows. The feedback model is described in section 2 and analyzed in section 3. In particular we obtain the joint distribution of the successive sojourn times (at each loop) of a tagged customer and the queue lengths found at the beginning of each sojourn. Sections 4, 5 and 6 are devoted to the sojourn time in the $M/G/1$ PS queue; section 4 considers the limiting procedure leading from the feedback queue to the PS queue, while sections 5 and 6 are concerned with the variance and the distribution of the sojourn time. Section 7 considers the $M/M/1$ feedback queue with some additional permanent customers – customers who are fed back after *each* service; its limiting PS counterpart is taken into consideration in section 8. Finally, in section 9 the $M/G/1$ PS queue *with feedback* is studied.

RELATED LITERATURE

We refer to the survey paper [32] of Yashkov for literature on queues with processor sharing. Concerning feedback queues, a pioneering study is Takács [26]. He considers the $M/G/1$ FCFS queue with so-called Bernoulli feedback: after each service, a customer leaves with fixed probability $1-p$ and returns to the end of the queue with probability p. His main result is a recurrence relation for the LST and generating function of the joint distribution of a customer's total sojourn time and the number of customers present in the system after a certain number of services. An implicitly used (although not explicitly mentioned) observation leading to this result is that for a tagged customer the joint process of successive service completion epochs and queue length at these epochs is a Markov renewal process. In fact, a similar observation is the basis for many other feedback studies, including ours. See Disney and Kiessler [13] for an extensive and fundamental discussion of Markov renewal processes in queueing networks, with an emphasis on traffic flows and queue lengths. Disney and König [14] give an overview of literature concerning Bernoulli feedback models. For our purposes it suffices to mention here the following feedback studies. Disney [12] emphasizes the Markov renewal approach in his formal derivation of the sojourn time distribution for the $M/G/1$ queue with Bernoulli feedback. Without going into details, he also observes that this approach allows a more general feedback mechanism than Bernoulli feedback. Doshi and Kaufman [15] derive, for the $M/G/1$ queue with Bernoulli feedback, the LST of the joint distribution of the sojourn times of a customer in his successive passes through the system. Lam and Shankar [22] consider basically the same $M/M/1$ feedback model with general feedback mechanism as we do. They derive the total sojourn time distribution, which becomes a special case of our result for the joint distribution of successive sojourn times. Hunter [18] considers single server queues with state-dependent feedback and finite waiting room. In particular, he studies an appropriately constructed Markov renewal process which describes the behaviour of the system starting at the arrival of a tagged customer; the sojourn time of the tagged customer relates to a first passage time in this process. For some special cases, like the $M/M/1/2$ queue with Bernoulli feedback, this approach leads to the derivation of explicit expressions for the LST of the distribution of the total sojourn time. Mean sojourn times are obtained for the $M/M/1/N$ queue with Bernoulli feedback. Hunter also gives a brief survey of the literature on sojourn times in feedback models.

The present paper is a sequel to [6] which studied the case of deterministic feedback (each customer makes exactly N loops), [4] which considered the case of general feedback, and [5] which analyzed the mean and the variance of the sojourn time distribution in an $M/G/1$ PS queue as the limit of a feedback queue. To provide a comprehensive overview of our approach and results, we have allowed some overlap with parts of [4] and [5]. We refer to the thesis [3] for a more detailed discussion of sojourn times in feedback and processor sharing queues.

2. Model description and preliminaries

We consider a single server queueing system with infinite waiting room, see fig. 1. 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 ith 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 ith time will be called a *type-i customer*. The service discipline is FCFS. It is assumed that 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

$$q(0) := 1, \quad q(i) := \prod_{j=0}^{i-1} p(j), \quad i = 1, 2, \ldots, \tag{2.1}$$

with

$$p(0) := 1.$$

Note that $\lambda q(i)$ is the arrival rate of type-i customers, $i = 1, 2, \ldots$. The total offered load to the queue per unit of time, denoted by ρ, is given by

$$\rho = \lambda\beta \sum_{i=1}^{\infty} q(i). \tag{2.2}$$

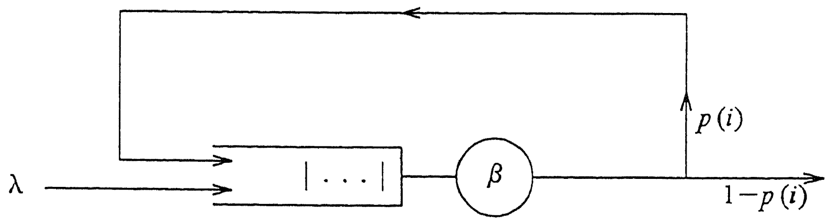

Fig. 1. The $M/M/1$ queue with general feedback.

For stability it is required that $\rho < 1$. We are interested in the following steady-state quantities:

- X_i: number of type-i customers in the system at an arbitrary epoch, $i = 1, 2, \dots$;
- $X_i^{(j)}$: number of type-i customers in the system at the jth service completion of a customer, $i = 1, 2, \dots, j = 1, 2, \dots$;
- $X_i^{(0)}$: number of type-i customers in the system at the arrival of a new customer, $i = 1, 2, \dots$;
- S_j: time required for the jth pass through the system (jth sojourn time), $j = 1, 2, \dots$;
- $S^{(k)}$: total sojourn time after k services: $S^{(k)} = \sum_{j=1}^{k} S_j$, $k = 1, 2, \dots$.

It is important to note that the system described above can be considered as a queueing network consisting of one queue with several types of customers. Type-i customers are fed back with probability $p(i)$ after service, and then change into type-$(i+1)$ customers, $i = 1, 2, \dots$. Because the service times are assumed to be independent exponentially distributed, the joint distribution of the number of type-i customers in the system at an arbitrary epoch, $i = 1, 2, \dots$, is of product-form type (Baskett et al. [1]): for $x_1, x_2, \dots = 0, 1, \dots$, and $x_1 + x_2 + \cdots < \infty$,

$$Pr\{X_1 = x_1, X_2 = x_2, \dots\} = (1-\rho)\left(\sum_{i=1}^{\infty} x_i\right)! \prod_{i=1}^{\infty} \frac{(\lambda\beta q(i))^{x_i}}{x_i!}. \tag{2.3}$$

It is convenient to have at our disposal the generating function of the joint queue length distribution. We have, for $|z_i| \leqslant 1$, $i = 1, 2, \dots$,

$$E\left\{\prod_{i=1}^{\infty} z_i^{X_i}\right\} = (1-\rho) \sum_{m=0}^{\infty} \underset{x_1+x_2+\cdots=m}{\sum_{x_1}\sum_{x_2}\cdots}\, m! \prod_{i=1}^{\infty} \frac{(\lambda\beta q(i) z_i)^{x_i}}{x_i!}$$

$$= (1-\rho) \sum_{m=0}^{\infty} \left(\sum_{i=1}^{\infty} \lambda\beta q(i) z_i\right)^m = \frac{1-\rho}{1 - \sum_{i=1}^{\infty} \lambda\beta q(i) z_i}. \tag{2.4}$$

The distribution of the total number of customers in the system coincides with the queue length distribution in an ordinary $M/M/1$ model:

$$E\{z^{\sum_{i=1}^{\infty} X_i}\} = \frac{1-\rho}{1-\rho z}, \quad |z| \leqslant 1,$$

i.e.

$$Pr\left\{\sum_{i=1}^{\infty} X_i = j\right\} = (1-\rho)\rho^j, \quad j = 0, 1, \dots. \tag{2.5}$$

We shall use these results in the next section.

3. The $M/M/1$ queue with feedback

In this section we present, in the form of Laplace–Stieltjes transforms and generating functions, an expression for the joint steady-state distribution of the successive sojourn times S_j, $j=1,\dots,k$, and the number of type-i customers, $X_i^{(j)}$, $i=1,2,\dots$, present at the jth service completion of a customer who is fed back at least $k-1$ times, $k=1,2,\dots$.

Let us follow a tagged customer from the moment he arrives as a type-1 customer until he completes his kth service. The PASTA property implies the equality of the joint queue length distribution at the epoch of a new arrival and at an arbitrary epoch. Hence, for Re $\omega_i \geqslant 0$, $|z_{i,j}| \leqslant 1$, $i=1,2,\dots, j=0,\dots,k$,

$$E\left\{e^{-(\omega_1 S_1+\dots+\omega_k S_k)}\left(\prod_{i=1}^{\infty} z_{i,0}^{X_i^{(0)}}\cdots\prod_{i=1}^{\infty} z_{i,k}^{X_i^{(k)}}\right)\right\}$$
$$=\sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty}\cdots Pr\{X_1=x_1, X_2=x_2,\dots\}E\Bigg\{e^{-(\omega_1 S_1+\dots+\omega_k S_k)}$$
$$\times\left(\prod_{i=1}^{\infty} z_{i,0}^{X_i^{(0)}}\cdots\prod_{i=1}^{\infty} z_{i,k}^{X_i^{(k)}}\right) \mid X_1^{(0)}=x_1, X_2^{(0)}=x_2,\dots\Bigg\}. \tag{3.1}$$

The conditional expectation in the RHS of (3.1) can be evaluated by using the fact that $(X_1^{(i+1)}, X_2^{(i+1)},\dots)$, which determines the distribution of S_{i+2}, is conditionally independent of $\{(X_1^{(j)}, X_2^{(j)},\dots), j=0,\dots,i-1; S_1,\dots,S_i\}$ given $\{(X_1^{(i)}, X_2^{(i)},\dots); S_{i+1}\}$, $i=1,\dots,k-1$, i.e., the joint process of successive service completion epochs and queue length vector at these service completion epochs is a *Markov renewal* process (cf. Çinlar [9, ch. 10]). The calculations, which are lengthy but quite straightforward, are omitted here; they can be found in appendix 2.1 of Van den Berg [3]. There it is shown that

$$E\left\{e^{-(\omega_1 S_1+\dots+\omega_k S_k)}\left(\prod_{i=1}^{\infty} z_{i,0}^{X_i^{(0)}}\cdots\prod_{i=1}^{\infty} z_{i,k}^{X_i^{(k)}}\right) \mid X_1^{(0)}=x_1, X_2^{(0)}=x_2,\dots\right\} \tag{3.2}$$
$$=\prod_{j=1}^{k} A_k(j,\omega,z)\prod_{i=1}^{\infty}\left(z_{i,0}f_k(i,\omega,z)\right)^{x_i},$$

with $\omega := (\omega_1,\dots,\omega_k)$, $z := ((z_{1,0}, z_{2,0},\dots),\dots,(z_{1,k}, z_{2,k},\dots))$, and

$$A_k(1,\omega,z) := \left[1+\beta\{\omega_k+\lambda(1-z_{1,k})\}\right]^{-1}, \tag{3.3}$$
$$A_k(2,\omega,z) := \left[1+\beta\{\omega_{k-1}+\lambda-\lambda z_{1,k-1}A_k(1,\omega,z)\right.$$
$$\left.\times\left[p(1)z_{2,k}+1-p(1)\right]\}\right]^{-1},$$
$$A_k(i,\omega,z) := \Big[1+\beta\big\{\omega_{k-i+1}+\lambda-\lambda z_{1,k-i+1}A_k(i-1,\omega,z)$$
$$\times\left[A_k(i-2,\omega,z)\left[\cdots\left[A_k(2,\omega,z)\right.\right.\right.$$

$$\times\left[A_k(1,\omega,z)\left[p(i-1)z_{i,k}+1-p(i-1)\right]\right.$$
$$\times p(i-2)z_{i-1,k-1}+1-p(i-2)\big]p(i-3)z_{i-2,k-2}$$
$$+1-p(i-3)\big]\cdots\big]$$
$$\times p(1)z_{2,k-i+2}+1-p(1)\big]\big\}\big]^{-1},\quad i=3,\ldots,k,$$

$$f_k(i,\omega,z):=A_k(k,\omega,z)\left[A_k(k-1,\omega,z)\left[\cdots\left[A_k(2,\omega,z)\right.\right.\right.$$
$$\times\left[A_k(1,\omega,z)\left[p(k+i-1)z_{k+i,k}+1-p(k+i-1)\right]\right.$$
$$\times p(k+i-2)z_{k+i-1,k-1}$$
$$+1-p(k+i-2)\big]p(k+i-3)z_{k+i-2,k-2}$$
$$+1-p(k+i-3)\big]\cdots\big]$$
$$\times p(i)z_{i+1,1}+1-p(i)\big],\quad i=1,2,\ldots. \tag{3.4}$$

Remark 3.1

The calculations leading to (3.2)–(3.4) reveal that the factor $(z_{i,0}f_k(i,\omega,z))^{x_i}$ in the RHS of (3.2) is due to the contribution to $\{(X_1^{(j)}, X_2^{(j)},\ldots),\ j=0,\ldots,k;\ S_1,\ldots,S_k\}$ induced by the x_i type-i customers present in the system upon the first arrival of the tagged customer, $i=1,2,\ldots$ (their own services and those of customers newly arriving during these services); the factor $\prod_{j=1}^{k}A_k(j,\omega,z)$ is due to the contribution induced by the tagged customer himself. These contributions are independent, cf. (3.2).

Substituting (2.3) and (3.2) into (3.1) and evaluating the summations we obtain our main result:

THEOREM 3.1

$$E\left\{e^{-(\omega_1S_1+\ldots+\omega_kS_k)}\left(\prod_{i=1}^{\infty}z_{i,0}^{X_i^{(0)}}\cdots\prod_{i=1}^{\infty}z_{i,k}^{X_i^{(k)}}\right)\right\}$$
$$=\frac{(1-\rho)\prod_{j=1}^{k}A_k(j,\omega,z)}{1-\lambda\beta\sum_{i=1}^{\infty}q(i)z_{i,0}f_k(i,\omega,z)},$$
$$\operatorname{Re}\omega_j\geqslant 0,\quad |z_{i,j}|\leqslant 1,\quad i=1,2,\ldots,\quad j=0,\ldots,k. \tag{3.5}$$

Introducing

$$M_k(i, \omega) := \prod_{j=1}^{i} \frac{1}{A_k(j, \omega, 1)}, \quad i = 1, \ldots, k,$$

$$M_k(0, \omega) := 1,$$

we prove

COROLLARY 3.1

The Laplace–Stieltjes transform 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

$$E\{e^{-(\omega_1 S_1 + \ldots + \omega_k S_k)}\} = \frac{1-\rho}{M_k(k, \omega) - \lambda\beta \sum_{l=2}^{k} q(k-l+1) M_k(l-1, \omega) - \left(\rho - \lambda\beta \sum_{i=1}^{k-1} q(i)\right)}, \tag{3.6}$$

with

$$M_k(i, \omega) = (1 + \beta\omega_{k-i+1}) M_k(i-1, \omega) + \lambda\beta \times \left[M_k(i-1, \omega) - q(i-1) - \sum_{j=2}^{i-1} q(i-j)(1-p(i-j)) M_k(j-1, \omega) \right],$$
$$i = 1, \ldots, k. \tag{3.7}$$

Proof

First substitute $z_{i,j} = 1$ into (3.3)–(3.5), $i = 1, 2, \ldots, j = 0, \ldots, k$, and subsequently rewrite these expressions in a form that is more suitable for obtaining sojourn time moments: Formula (3.3) leads to (3.7), and (3.4) leads for $i = 1, 2, \ldots$ to

$$f_k(i, \omega, 1) = \frac{1}{q(i)} \frac{1}{M_k(k, \omega)} \times \left[q(k+i-1) + \sum_{l=2}^{k} q(k+i-l)(1-p(k+i-l)) M_k(l-1, \omega) \right]. \tag{3.8}$$

Substitution of (3.8) into (3.5) proves the corollary. □

COROLLARY 3.2

The joint distribution of the number $X_i^{(j)}$ of type-i customers, $i = 1, 2, \ldots$, is the same for all $j = 0, 1, \ldots$; its generating function is given by (2.4).

Proof

The PASTA property proves that the generating function for $j = 0$ is given by (2.4). A simple calculation shows that the same expression holds for $j = 1$; it now readily follows that the statement holds for all $j = 0, 1, \ldots$. □

Remark 3.2

Corollary 3.2 is, in a more general context, known as the "*arrival theorem*" for product-form networks, see e.g. Walrand [30, section 4.4]. This theorem implies that an arriving type-i customer (who has just completed his $(i-1)$th service) "sees" the system as at an arbitrary epoch.

The fact that the joint queue length distribution at the arrival of a customer and after each of his passes is the same (cf. corollary 3.2), implies that the sojourn times S_j, $j = 1, \ldots, k$ have the same marginal distribution. S_1 can easily be obtained from (3.6) and (3.7) by taking $k = 1$. It is found that the sojourn times are negative exponentially distributed with mean $\beta/(1-\rho)$:

$$E\{e^{-\omega_j S_j}\} = \frac{1-\rho}{1-\rho+\beta\omega_j}, \quad j = 1, \ldots, k. \tag{3.9}$$

Note that this coincides with the sojourn time transform in an ordinary $M/M/1$ queue with mean service time β and arrival rate $\lambda\sum_{i=1}^{\infty} q(i)$, cf. (2.5).

In order to investigate the dependence between the ith and jth sojourn times we have computed the Laplace–Stieltjes transform of the joint distribution of S_i and S_j, $1 \leqslant i < j \leqslant k$. It is found from (3.6) and (3.7) that

$$E\{e^{-(\omega_i S_i + \omega_j S_j)}\} = \frac{1-\rho}{1-\rho+\beta\omega_i+\beta\omega_j+\beta^2\omega_i\omega_j C_{j-i}}, \quad 1 \leqslant i < j \leqslant k, \tag{3.10}$$

where C_{j-i} is determined by

$$C_1 = 1, \tag{3.11}$$

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

The last equation can be rewritten as

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

Using $C_1 = 1$ and extending (3.11) to $n = k,\ k + 1, \ldots$ it is easily seen that

$$C_1 = 1, \tag{3.12}$$

$$C_n = 1 + \lambda\beta \sum_{l=1}^{n-1} q(n-l)C_l, \quad n = 2, 3, \ldots.$$

For future use we determine the generating function of $\{C_1, C_2, \ldots\}$. Let

$$Q(z) := \sum_{i=1}^{\infty} q(i)(1-p(i))z^i, \quad |z| \leqslant 1, \tag{3.13}$$

$$C(z) := \sum_{i=1}^{\infty} C_i z^i, \quad |z| < 1. \tag{3.14}$$

Taking generating functions in (3.12) yields:

$$C(z) = \frac{z}{(1-z)\left(1 - \lambda\beta \frac{z}{1-z}(1-Q(z))\right)}, \quad |z| < 1. \tag{3.15}$$

Remark 3.3

It was pointed out by Prof. J.W. Cohen that the two-dimensional Laplace–Stieltjes transform given by (3.10) is of a type for which the corresponding joint probability density function, $f_{i,j}(\cdot, \cdot)$, is known. From the formula given in entry 8 of table B in Voelker and Doetsch [28, p. 208] it is found that, for $1 \leqslant i < j \leqslant k$,

$$f_{i,j}(x, y) = \frac{1-\rho}{\beta^2 C_{j-i}} \mathrm{e}^{-(x+y)/(\beta C_{j-i})} \times \sum_{m=0}^{\infty} \left(\frac{-xy}{\beta^2 C_{j-i}}\right)^m (1-\rho+1/C_{j-i})^m (1/m!)^2, \quad x, y \geqslant 0. \tag{3.16}$$

From (3.10) the correlation coefficient, $corr(S_i, S_j)$, can easily be obtained:

$$corr(S_i, S_j) = 1 - C_{j-i}(1-\rho), \quad 1 \leqslant i < j \leqslant k. \tag{3.17}$$

Note that $E\{\mathrm{e}^{-(\omega_i S_i + \omega_j S_j)}\}$ and $corr(S_i, S_j)$ only depend on i and j through the difference $j - i$. This property might also have been derived from corollary 3.2. Observing that in (3.12) $\Sigma_{l=2}^{n-1} q(n-l)(1-p(n-l)) \leqslant 1$ (remember that $q(n-l)(1-p(n-l))$ is the probability that a customer receives exactly $n-l$ services) it follows by induction that the row $\{C_n, n = 1, 2, \ldots\}$ is monotonically increasing. Hence, from (3.17), $corr(S_i, S_j)$ decreases if $j - i$ grows. In particular it can be proven, using (3.15), that $\lim_{n\to\infty} C_n = 1/(1-\rho)$, yielding $\lim_{j-i\to\infty} corr(S_i, S_j) = 0$. For $j - i = 1$, $corr(S_i, S_j) = \rho$. So, the successive sojourn times of a tagged customer are always correlated positively.

The Laplace–Stieltjes transform of the distribution of a customer's total time spent in the system until the end of his kth pass $S^{(k)} := S_1 + \dots + S_k$, can be obtained from (3.6) by substituting $\omega_j = \omega_0$, $j = 1, \dots, k$. Replacing the term $M_k(k, \omega)$ in the denominator of (3.6) by the RHS of (3.7) (with $i = k$) and substituting $\omega_j = \omega_0$, $j = 1, \dots, k$, it is found that for Re $\omega_0 \geq 0$, $k = 1, 2, \dots,$

$$E\{e^{-\omega_0 S^{(k)}}\} = \frac{1-\rho}{(1+\beta\omega_0)M_{k-1} - \lambda\beta \sum_{j=1}^{k-2} q(k-j-1)M_j - \left(\rho - \lambda\beta \sum_{i=1}^{k-2} q(i)\right)}, \tag{3.18}$$

where, for $n \leq k$, $M_n := M_k(n, \omega)$ is given by

$$M_0 := 1, \tag{3.19}$$

$$M_n := (1 + \beta\omega_0 + \lambda\beta)M_{n-1}$$
$$- \lambda\beta\left[q(n-1) + \sum_{l=2}^{n-1} q(n-l)(1-p(n-l))M_{l-1}\right], \quad n=1, 2, \dots.$$

$(q(0) := 1.)$

For future use we also introduce the generating function of the M_n's. From (3.19) it follows that

$$M(z) := \sum_{n=1}^{\infty} M_n z^n = z\frac{1 + \beta\omega_0 - \lambda\beta\frac{z}{1-z}(1-Q(z))}{1 - z(1+\beta\omega_0 + \lambda\beta) + \lambda\beta z Q(z)}. \tag{3.20}$$

From (3.9) it follows immediately that $E\{S^{(k)}\}$ is linear in k:

$$E\{S^{(k)}\} = \sum_{i=1}^{k} E\{S_i\} = k\frac{\beta}{1-\rho}. \tag{3.21}$$

The variance of the sojourn time, $var(S^{(k)})$, is determined by:

$$var(S^{(k)}) = \sum_{i=1}^{k} var(S_i) + 2\sum_{i=1}^{k} \sum_{j=i+1}^{k} cov(S_i, S_j)$$
$$= \left(\frac{\beta}{1-\rho}\right)^2 \left[k^2 - 2(1-\rho)\sum_{j=1}^{k-1} jC_{k-j}\right], \tag{3.22}$$

with $C_1, \dots, C_{k-1}$ given by (3.11). The Laplace–Stieltjes transform of the distribution of the total sojourn time S of an arbitrary customer is given by

$$E\{e^{-\omega_0 S}\} = \sum_{k=1}^{\infty} q(k)(1-p(k))E\{e^{-\omega_0 S^{(k)}}\}. \tag{3.23}$$

4. The $M/G/1$ processor sharing queue

In the previous section we have regarded the feedback model as an $M/M/1$ queue in which after each service it is decided whether or not the customer is fed back. In this section we consider the same model from another point of view, viz., as a round robin (time sharing) model in which a customer's service demand requires a stochastic number of exponentially distributed service quanta with mean length β. Obviously, the service requirements are completely determined by the feedback probabilities $p(1)$, $p(2), \dots,$ as defined in section 2. From this point of view it is intuitively clear that if the mean service time β shrinks to zero while the feedback probabilities go to one such that a customer's total required service time remains unchanged, the behaviour of the feedback queue approaches that of the $M/G/1$ processor sharing (PS) queue. Different choices of the feedback probabilities lead to different service time distributions in the PS queue.

The queue length process in a round robin type of queue is usually less amenable to mathematical analysis than the queue length process in its limiting case, a PS queue. It is well known (Kleinrock [20]) that the stationary distribution of the queue length, X^{PS}, in the $M/G/1$ PS queue is independent of the distribution of the required service time apart from its first moment:

$$Pr\{X^{PS}=j\} = (1-\rho)\rho^j, \quad j=0, 1, 2, \dots, \tag{4.1}$$

with ρ the offered load per unit of time. 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. Yashkov [31], Ott [23], Schassberger [25], and the survey of Yashkov [32]. The essence of Yashkov's approach [31] is a decomposition of the sojourn time of a (tagged) customer as the sum of "time delays", which are induced by the customers present in the system at the arrival of the tagged customer and by the tagged customer himself. These time delays include the influence of customers who arrive during the sojourn time of the tagged customer. It is shown that the time delays can be interpreted as lifetimes of some terminating branching process. The dynamics of the time delays is described by a set of integro-differential equations, derived by using ideas from branching theory. Ott [23] independently follows a similar approach, slightly generalizing Yashkov's result by obtaining the transform of the joint distribution of a customer's sojourn time and the number of other customers present at his departure. Schassberger [25] derives the sojourn time LST by analyzing a discrete-time queue with deterministic service quanta under a slight variation of the standard round robin discipline: a newly arriving customer immediately receives a quantum of service and only then joins the end of the queue. Using his sojourn time results for this round robin model and letting the quantum size shrink to zero he finds results for the corresponding sojourn times in the $M/G/1$ PS queue. He also gives the theoretical background of the weak convergence of the sojourn time distribution for the discrete-time round robin model to the distribution of the sojourn time in the PS model.

In this section and the next two we present a novel approach, which uses a similar idea as [25]: via a limiting procedure we obtain sojourn time results for the $M/G/1$ PS queue from known sojourn time results (obtained in section 3) for the $M/M/1$ queue with general feedback. The limiting procedure described above was first proposed by Van den Berg et al. [6]. In that paper it is shown how the distribution of the sojourn time in the $M/D/1$ PS queue follows immediately, by taking appropriate limits, from the sojourn time distribution in the $M/M/1$ queue with so-called deterministic feedback, in which each customer receives exactly N services. In Van den Berg and Boxma [5] this method has been used for the analysis of the sojourn time *mean* and *variance* in the processor sharing queue with general service times. In these papers the authors concluded on *intuitive* grounds that performance measures such as the sojourn time in the feedback model converge to the corresponding performance measures in the processor sharing queue. Only very recently a formal proof of this convergence has been given by Resing et al. [24]. They present a probabilistic coupling between the $M/G/1$ PS queue and the approximating sequence of $M/M/1$ feedback queues, which shows that the sojourn time of the nth customer in the feedback model converges almost surely to the corresponding quantity in the PS model. From this result they conclude the distributional convergence of the steady state sojourn times. The proof partially follows the same line of thought as Schassberger [25].

In this section we describe the limiting procedure that gives rise to processor sharing, and we derive the mean sojourn time. The sojourn time variance is derived in section 5, and in section 6 it is shown how the LST of the distribution of the sojourn time in the $M/G/1$ PS queue can be obtained.

THE LIMITING PROCEDURE

To go from feedback to processor sharing 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 the discussion to those service times, τ^{PS}, in the PS queue which are composed of negative exponentially distributed stages:

$$E\{\exp(-\omega_0 \tau^{PS})\} = \sum_{j=1}^{m} \alpha_j \prod_{i=1}^{r_j} \frac{1}{1+\hat{\beta}_{ij}\omega_0}, \tag{4.2}$$

with $\alpha_1, \dots, \alpha_m > 0$, $\sum_{j=1}^{m} \alpha_j = 1$, $r_1, \dots, r_m$ positive integers (cf. Kleinrock [19, 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 (cf. Tijms [27, p. 398]). This choice of service time distribution for the PS queue enables us to choose the feedback probabilities (hence $Q(z)$) such that τ^{PS} and the total required service time τ^{FB} in the 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. (3.13),

$$E\{\exp(-\omega_0\tau^{FB})\} = \sum_{i=1}^{\infty} q(i)(1-p(i))\left(\frac{1}{1+\beta\omega_0}\right)^i = Q\left(\frac{1}{1+\beta\omega_0}\right), \quad \text{Re } \omega_0 \geqslant 0. \tag{4.3}$$

Now choose

$$Q(z) = \sum_{j=1}^{m} \alpha_j \prod_{i=1}^{r_j} \frac{(1-p_{ij})z}{1-p_{ij}z}, \tag{4.4}$$

with

$$p_{ij} = 1 - \beta/\hat{\beta}_{ij} > 0, \quad i = 1,\dots,r_j, \quad j = 1,\dots,m. \tag{4.5}$$

Then

$$E\{\exp(-\omega_0\tau^{FB})\} = \sum_{j=1}^{m} \alpha_j \prod_{i=1}^{r_j} \frac{\beta/\hat{\beta}_{ij}}{1+\beta\omega_0 - (1-\beta/\hat{\beta}_{ij})} = \sum_{j=1}^{m} \alpha_j \prod_{i=1}^{r_j} \frac{1}{1+\hat{\beta}_{ij}\omega_0} = E\{\exp(-\omega_0\tau^{PS})\}. \tag{4.6}$$

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

$$E\{\exp(-\omega_0\tau^{PS})\} = E\{\exp(-\omega_0\tau^{FB})\} = \frac{1}{1+(\beta/(1-p))\omega_0} = \frac{1}{1+\hat{\beta}\omega_0}. \tag{4.7}$$

Hence the total required service times in both the feedback queue and the PS queue are negative exponentially distributed with mean $\hat{\beta} = \beta/(1-p)$.

When $\beta \to 0$, performance measures in the feedback queue clearly approach corresponding performance measures in the PS queue. Resing et al. [24] give a formal proof of the convergence of the sojourn time. Note that the queue length distribution in both models is the same for the *whole* range of possible β values, cf. (4.1) and (2.5). Below we shall focus mainly on sojourn times. In particular we are interested in the sojourn time of a customer *conditioned* on his required service time. This is an important performance measure for time sharing systems like PS queues, cf. Kleinrock [20]. We define for the PS queue

$S^{PS}(x)$: conditional sojourn time of a customer with service demand x;
S^{PS}: sojourn time of an arbitrary customer.

Obviously,

$$Pr\{S^{PS} < s\} = \int_{x=0}^{\infty} Pr\{S^{PS}(x) < s\} \, \mathrm{d}Pr\{\tau^{PS} < x\}, \quad s \geq 0. \tag{4.8}$$

The conditional sojourn time $S^{PS}(x)$ can be derived from the total sojourn time after k services, $S^{(k)}$, in the feedback queue in the following way. Choose $Q(z)$ for the feedback queue as in (4.4), (4.5), and consider a newly arriving customer, say C, who requires exactly k services. Then take $\beta = x/k$ and let $k \to \infty$. It is easily seen that the total required service time of C approaches the constant x. Indeed, the LST of C's total required service time equals $(1 + \beta\omega_0)^{-k} = (1 + x\omega_0/k)^{-k} \to \mathrm{e}^{-x\omega_0}$. *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* (4.2).

The limiting procedure described above will be applied to obtain results for the mean, the variance and the LST of the sojourn time in the PS queue from $E\{S^{PS}(x)\} = \lim_{k\to\infty} E\{S^{(k)}\}$, $var(S^{PS}(x)) = \lim_{k\to\infty} var(S^{(k)})$ and $E\{\mathrm{e}^{-\omega_0 S^{PS}(x)}\} = \lim_{k\to\infty} E\{\mathrm{e}^{-\omega_0 S^{(k)}}\}$ respectively. The results to be presented for the mean and the variance of the sojourn time are more general and more detailed than the results for the LST.

THE MEAN SOJOURN TIME

In the $M/G/1$ PS queue, the mean sojourn time of a customer with service demand x is linear in x (cf. Kleinrock [20]):

$$E\{S^{PS}(x)\} = \frac{x}{1-\rho}. \tag{4.9}$$

This well known result, which is sometimes proved rather heuristically, can be easily obtained from the feedback results of the previous section. The mean total sojourn time $E\{S^{(k)}\}$ of a customer who requires k services is linear in k, see (3.21). Apply the limiting procedure described above, taking $\beta = x/k$ and letting $k \to \infty$. Formula (4.9) now immediately follows from (3.21) which is essentially a result from product form theory.

5. The variance of the sojourn time

The sojourn time variance for a customer with service request x in the $M/G/1$ PS queue, $var(S^{PS}(x))$, can be obtained by applying the limiting procedure to (3.22). First, as an example, we derive $var(S^{PS}(x))$ for the $M/M/1$

PS queue. Next the analysis is extended to the PS queue with service time LST given by (4.2).

THE $M/M/1$ PS QUEUE

As observed in (4.7), the choice $Q(z)=(1-p)z/(1-pz)$ leads, in the feedback queue as well as in the limiting PS queue, to a negative exponentially distributed total service time with mean $\beta/(1-p)=\hat{\beta}$. To obtain an explicit expression for $var(S^{(k)})$, see (3.22), we derive C_n, $n=1,2,\dots,$ from (3.15). Substituting $Q(z)=(1-p)z/(1-pz)$ into (3.15) yields

$$C(z)=z\frac{1-pz}{(1-z)(1-(\lambda\beta+p)z)}. \tag{5.1}$$

Rewriting the right-hand side of (5.1) as

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

it follows that

$$C_n=U_1+U_2x_2^{n-1},\quad n=1,2,\dots, \tag{5.2}$$

with $U_1=1/(1-\rho)$, $U_2=-\rho/(1-\rho)$, $x_2=\lambda\beta+p$. Substituting (5.2) into (3.22) yields

$$\begin{aligned} var(S^{(k)}) &= \left(\frac{\beta}{1-\rho}\right)^2\left[k-2(1-\rho)U_2\frac{x_2^k+k(1-x_2)-1}{(1-x_2)^2}\right] \\ &= \left(\frac{\beta}{1-\rho}\right)^2\left[k+\frac{2\rho}{1-p}\left(\frac{k}{1-\rho}-\frac{1-(\lambda\beta+p)^k}{(1-p)(1-\rho)^2}\right)\right]. \end{aligned} \tag{5.3}$$

Let x be the service time of a tagged customer (cf. section 4). Substitute $\beta=x/k$ and $p=1-x/k\hat{\beta}$ into (5.3). Letting $k\to\infty$ leads to $var(S^{PS}(x))$:

$$var(S^{PS}(x))=\lim_{k\to\infty}var(S^{(k)})=\frac{2\rho\hat{\beta}x}{(1-\rho)^3}-\frac{2\rho\hat{\beta}^2}{(1-\rho)^4}\left[1-e^{-x(1-\rho)/\hat{\beta}}\right], \tag{5.4}$$

a result previously obtained by Ott [23]. Note that the sojourn time variance depends linearly on x for $x\to\infty$:

$$var(S^{PS}(x))\sim\frac{2\rho\hat{\beta}}{(1-\rho)^3}x-\frac{2\rho\hat{\beta}^2}{(1-\rho)^4},\quad x\to\infty, \tag{5.5}$$

(see also Kleinrock [20, p. 170]), whereas it depends quadratically on x for $x\to 0$:

$$var(S^{PS}(x))\sim\frac{\rho}{(1-\rho)^2}x^2-\frac{\lambda}{3(1-\rho)}x^3,\quad x\to 0. \tag{5.6}$$

THE $M/G/1$ PS QUEUE

We now derive an expression for $var(S^{PS}(x))$ for the $M/G/1$ PS queue. We consider service time distributions with LST as in (4.2), by choosing $Q(z)$ as in (4.4), (4.5):

$$Q(z) = \sum_{j=1}^{m} \alpha_j \prod_{i=1}^{r_j} \frac{(1-p_{ij})z}{1-p_{ij}z} = \sum_{j=1}^{m} \alpha_j \prod_{i=1}^{r_j} \frac{\beta z/\hat{\beta}_{ij}}{1-\left(1-\beta/\hat{\beta}_{ij}\right)z}. \tag{5.7}$$

Analogously to the $M/M/1$ case analyzed above, (3.15) and (5.7) lead to:

$$C_n = U_1 + U_2 x_2^{n-1} + \cdots + U_L x_L^{n-1}, \quad n = 1, 2, \ldots, \tag{5.8}$$

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

$$1 - \lambda\beta \frac{z}{1-z}(1 - Q(z)) = 0. \tag{5.9}$$

$U_1, \ldots, U_L$ are determined by

$$\frac{U_1 z}{1 - x_1 z} + \ldots + \frac{U_L z}{1 - x_L z} = C(z). \tag{5.10}$$

Note that in (5.8)–(5.10) we have used the following assumption:

ASSUMPTION 5.1

The roots $1/x_2, \ldots, 1/x_L$ of (5.9) are all disctinct.

Remark 5.1

Assumption 5.1 can be easily proved to hold for the Erlang and hyperexponential cases. We have found no example for which the roots are *not* distinct.

Remark 5.2

$1/x_2, \ldots, 1/x_L$ are the roots of a polynomial of degree $L - 1 \leqslant \Sigma_{j=1}^{m} r_j$, see (5.7), (5.9); (5.10) leads to a set of L linear equations from which $U_1, \ldots, U_L$ can be obtained.

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

LEMMA 5.1

(i) $|x_i| < 1, \quad i = 2, \ldots, L$;

(ii) x_i can be written as

$$x_i = 1 - \beta a_i, \tag{5.11}$$

with a_i independent of β, and Re $a_i > 0$, $i = 2, \ldots, L$;

(iii) U_i is independent of β, $i = 1, \ldots, L$, and $U_1 = 1/(1-\rho)$.

Proof

Noting that (see (3.15)),

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

and $\lambda\beta\sum_{i=1}^{\infty}q(i)=\rho<1$, it follows immediately that $|x_i|<1$, $i=2,\ldots,L$. To prove (ii), substitute (5.7) into (5.9) and replace z by $1/(1-\beta\tilde{z})$. Then (5.9) reduces to

$$1+\frac{\lambda}{\tilde{z}}-\frac{\lambda}{\tilde{z}}\sum_{j=1}^{m}\alpha_j\prod_{i=1}^{r_j}\frac{1}{1-\hat{\beta}_{ij}\tilde{z}}=0. \tag{5.12}$$

Since $1/x_i$ is a root of (5.9), $(1-x_i)/\beta=a_i$ is a root of (5.12). The fact that β does not occur in the left-hand side of (5.12) implies that $1-x_i$ depends linearly on β. The statement concerning Re $a_i>0$ now follows from (i).
It follows from (5.10) that

$$U_i\frac{1}{x_i}=\lim_{z\to 1/x_i}(1-zx_i)C(z)=\lim_{\tilde{z}\to a_i}\left(1-\frac{x_i}{1-\beta\tilde{z}}\right)C\left(\frac{1}{1-\beta\tilde{z}}\right). \tag{5.13}$$

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

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

is independent of β. Formula (3.15) implies that $\lim_{n\to\infty}C_n=1/(1-\rho)$; together with (5.8) and (i) this yields $U_1=1/(1-\rho)$. □

Substituting (5.8) into (3.22) gives (cf. (5.3))

$$var(S^{(k)})=\left(\frac{\beta}{1-\rho}\right)^2\left[k-2(1-\rho)\sum_{j=2}^{L}U_j\frac{x_j^k+k(1-x_j)-1}{(1-x_j)^2}\right]. \tag{5.14}$$

Now, let x be the service time of a tagged customer, and take $\beta=x/k$. For $k\to\infty$, $var(S^{PS}(x))$ follows from (5.14) and (i) of lemma 5.1; integrating $E\{(S^{PS}(x))^2\}=var(S^{PS}(x))+x^2/(1-\rho)^2$ over x and subtracting $(E\{S^{PS}\})^2=\hat{\beta}^2/(1-\rho)^2$ yields the unconditional sojourn time variance. We collect these results in

THEOREM 5.1

In the $M/G/1$ PS queue with service time LST given by (4.2),

$$var(S^{PS}(x))=\frac{2}{1-\rho}\sum_{j=2}^{L}(1/a_j)^2U_j\left[1-xa_j-e^{-xa_j}\right], \tag{5.15}$$

$$var(S^{PS}) = \frac{2}{1-\rho} \sum_{j=2}^{L} (1/a_j)^2 U_j\left[1 - \hat{\beta} a_j - E\{e^{-a_j \tau^{PS}}\}\right] + \frac{E\{(\tau^{PS})^2\} - \hat{\beta}^2}{(1-\rho)^2}, \tag{5.16}$$

with $a_2, \dots, a_L$ the roots of (5.12) and $U_2, \dots, U_L$ determined by (5.10), cf. remark 5.2; a_j and U_j are independent of x, $j = 2, \dots, L$.

Formula (5.15) shows that $var(S^{PS}(x))$ depends on the required service time x in a very simple way. It is convenient to use this formula for the analysis of the behaviour of the sojourn time variance when x varies. In [2] asymptotic results for $x \to \infty$ and $x \to 0$ have thus been derived. In particular, it has been shown there that (5.6) holds for the class of service time distributions given by (4.2).

6. The distribution of the sojourn time

Application of the limiting procedure to (3.18) yields the LST of the distribution of the sojourn time in the $M/G/1$ PS queue. The analysis can be performed along the same lines as the analysis of the sojourn time variance. It appears that the M_n's in (3.18) have similar properties as the C_n's in the previous section. However, there are some difficulties which did not arise in the analysis of the variance. These problems are due to the presence of the individual feedback probabilities contained in the $q(n)$'s in the denominator of (3.18). In general the $q(n)$'s are given by very complicated expressions and can not be explicitly determined for the whole class of service time distributions given by (4.2) (cf. (4.4), (4.5) and (4.6)). Therefore, we shall restrict ourselves below to a subclass of these service times, viz. mixtures of Erlang distributions: (cf. (4.2))

$$E\{e^{-\omega_0 \tau^{PS}}\} = \sum_{j=1}^{m} \alpha_j \left(\frac{1}{1 + \hat{\beta}_j \omega_0} \right)^{r_j}, \tag{6.1}$$

with $\alpha_1, \dots, \alpha_m \geq 0$, $\sum_{j=1}^{m} \alpha_j = 1$, $r_1, \dots, r_m$ positive integers. The corresponding feedback probabilities are determined by (cf. (4.4), (4.5))

$$Q(z) = \sum_{j=1}^{m} \alpha_j \left(\frac{(1-p_j)z}{1 - p_j z} \right)^{r_j}, \tag{6.2}$$

with $p_j = 1 - \beta/\hat{\beta}_j > 0$.
From (6.2) we find

$$q(l)(1 - p(l)) = \sum_{j=1}^{m} \alpha_j (1-p_j)^{r_j} \binom{l-1}{r_j - 1} p_j^{l - r_j}, \tag{6.3}$$

from which the $q(n)$'s can be obtained via:

$$q(n) = \sum_{l=n}^{\infty} q(l)(1-p(l)), \quad n = 1, 2, \dots. \tag{6.4}$$

Note that the (sub)class of distribution functions determined by (6.1) is still large enough to approximate the distribution of any non-negative random variable arbitrarily closely (cf. Tijms [27, p. 398]).

We start the analysis with a lemma that states some properties of the M_n's given by (3.19) (see also (3.20)). Then, as an example, we consider the $M/M/1$ PS queue and show how these properties can be exploited to derive from (3.18) the LST of the sojourn time distribution. Next, the general case is treated.

To obtain closed expressions for the M_n's determined by (3.19) we introduce the following assumption (cf. assumption 5.1):

ASSUMPTION 6.1

The zeros $1/y_1, \dots, 1/y_L$ of the denominator of $M(z)$ given by (3.20) are all distinct.

Under this assumption it is easily seen that we can write, cf. (5.8),

$$M_n = A_1 y_1^n + \cdots + A_L y_L^n, \quad n = 1, 2, \dots, \tag{6.5}$$

with $A_1, \dots, A_L$ determined by

$$\frac{A_1 y_1}{1 - y_1 z} + \cdots + \frac{A_L y_L}{1 - y_L z} = M(z). \tag{6.6}$$

Remark 6.1

$1/y_1, \dots, 1/y_L$ are the roots of a polynomial of degree $L \leqslant \sum_{j=1}^{m} r_j + 1$, see (3.20), (6.2); (6.6) leads to a set of L linear equations from which $A_1, \dots, A_L$ can be obtained (cf. remark 5.2).

Analogously to the proof of (ii) and (iii) of lemma 5.1 it can be shown that

LEMMA 6.1

(i) y_i can be written as

$$y_i = 1 - \beta d_i, \tag{6.7}$$

with d_i independent of β, $i = 1, \dots, L$;

(ii) A_i is independent of β, $i = 1, \dots, L$.

Note that, in fact, $d_i = (1 - y_i)/\beta$, $i = 1, \ldots, L$ are the roots of

$$\tilde{z} + \omega_0 + \lambda - \sum_{j=1}^{m} \alpha_j \left[\frac{1}{1 - \hat{\beta}_j \tilde{z}} \right]^{r_j} = 0. \tag{6.8}$$

The properties stated in lemma 6.1 will be used below. Before treating the general case we first give an example.

THE $M/M/1$ PS QUEUE

For exponential service times ($Q(z) = (1-p)z/(1-pz)$, with $p = 1 - \beta/\hat{\beta}$),

$$M(z) = z \frac{1 + \beta\omega_0 - \lambda\beta \frac{z}{1-z}(1 - (1-p)z/(1-pz))}{1 - z(1 + \beta\omega_0 + \lambda\beta) + \lambda\beta z(1-p)z/(1-pz)}. \tag{6.9}$$

It is easily seen that the zeros $1/y_1$ and $1/y_2$ of the denominator of (6.9) are given by

$$y_1 = 1 + \tfrac{1}{2}\beta\left[\omega_0 + \lambda - 1/\hat{\beta} + \sqrt{\left(\omega_0 + \lambda - 1/\hat{\beta}\right)^2 + 4\omega_0/\hat{\beta}}\right], \tag{6.10}$$

$$y_2 = 1 + \tfrac{1}{2}\beta\left[\omega_0 + \lambda - 1/\hat{\beta} - \sqrt{\left(\omega_0 + \lambda - 1/\hat{\beta}\right)^2 + 4\omega_0/\hat{\beta}}\right].$$

We can write (cf. (6.5))

$$M_n = A_1 y_1^n + A_2 y_2^n, \quad n = 1, 2, \ldots, \tag{6.11}$$

with

$$A_1 = \frac{y_2 - (1 + \beta\omega_0)}{y_2 - y_1}, \quad A_2 = \frac{y_1 - (1 + \beta\omega_0)}{y_1 - y_2}. \tag{6.12}$$

Now substitute (6.11) into (3.18) and evaluate the summations in the denominator (take $q(i) = p^{i-1}$). Taking in the resulting expressions $y_i = 1 - \beta d_i$, $i = 1, 2$, $p = 1 - \beta/\hat{\beta}$, $\beta = x/k$ and using that d_i is independent of β it is easily seen that

$$\lim_{k\to\infty} (1 + \beta\omega_0) M_{k-1} = \sum_{h=1}^{2} A_h \, e^{-x d_h},$$

$$\lim_{k\to\infty} \lambda\beta \sum_{j=1}^{k-2} q(k-j-1) M_j = \lambda \sum_{h=1}^{2} A_h \hat{\beta} \frac{1}{1 - \hat{\beta} d_h} \left[e^{-x d_h} - e^{-x/\hat{\beta}} \right],$$

$$\lim_{k\to\infty} \lambda\beta \sum_{i=1}^{k-2} q(i) = \lambda\hat{\beta}(1 - e^{-x/\hat{\beta}}).$$

Hence, cf. (4.4),

$$E\{e^{-\omega_0 S^{PS}(x)}\}$$
$$= \lim_{k\to\infty} E\{e^{-\omega_0 S^{(k)}}\}$$
$$= \frac{1-\rho}{\sum_{h=1}^{2} A_h e^{-xd_h} - \rho e^{-x/\hat{\beta}} - \lambda \sum_{h=1}^{2} A_h \hat{\beta} \frac{1}{1-\hat{\beta} d_h}\left[e^{-xd_h} - e^{-x/\hat{\beta}}\right]}. \quad (6.13)$$

It is easily shown that this result coincides with the result obtained in Coffman et al. [10] (formula (30) on p. 128). Note that formula (30) of that paper represents the LST of the distribution of the total *delay* of a customer with a specific service demand. To match it with our result it has to be multiplied by the LST of the required service time (given by $\exp(-\omega_0 x)$).

THE $M/G/1$ PS QUEUE

Now we shall treat the general case, i.e., the case that the service times are determined by (6.1). Consider in the corresponding feedback queue the total sojourn time after k services given by (3.18). As in the $M/M/1$ case, we evaluate the terms $(1+\beta\omega_0)M_{k-1}$, $\lambda\beta\Sigma_{j=1}^{k-2} q(k-j-1)M_j$ and $\lambda\beta\Sigma_{i=1}^{k-2} q(i)$ in the denominator and take the limit $k\to\infty$ independently for each term. The first term is simple: from (6.5) and (6.7) it is easily seen that

$$\lim_{k\to\infty}(1+\beta\omega_0)M_{k-1} = \lim_{k\to\infty}\left(1+\frac{x}{k}\omega_0\right)\sum_{h=1}^{L} A_h\left(1-\frac{x}{k}d_h\right)^{k-1} = \sum_{h=1}^{L} A_h e^{-xd_h}. \quad (6.14)$$

The second one needs more effort. Using (6.3)–(6.5) and (6.7) it is found after extensive calculations that

$$\lambda\beta\sum_{j=1}^{k-2} q(k-j-1)M_j$$
$$= \lambda\beta\sum_{j=1}^{k-2} M_j \sum_{n=1}^{m}\alpha_n \sum_{i=0}^{r_n-1}\binom{k-j-2}{r_n-1-i}(1-p_n)^{r_n-1-i} p_n^{k-j-2-(r_n-1-i)}$$
$$= \lambda\beta\sum_{h=1}^{L} A_h \sum_{n=1}^{m}\alpha_n \sum_{i=0}^{r_n-1}\sum_{j=1}^{k-2}\binom{k-j-2}{r_n-1-i}(1-p_n)^{r_n-1-i} p_n^{k-j-2-(r_n-1-i)} y_h^j$$
$$= \lambda\sum_{h=1}^{L} A_h \sum_{n=1}^{m}\alpha_n \sum_{i=0}^{r_n-1}\frac{\left(1-x/\left(k\beta_n\right)\right)^{k-2-(r_n-1-i)}}{(r_n-1-i)!}$$

$$\times \sum_{j=1}^{k-r_n-1+i} \frac{x}{k}(k-j-2)\cdots(k-j-2-(r_n-2-i))$$

$$\times\left(x/\left(k\hat{\beta}_n\right)\right)^{r_n-1-i}\left(\frac{1-xd_h/k}{1-x/\left(k\hat{\beta}_n\right)}\right)^j.$$

The last equality is obtained by substituting $p_n = 1-\beta/\hat{\beta}_n$, $\beta = x/k$ and noting that $\binom{k-j-2}{r_n-1-i} = 0$ if $k-j-2 < r_n-1-i$. Using that, actually by definition,

$$\lim_{k\to\infty} \sum_{j=1}^{k-r_n-1+i} \frac{x}{k}(k-j-2)(k-j-3)\cdots(k-j-2-(r_n-2-i)) \times \tag{6.15}$$

$$\left(x/\left(k\hat{\beta}_n\right)\right)^{r_n-1-i}\left(\frac{1-xd_h/k}{1-x/\left(k\hat{\beta}_n\right)}\right)^j = \int_{s=0}^{x}\left(\frac{x}{\hat{\beta}_n}-\frac{s}{\hat{\beta}_n}\right)^{r_n-1-i} \mathrm{e}^{-s(d_h-1/\hat{\beta}_n)}\,\mathrm{d}s,$$

we obtain

$$\lim_{k\to\infty} \lambda\beta \sum_{j=1}^{k-2} q(k-j-1)M_j$$

$$= \lambda \sum_{h=1}^{L} A_h \sum_{n=1}^{m} \alpha_n \sum_{i=0}^{r_n-1} \frac{\mathrm{e}^{-x/\hat{\beta}_n}}{(r_n-1-i)!} \int_{s=0}^{x}\left(\frac{x}{\hat{\beta}_n}-\frac{s}{\hat{\beta}_n}\right)^{r_n-1-i} \mathrm{e}^{-s(d_h-1/\hat{\beta}_n)}\,\mathrm{d}s. \tag{6.16}$$

The evaluation of the third term is analogous to that of the second term:

$$\lambda\beta \sum_{i=1}^{k-2} q(i) = \lambda\beta \sum_{i=1}^{k-2} \sum_{n=1}^{m} \alpha_n \sum_{j=0}^{r_n-1} \binom{i-1}{r_n-1-j}(1-p_n)^{r_n-1-j} p_n^{i-1-(r_n-1-j)}$$

$$= \lambda \sum_{n=1}^{m} \alpha_n \sum_{j=0}^{r_n-1} \frac{\left(1-x/\left(k\hat{\beta}_n\right)\right)^{-(r_n-j)}}{(r_n-1-j)!}$$

$$\times \sum_{i=r_n-j}^{k-2} \frac{x}{k}(i-1)\cdots(i-1-(r_n-2-j))$$

$$\times\left(x/\left(k\hat{\beta}_n\right)\right)^{r_n-1-j}\left(1-x/\left(k\hat{\beta}_n\right)\right)^i.$$

Hence, cf. (6.15),

$$\lim_{k\to\infty} \lambda\beta \sum_{i=1}^{k-2} q(i) = \lambda \sum_{n=1}^{m} \alpha_n \sum_{j=0}^{r_n-1} \frac{1}{(r_n-1-j)!} \int_{s=0}^{x}\left(\frac{s}{\hat{\beta}_n}\right)^{r_n-1-j} \mathrm{e}^{-s/\hat{\beta}_n}\,\mathrm{d}s. \tag{6.17}$$

In the derivation of (6.16) and (6.17) one recognizes the convergence of the binomial distribution to the Poisson distribution, cf. Feller [16, ch. 6]. The integrals in (6.16) and (6.17) can be evaluated by noting that

$$\int_{s=0}^{x} s^n \, \mathrm{e}^{-s/c} \, \mathrm{d}s = n!c^{n+1}(1-\mathrm{e}^{-x/c}) - \mathrm{e}^{-x/c} \sum_{j=0}^{n-1} \frac{n!}{(n-j)!} x^{n-j} c^{j+1}.$$

Using the resulting expressions and (6.14) we obtain from (3.18):

THEOREM 6.1

In the $M/G/1$ PS queue with service time LST given by (6.1), for Re $\omega_0 \geqslant 0$,

$$\begin{aligned} E\{\mathrm{e}^{-\omega_0 S^{PS}(x)}\} &= \lim_{k\to\infty} E\{\mathrm{e}^{-\omega_0 S^{(k)}}\} \\ &= (1-\rho)\Bigg[\sum_{h=1}^{L} A_h \, \mathrm{e}^{-xd_h} - \lambda \sum_{n=1}^{m} \alpha_n \hat{\beta}_n \, \mathrm{e}^{-x/\hat{\beta}_n} \sum_{j=0}^{r_n-1} \sum_{i=0}^{j} \frac{\left(x/\hat{\beta}_n\right)^i}{i!} \\ &\quad - \lambda \sum_{h=1}^{L} A_h \sum_{n=1}^{m} \alpha_n \hat{\beta}_n \left(\frac{1}{1-\hat{\beta}_n d_h}\right)^{r_n} \sum_{j=0}^{r_n-1} \left(1-\hat{\beta}_n d_h\right)^j \\ &\quad \times \left(\mathrm{e}^{-xd_h} - \mathrm{e}^{-x/\hat{\beta}_n} \sum_{i=0}^{r_n-1-j} \frac{\left(x(1-\hat{\beta}_n d_h)/\hat{\beta}_n\right)^i}{i!}\right)\Bigg]^{-1}, \end{aligned} \tag{6.18}$$

with $d_1, \ldots, d_L$ the roots of (6.8) and $A_1, \ldots, A_L$ determined by (6.6), cf. remark 6.1; d_h and A_h are independent of x, $h = 1, \ldots, L$.

For hyperexponentially (H_m) distributed service times ($r_j = 1$, $j = 1, \ldots, m$, cf. (6.1)) (6.18) reduces to a much simpler expression. It is easily verified that for $m = 1$, $r_1 = 1$, the $M/M/1$ case, (6.18) reduces to (6.13).

The deterministic distribution is not contained in the class of service time distributions determined by (6.1), so the above analysis does not apply to the $M/D/1$ PS queue. Deterministic service times can be approximated by an Erlang-n distribution (for large n) but this leads to the problem of finding the roots of an $(n+1)$th degree polynomial and the solution of a set of $n+1$ linear equations (cf. (6.5), (6.6)). However, another approach is possible [6]: explicit formulas for the sojourn time in the $M/D/1$ PS queue can be easily obtained from the sojourn time in the $M/M/1$ queue with deterministic feedback in which each customer receives exactly N services. Taking $N = \lceil \hat{\beta}/\beta \rceil$ and $\beta = x/k$ it is clear that the total sojourn time after k services in the feedback queue approaches, for $k \to \infty$, the sojourn time of a (special) customer with service demand x in the $M/D/1$ PS queue with service time $\hat{\beta}$.

Remark 6.2

Our sojourn time results in theorems 5.1 and 6.1 are given in terms of the roots of a polynomial and the solution of a set of linear equations. The corresponding formulas collected in Yashkov [32] are given in terms of multiple integrals. In general both types of formulas can only be evaluated numerically. For obtaining numerical results it seems in our case to be more convenient to use the feedback results (3.18) and (3.22) with $\beta = x/k$ and k sufficiently large. Preliminary tests suggest that this procedure works quite well even for reasonably small k. Thus the feedback queue might in a natural way lead to sojourn time approximations for the $M/G/1$ PS queue. This is a promising topic for further research.

Remark 6.3

From corollary 3.2 and application of the limiting procedure, see section 4, it follows that for the $M/G/1$ PS queue the random state of the system (the number of customers present *and* their residual service requests) just after the departure of a tagged customer who has received an amount $x \geqslant 0$ of service is described by the stationary distribution of the state of the system at an arbitrary epoch, *independent* of x. This result slightly extends theorem 2.3 of Ott [23]. Ott's theorem concerns only the distribution of the number of customers at a departure epoch of a tagged customer with initial service demand x.

7. The $M/M/1$ feedback queue with additional permanent customers

In this section we consider the same $M/M/1$ queue with general feedback as in section 3 but with $K \geqslant 1$ additional permanent customers. This model is pictured in fig. 2. A model of a single server queue with an additional class of permanent customers exposes a structure that appears in many representations of computer and communication networks. The prime interest of such models is to determine and understand the influence of one class of customers on another one. This is useful for perceiving the operation of more complex queueing

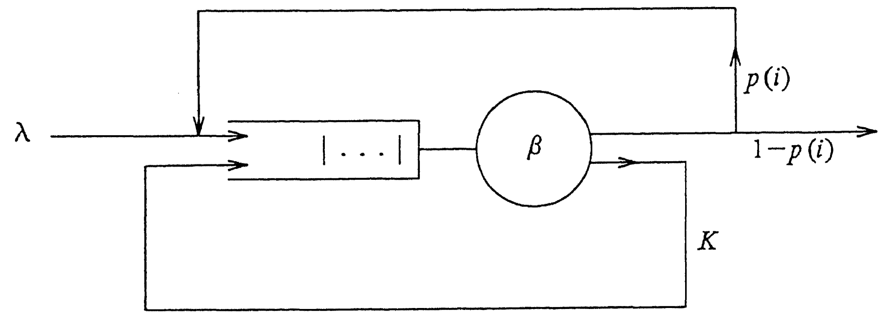

Fig. 2. The $M/M/1$ feedback queue with K additional permanent customers.

systems. See [7] for a study of an $M/G/1$ queue with additional customers whose service time distributions may differ from those of the Poisson customers; the latter model generalizes certain vacation queues.

In the present study it is assumed that the service times of the Poisson customers and the permanent customers are independent, negative exponentially distributed random variables, all with mean β. For the Poisson customers the assumptions about the feedback mechanism, notations, terminology, etc. are the same as for the model without permanent customers, see section 2. Our main goal is to study the influence of the presence of the permanent customers on the joint distribution of the successive sojourn times of a tagged Poisson customer and to use the results for the analysis of the sojourn time in the $M/G/1$ PS queue with additional permanent customers (section 8). The results for this latter model are obtained by applying the same limiting procedure as used for the case without permanent customers, see section 3.

Because the Poisson customers and the permanent customers have the *same* exponential service time distribution, the joint stationary distribution of the number of type-i (Poisson) customers, X_i, $i = 1, 2, \dots$, in the system at an arbitrary epoch is of product-form type. From the queue length results for general product-form networks (see Baskett et al. [1]) it is found that for our model, cf. (2.3),

$$Pr\{X_1 = x_1, X_2 = x_2, \dots\} = (1-\rho)^{K+1} \frac{(K + x_1 + x_2 + \cdots)!}{K!}$$

$$\prod_{i=1}^{\infty} \frac{(\lambda\beta q(i))^{x_i}}{x_i!}, \quad x_1, x_2, \dots = 0, 1, \dots, \quad x_1 + x_2 + \cdots < \infty. \tag{7.1}$$

(Remember that $q(i)$ represents the relative arrival rate of type-i (Poisson) customers, $i = 1, 2, \dots$, (cf. (2.1)), and that ρ denotes the total offered load to the system per unit of time due to the Poisson customers: $\rho = \lambda\beta\sum_{i=1}^{\infty} q(i)$.)

The generating function of the joint queue length distribution is given by (cf. the derivation of (2.4)):

$$E\left\{\prod_{i=1}^{\infty} z_i^{X_i}\right\} = (1-\rho)^{K+1} \frac{1}{K!} \sum_{m=0}^{\infty} \sum_{x_1} \sum_{x_2} \cdots_{x_1 + x_2 + \cdots = m} (m+K)! \prod_{i=1}^{\infty} \frac{(\lambda\beta q(i) z_i)^{x_i}}{x_i!}$$

$$= (1-\rho)^{K+1} \sum_{m=0}^{\infty} \binom{m+K}{K} \left(\sum_{i=1}^{\infty} \lambda\beta q(i) z_i\right)^m$$

$$= \left(\frac{1-\rho}{1 - \sum_{i=1}^{\infty} z_i \lambda\beta q(i)}\right)^{K+1}, \quad |z_i| \leqslant 1, \quad i = 1, 2, \dots. \tag{7.2}$$

Comparing this result with (2.4) we observe the following phenomenon: *the presence of the K permanent customers in the M/M/1 feedback queue leads to a joint queue length distribution which is the $(K+1)$-fold convolution of the joint queue length distribution in the same model without permanent customers.* In this section we shall use (7.2) for the analysis of the sojourn time distribution.

SOJOURN TIME DISTRIBUTION

We present, in the form of Laplace–Stieltjes transforms and generating functions, an expression for the joint steady state distribution of the successive sojourn times S_j, $j=1,\dots,k$, and the number of type-i customers, $X_i^{(j)}$, $i=1,2,\dots,$ present at the jth service completion of a customer who is fed back at least $k-1$ times, $k=1,2\dots.$ It will appear that for the derivation of this quantity we can largely rely on the analysis of the sojourn time in the model without permanent customers given in section 3.

Consider a newly arriving (tagged) customer, say C, and suppose that he finds $X_i^{(0)}=x_i$ type-i (Poisson) customers in the system, $i=1,2,\dots,$ together with the K permanent customers. It is easily seen that the determination of the (conditional) joint sojourn time distribution of C can be performed in almost exactly the same way as for the original $M/M/1$ feedback queue without permanent customers leading to theorem 3.1, the only difference being that for the present model one has to take into account that after *each* of his services C finds K additional permanent customers in the queue (besides the different types of Poisson customers). Realizing this it can be shown in a straightforward manner that, for Re $\omega_j \geqslant 0$, $|z_{i,j}| \leqslant 1$, $i=1,2,\dots,j=0,\dots,k$,

$$E\left\{e^{-(\omega_1 S_1+\dots+\omega_k S_k)}\left(\prod_{i=1}^{\infty} z_{i,0}^{X_i^{(0)}}\cdots\prod_{i=1}^{\infty} z_{i,k}^{X_i^{(k)}}\right) \mid X_1^{(0)}=x_1,\ X_2^{(0)}=x_2,\dots\right\}$$

$$=\left[\prod_{j=1}^{k} A_k(j,\omega,z)\right]^{K+1}\prod_{i=1}^{\infty}\left(z_{i,0}f_k(i,\omega,z)\right)^{x_i}, \tag{7.3}$$

with $\omega := (\omega_1,\dots,\omega_k)$, $z := ((z_{1,0}, z_{2,0},\dots),\dots,(z_{1,k}, z_{2,k},\dots))$, and with $A_k(\cdot,\cdot,\cdot)$ and $f_k(\cdot,\cdot,\cdot)$ defined by (3.3) and (3.4). Note that the $(K+1)$st power term in (7.3) is the contribution induced by the tagged customer and the K permanent customers, cf. remark 3.1.

Using the PASTA property and deconditioning we obtain from (7.2) and (7.3) our main result:

THEOREM 7.1

The joint distribution of the successive sojourn times and the number of Poisson customers of each type present in the system at the service completion epochs of a tagged Poisson customer is the $(K+1)$-fold convolution of the corresponding joint distribution in the model without permanent customers, cf. (3.5):

$$E\left\{e^{-(\omega_1 S_1+\ldots+\omega_k S_k)}\left(\prod_{i=1}^{\infty} z_{i,0}^{X_i^{(0)}} \cdots \prod_{i=1}^{\infty} z_{i,k}^{X_i^{(k)}}\right)\right\}$$
$$=\left[\frac{(1-\rho)\prod_{j=1}^{k} A_k(j,\omega,z)}{1-\lambda\beta\sum_{i=1}^{\infty} q(i) z_{i,0} f_k(i,\omega,z)}\right]^{K+1},$$

$$\operatorname{Re}\omega_j \geqslant 0, \quad |z_{i,j}| \leqslant 1, \quad i=1,2,\ldots, \quad j=0,\ldots,k. \tag{7.4}$$

Using theorem 7.1 most of the sojourn time characteristics can be immediately obtained from the results given in section 3. Here we shall restrict ourselves to a summary of the most important characteristics.

- The jth sojourn time S_j of a Poisson customer has a $(K+1)$-stage Erlang distribution (E_{K+1}) with mean $(K+1)\beta/(1-\rho)$ (cf. (3.9)):

$$E\{e^{-\omega_j S_j}\}=\left(\frac{1-\rho}{1-\rho+\beta\omega_j}\right)^{K+1}, \quad j=1,\ldots,k. \tag{7.5}$$

- The correlation coefficient, $corr(S_i, S_j)$, of the ith and the jth sojourn time of a Poisson customer is independent of the number of permanent customers in the system (cf. (3.17)):

$$corr(S_i, S_j)=1-C_{j-i}(1-\rho), \quad 1 \leqslant i<j \leqslant k, \tag{7.6}$$

  with C_n, $n=1,\ldots,k-1$, determined by (3.12).
- The variance of the total sojourn time after k services,. $var(S^{(k)})$, is given by (cf. (3.22)):

$$var(S^{(k)})=(K+1)\left(\frac{\beta}{1-\rho}\right)^2\left[k^2-2(1-\rho)\sum_{j=1}^{k-1} C_{k-j}\right], \quad k=1,2,\ldots. \tag{7.7}$$

Remark 7.1

Noting that in the present product-form model a departing (and hence arriving) permanent customer sees the system in equilibrium with one less customer of his own type (see e.g. Walrand [30, section 4.4]) the characteristics of the successive cycle times of a particular permanent customer can be immediately obtained from the above sojourn time results for the Poisson customers. For example, the cycle times have a K-stage Erlang distribution (E_K) with mean $K\beta/(1-\rho)$, cf. (7.5).

8. The $M/G/1$ PS queue with additional permanent customers

In sections 4–6 it has been shown how queue length and sojourn time results for the $M/G/1$ processor sharing queue can be obtained from queue length and sojourn time results for the $M/M/1$ queue with general feedback. We have applied a limiting procedure in which the mean service time $\beta \to 0$ while the feedback probabilities approach one in such a way that a customer's total required service time remains constant, see section 4. It is easily seen that application of the same limiting procedure to the present $M/M/1$ feedback model with K permanent customers leads to the $M/G/1$ PS queue with K permanent customers. Note that the behaviour of the latter model is independent of the service time distribution(s) of the permanent customers (the permanent customers are *always* in service). From (7.2) it follows immediately that *for the $M/G/1$ PS queue with K permanent customers, the distribution of the queue length X^{PS} at an arbitrary epoch is the $(K+1)$-fold convolution of the queue length distribution in the same model without permanent customers* (cf. (4.1)):

$$E\{z^{X^{PS}}\} = \left(\frac{1-\rho}{1-\rho z}\right)^{K+1}, \quad |z| \leqslant 1, \tag{8.1}$$

i.e.

$$Pr\{X^{PS} = n\} = (1-\rho)^{K+1}\binom{n+K}{K}\rho^n, \quad n = 0, 1, \ldots, \tag{8.2}$$

with ρ the offered load to the system per unit of time due to the Poisson customers. From theorem 7.1 we obtain the following remarkable sojourn time result:

THEOREM 8.1

For the $M/G/1$ PS queue with K permanent customers the distribution of the conditional sojourn time $S^{PS}(x)$ of a Poisson customer with given service demand x is the $(K+1)$-fold convolution of the distribution of the conditional sojourn time in the same model without permanent customers. This also holds for the unconditional sojourn time S^{PS} of an arbitrary Poisson customer.

Theorem 8.1 implies (cf. (4.9)):

$$E\{S^{PS}(x)\} = (K+1)\frac{x}{1-\rho}, \quad x \geqslant 0. \tag{8.3}$$

Remark 8.1

For the present PS model it is interesting to study the influence of the presence of the Poisson customers on the "speed" with which the permanent customers are served. For $x \geqslant 0$ let $C^{PS}(x)$ be the time required to give the permanent customers an amount x of service. From the discussion in remark 7.1 and application of the limiting procedure it follows that $C^{PS}(x)$ is distributed as the conditional sojourn time of a tagged Poisson customer with service demand x in the same model but with one less permanent customer. For example, from (8.3),

$$E\{C^{PS}(x)\} = K\frac{x}{1-\rho}, \quad x \geqslant 0. \tag{8.4}$$

This formula shows that the influence of the Poisson customer stream on $E\{C^{PS}(x)\}$ is simply a reduction of the capacity of the server by an amount ρ, the load offered by the Poisson customers. Moreover, (8.4) implies that the mean *total* amount of service obtained by the permanent customers per unit of time (given by $Kx/E\{C^{PS}(x)\}$) is independent of K.

Remark 8.2

In remark 6.3 we concluded that for the $M/G/1$ PS queue (without permanent customers) the queue length distribution just after the departure of a tagged customer who has received an amount x of service is the same as at an arbitrary epoch, *independent* of x. From (8.1) it follows that for the $M/G/1$ PS queue with one permanent customer the queue length distribution at an arbitrary epoch is the two-fold convolution of the queue length distribution in the PS queue without permanent customers. Since one would expect that, when the required service time x of a tagged customer becomes very large, the behaviour of the $M/G/1$ PS queue approaches that of the corresponding PS queue with one permanent customer, it seems paradoxical that both statements are true. However, viewing the $M/G/1$ PS queue as the limiting case of the $M/M/1$ queue with general feedback this is immediately clear (a departure in the PS model corresponds with a certain service completion in the feedback model which is more likely to occur when there are fewer customers in the system). A similar "paradox" for queue lengths in PS queues is discussed in Foley and Klutke [17].

Remark 8.3

Cohen [11] has studied *generalized processor sharing* (GPS), which is the following generalization of the PS service discipline: When j customers are present in the system, then the service rate for each of them is $f(j) > 0$. The $M/G/1$ PS queue with K permanent customers can be viewed as a special case of generalized processor sharing, with $f(j) = 1/(j+K)$, $j = 1, 2, \ldots$. Formulas (8.2) and (8.3) have thus already been obtained by Cohen [11]; theorem 8.1 is a new result. Another approach to the $M/G/1$ GPS queue is to start with an $M/M/1$ feedback queue in which the feedback probabilities are chosen as in section 6 to obtain the service request distribution of the $M/G/1$ GPS queue, but with state-dependent service rates $\mu(j) = jf(j)$ when j customers are present in the feedback queue. For this feedback queue the Markov renewal approach of section 3 no longer works. However, the feedback queue still has queue-length product form and taking the limits in the usual way to arrive at processor sharing leads to (8.2) and (8.3).

The above results for the queue length and the sojourn time in the $M/G/1$ PS queue with permanent customers are interesting both from a theoretical and a practical point of view. One example where this queueing model may arise is provided by a "Stored Program Controlled" (SPC) telephone exchange that is offered two types of jobs: (i) call requests, and (ii) operator tasks (see De Waal [29]). To guarantee a certain quality of service of the call requests only a limited number (K) of operator tasks is allowed to be in service at the same time. It is clear that under heavy traffic conditions of the operator tasks and for appropriate assumptions about the system parameters the above formulas (8.1)–(8.3) (approximately) reflect the influence of the choice of the control parameter K on the queue length and the delay of the call requests. From the discussion in remark 8.1 it follows that under certain conditions the *maximum* throughput of the operator tasks is independent of K. So, if the objective is to minimize the delay of the call requests and to maximize the throughput of the operator tasks one should take K as small as possible, i.e., $K = 1$.

9. The $M/G/1$ processor sharing queue with feedback

In this section we consider an $M/G/1$ PS queue with feedback. The feedback mechanism has the same structure as described in section 2 for the $M/M/1$ FCFS queue, i.e., the probability that a customer is fed back after completing his service may depend on the number of times he has already been served. We shall study the successive sojourn times of a tagged customer. In particular we are interested in dependencies between these sojourn times.

The PS queue with feedback has been studied before by Klutke et al. [21]. They consider the special case of Bernoulli feedback and analyze the behaviour of the internal input and output processes. In particular they study the influence of the shape of the service time distribution on the interoutput time distribution. Their main result is that when service time distributions with the same mean are convexly ordered, so are interoutput time distributions. The purpose of their study is to gain insight into the properties of traffic processes in general queueing networks with processor sharing nodes.

In Klutke et al. [21] it is remarked that the study of flow processes is crucial for understanding the behaviour of more complicated processes in the system.

As an example the authors mention the sojourn time process and say that "*this is still an open problem*". In this section we shall show that sojourn times in the $M/G/1$ PS queue with feedback can be obtained from the sojourn time results for the $M/M/1$ FCFS feedback queue derived in section 3.

MODEL DESCRIPTION AND NOTATIONS

We consider an $M/G/1$ PS queue with feedback (PSFB). When a customer has completed his ith service, he departs from the system with probability $1-\tilde{p}(i)$ and is fed back with probability $\tilde{p}(i)$, $i=1, 2, \dots$. Fed back customers return instantaneously, and due to the PS service discipline a returning customer is immediately taken into service again. The successive service requests $\tilde{\tau}_1, \tilde{\tau}_2, \dots$ of a customer are independent random variables with distribution function $\tilde{B}_1(\cdot), \tilde{B}_2(\cdot), \dots$ and means $\tilde{\beta}_1, \tilde{\beta}_2, \dots$ respectively. New customers arrive according to a Poisson process with intensity λ. Obviously, for stability it is required that the load $\rho = \lambda \sum_{j=1}^{\infty}((1-\tilde{p}(j))\prod_{i=1}^{j-1}\tilde{p}(i))(\tilde{\beta}_1 + \cdots + \tilde{\beta}_j) < 1$.

We are interested in the successive sojourn times $\tilde{S}_1(T_1), \dots, \tilde{S}_N(T_N)$ of a (tagged) customer in the PSFB queue who requires at least $N \geq 1$ services of length $T_1, \dots, T_N \geq 0$ respectively. In particular we shall derive an expression for the correlation coefficient, $corr(\tilde{S}_i(T_i), \tilde{S}_j(T_j))$, of the ith and the jth sojourn time of a tagged customer, $i, j = 1, \dots, N$.

For the analysis of the successive sojourn times in the PSFB queue we shall consider corresponding sojourn times in an associated processor sharing queue *without* feedback. Let $\hat{B}(\cdot)$ denote the distribution function of the total required service time, i.e.,

$$\hat{B}(t) := \sum_{j=1}^{\infty}\left((1-\tilde{p}(j))\prod_{i=1}^{j-1}\tilde{p}(i)\right)\left(\tilde{B}_1(t) * \cdots * \tilde{B}_j(t)\right), \quad t \geq 0. \tag{9.1}$$

It is easily seen that the behaviour of the $M/G/1$ PS queue with service time distribution $\hat{B}(\cdot)$ is exactly the same as the behaviour of the PSFB queue described above. In the sequel the PS queue with service time distribution $\hat{B}(\cdot)$ will be called "the associated PS queue" (or shortly "the PS queue"). For a tagged customer with initial service demand $\tau^{PS} \geq T_1 + \cdots + T_N$, $T_1, \dots, T_N \geq 0$, in the associated PS queue we define:

$S_i^{PS}(T_i)$: time during which the remaining service demand of the tagged customer is in the range

$$\left(\tau^{PS} - \sum_{j=1}^{i} T_j, \tau^{PS} - \sum_{j=1}^{i-1} T_j\right], \quad i = 1, \dots, N.$$

Obviously, the joint distribution of $S_1^{PS}(T_1), \dots, S_i^{PS}(T_i)$ does not depend on $T_{i+1}, \dots, T_N$, $i = 1, \dots, N-1$; $S_1^{PS}(T_1)$ is distributed as the conditional sojourn time of a tagged customer with service demand T_1:

$$E\{e^{-\omega_0 S_1^{PS}(T_1)}\} = E\{e^{-\omega_0 S^{PS}(T_1)}\}, \quad T_1 \geq 0, \quad \text{Re}\ \omega_0 \geq 0. \tag{9.2}$$

It is clear that the quantities $S_i^{PS}(T_i)$, $i = 1, \ldots, N$ in the associated PS queue have the same joint distribution as the successive sojourn times $\tilde{S}_1(T_1), \ldots, \tilde{S}_N(T_N)$ in the PSFB queue. In particular,

$$corr\left(\tilde{S}_i(T_i),\, \tilde{S}_j(T_j)\right) = corr\left(S_i^{PS}(T_i),\, S_j^{PS}(T_j)\right), \quad i, j = 1, \ldots, N. \tag{9.3}$$

So below we shall focus on the sojourn times $S_i^{PS}(T_i)$, $T_i \geqslant 0$, $i = 1, \ldots, N$, in the PS queue.

ANALYSIS

Consider the $M/G/1$ PS queue with service time distribution $\hat{B}(\cdot)$. We assume that $\hat{B}(\cdot)$ belongs to the class of distributions given by (4.2). The first moment of $\hat{B}(\cdot)$ is denoted by $\hat{\beta}$. From remark 6.3 it follows immediately that for $2 \leqslant i \leqslant N$ the joint distribution of $S_i^{PS}(T_i), \ldots, S_N^{PS}(T_N)$ does not depend on $T_1, \ldots, T_{i-1}$. This implies that, cf. (9.2),

$$E\{e^{-\omega_0 S_i^{PS}(T_i)}\} = E\{e^{-\omega_0 S^{PS}(T_i)}\}, \quad T_i \geqslant 0, \quad i = 1, \ldots, N, \quad \text{Re } \omega_0 \geqslant 0. \tag{9.4}$$

So, means are simply given by, see (4.9),

$$E\{S_i^{PS}(T_i)\} = \frac{T_i}{1-\rho}, \quad i = 1, \ldots, N, \tag{9.5}$$

with offered load $\rho = \lambda\hat{\beta}$. It also follows that $corr(S_i^{PS}(T_i), S_j^{PS}(T_j))$, $T_1, \ldots, T_N \geqslant 0$ depends only on T_i, T_j and $\sum_{n=i+1}^{j-1} T_n$, $1 \leqslant i < j \leqslant N$. Hence, for the analysis of $corr(S_i^{PS}(T_i), S_j^{PS}(T_j))$, $T_1, \ldots, T_N \geqslant 0$, $i,j = 1, \ldots, N$ we can restrict ourselves to the determination of $corr(S_1^{PS}(T_1), S_3^{PS}(T_3))$, T_1, T_2, $T_3 \geqslant 0$, without loss of generality. Below we shall derive an expression for the latter correlation. We shall consider corresponding sojourn times in the $M/M/1$ FCFS feedback queue and apply the limiting procedure described in section 4. The analysis is largely analogous to the derivation of the sojourn time variance in the $M/G/1$ PS queue, see theorem 5.1.

Consider the $M/M/1$ FCFS feedback queue with mean service time β and feedback probabilities $p(i)$, $i = 1, 2, \ldots$ related with β such that the total required service time has distribution function $\hat{B}(\cdot)$, see (4.4)–(4.6). We follow a tagged customer during his first $k_1 + k_2 + k_3$ successive sojourn times $S_1, \ldots, S_{k_1+k_2+k_3}$. Define

$$S_1(k_1) := S_1 + \cdots + S_{k_1},$$

$$S_2(k_2) := S_{k_1+1} + \cdots + S_{k_1+k_2},$$

$$S_3(k_3) := S_{k_1+k_2+1} + \cdots + S_{k_1+k_2+k_3}.$$

Clearly, when we take $k_2 = \lceil T_2/\beta \rceil$, $k_3 = \lceil T_3/\beta \rceil$, $\beta = T_1/k_1$ and let $k_1 \to \infty$ then $S_1(k_1)$, $S_2(k_2)$ and $S_3(k_3)$ correspond to the PS quantities $S_1^{PS}(T_1)$, $S_2^{PS}(T_2)$ and $S_3^{PS}(T_3)$ respectively (cf. section 4; note that, for $k_1 \to \infty$, $k_i\beta \to T_i$, $i = 1, 2, 3$).

We shall first derive $corr(S_1(k_1), S_3(k_3))$ for general $k_1, k_2, k_3 \geqslant 0$. Next, taking k_2, k_3 and β as indicated above we use

$$corr\left(S_1^{PS}(T_1), S_3^{PS}(T_3)\right) = \lim_{k_1 \to \infty} corr\left(S_1(k_1), S_3(k_3)\right). \tag{9.6}$$

From the definition of $S_i(k_i)$, $i = 1, 2, 3$, it follows that the covariance of $S_1(k_1)$ and $S_3(k_3)$ can be written as

$$cov\left(S_1(k_1), S_3(k_3)\right) = \sum_{i=1}^{k_1} \sum_{j=1}^{k_3} cov\left(S_i, S_{k_1+k_2+j}\right), \quad k_1, k_2, k_3 \geqslant 0. \tag{9.7}$$

Formula (3.17) expresses the covariance of S_i and S_m in C_{m-i}, cf. (3.12). Writing C_{m-i} as in (5.8) it follows that, for $k_1, k_2, k_3 \geqslant 0$,

$$\begin{aligned} &cov\left(S_1(k_1), S_3(k_3)\right) \\ &= \left(\frac{\beta}{1-\rho}\right)^2 \sum_{i=1}^{k_1} \sum_{j=1}^{k_3} \left(1 - (1-\rho) C_{k_1+k_2+j-i}\right) \\ &= -\frac{\beta^2}{1-\rho} \sum_{l=2}^{L} \sum_{i=1}^{k_1} \sum_{j=1}^{k_3} U_l x_l^{k_1+k_2-1} x_l^{-i} x_l^{j} \\ &= -\frac{\beta^2}{1-\rho} \sum_{l=2}^{L} U_l x_l^{k_2} \left(\frac{x_l^{k_1} - x_l^{-1}}{x_l - 1} - 1\right)\left(\frac{1 - x_l^{k_3+1}}{1 - x_l} - 1\right). \end{aligned} \tag{9.8}$$

Replacing in (9.8) x_l by $1 - \beta a_l$, $l = 2, \ldots, L$, see (5.11), and taking appropriate limits, i.e. $k_2 = \lceil T_2/\beta \rceil$, $k_3 = \lceil T_3/\beta \rceil$, $\beta = T_1/k_1$ and $k_1 \to \infty$, we find, cf. (9.6):

$$\begin{aligned} &cov\left(S_1^{PS}(T_1), S_3^{PS}(T_3)\right) \\ &= -\frac{1}{1-\rho} \sum_{l=2}^{L} U_l (1/a_l)^2 \,\mathrm{e}^{-T_2 a_l}\left(1 - \mathrm{e}^{-T_1 a_l}\right)\left(1 - \mathrm{e}^{-T_3 a_l}\right), \quad T_1, T_2, T_3 \geqslant 0. \end{aligned} \tag{9.9}$$

Returning to the PS queue with feedback we have from (9.9) and (5.15):

THEOREM 9.1

For the successive sojourn times $\tilde{S}_1(T_1), \ldots, \tilde{S}_N(T_N)$, $T_1, \ldots, T_N \geqslant 0$, of a tagged customer in the $M/G/1$ PSFB queue with total service request LST given by (4.2),

$$\begin{aligned} &corr\left(\tilde{S}_i(T_i), \tilde{S}_j(T_j)\right) \\ &= \left\{ -\sum_{l=2}^{L} U_l (1/a_l)^2 \,\mathrm{e}^{-T_{i,j} a_l}\left(1 - \mathrm{e}^{-T_i a_l}\right)\left(1 - \mathrm{e}^{-T_j a_l}\right) \right\} \end{aligned}$$

$$\times\left\{2\left[\sum_{l=2}^{L}(1/a_l)^2U_l\left(1-T_ia_l-e^{-T_ia_l}\right)\right]^{1/2}\right.$$

$$\left.\times\left[\sum_{l=2}^{L}(1/a_l)^2U_l\left(1-T_ja_l-e^{-T_ja_l}\right)\right]^{1/2}\right\}^{-1},$$

$$1\leqslant i<j\leqslant N, \qquad (9.10)$$

with $T_{i,j}=\sum_{n=i+1}^{j-1}T_n$. $a_2,\dots,a_L$ are the roots of (5.12) and $U_2,\dots,U_L$ are determined by (5.10), cf. remark 5.2; a_l and U_l are independent of T_n, $n=1,\dots,N$, $l=2,\dots,L$.

It is interesting to consider some asymptotic properties of $corr(\tilde{S}_i(T_i),\ \tilde{S}_j(T_j))$. First, noting that in (9.10) Re $a_l>0$, $l=2,\dots,L$, see lemma 5.1, we obtain

$$corr\left(\tilde{S}_i(T_i),\ \tilde{S}_j(T_j)\right)\to 0,\quad T_i,\ T_j\geqslant 0,\ T_{i,j}\to\infty,\ 1\leqslant i<j\leqslant N, \qquad (9.11)$$

which is intuitively clear. Another asymptotic result applies to the case that T_i, T_j and $T_{i,j}$ become very small. Using $\sum_{l=2}^{L}U_l=1-1/(1-\rho)$ and (5.15) and the fact that (5.6) holds for general service time distribution (as observed below (5.16)), it follows from (9.10) that

$$corr\left(\tilde{S}_i(T_i),\ \tilde{S}_j(T_j)\right)\to 1,\quad T_i,\ T_j,\ T_{i,j}\to 0,\ 1\leqslant i<j\leqslant N. \qquad (9.12)$$

This result can be explained as follows. Suppose a tagged customer starts his ith service at time t. For T_i, T_j and $T_{i,j}$ close to zero it may be expected that the successive sojourn times $\tilde{S}_i(T_i),\dots,\tilde{S}_j(T_j)$ of the tagged customer are small (cf. (9.5)) and that no new arrivals or departures occur during the time interval $[t,\ t+\tilde{S}_i(T_i)+\dots+\tilde{S}_j(T_j)]$. Hence, due to the PS service discipline $\tilde{S}_j(T_j)=T_j\tilde{S}_i(T_i)/T_i$, i.e. $\tilde{S}_j(T_j)$ is completely determined by $\tilde{S}_i(T_i)$.

We conclude this section with an example.

THE $M/M/1$ PS QUEUE WITH BERNOULLI FEEDBACK

Consider the $M/M/1$ PS queue with Bernoulli feedback, i.e. $\tilde{B}(t)=1-e^{-t/\tilde{\beta}}$, $\tilde{p}(i)\equiv p$, $0\leqslant p<1$. For this case the total required service time is exponentially distributed with mean $\hat{\beta}=\tilde{\beta}/(1-p)$. From the calculations for the determination of the sojourn time variance in the $M/M/1$ PS queue, see section 5, we have in (9.10) $L=2$, $U_2=-1/(1-\rho)$, $a_2=(1-x_2)/\beta=(1-\rho)/\hat{\beta}$. Hence, for the $M/M/1$ queue with Bernoulli feedback (9.10) reduces to

$$corr\left(\tilde{S}_i(T_i),\ \tilde{S}_j(T_j)\right)$$

$$=\frac{e^{-T_{i,j}(1-\rho)/\hat{\beta}}\left(1-e^{-T_i(1-\rho)/\hat{\beta}}\right)\left(1-e^{-T_j(1-\rho)/\hat{\beta}}\right)}{2\left(e^{-T_i(1-\rho)/\hat{\beta}}-1+T_i(1-\rho)/\hat{\beta}\right)^{1/2}\left(e^{-T_j(1-\rho)/\hat{\beta}}-1+T_j(1-\rho)/\hat{\beta}\right)^{1/2}},$$

$$1\leqslant i<j\leqslant N. \qquad (9.13)$$

Acknowledgement

The authors are indebted to professor J.W. Cohen for several useful discussions.

References

[1] F. Baskett, K.M. Chandy, R.R. Muntz and F.G. Palacios, Open, closed, and mixed networks of queues with different classes of customers, J. ACM 22 (1975) 248–260.
[2] J.L. van den Berg, Simple approximations for second moment characteristics of the sojourn time in the $M/G/1$ processor sharing queue, in: *Messung, Modellierung und Bewertung von Rechensystemen und Netzen*, eds. G. Stiege and J.S. Lie (Springer, Berlin, 1989) pp. 105–120.
[3] J.L. van den Berg, Sojourn times in feedback and processor sharing queues, Ph.D. Thesis, Centre for Mathematics and Computer Science, Amsterdam (1990).
[4] J.L. van den Berg and O.J. Boxma, Sojourn times in feedback queues, in: *Operations Research Proceedings 1988*, eds. D. Pressmar et al. (Springer, Berlin, 1989) pp. 247–257.
[5] J.L. van den Berg and O.J. Boxma, Sojourn times in feedback and processor sharing queues, in: *Teletraffic Science for New Cost-Effective Systems, Networks and Services*, ed. M. Bonatti (North-Holland, Amsterdam, 1989) pp. 1467–1475.
[6] J.L. van den Berg, O.J. Boxma and W.P. Groenendijk, Sojourn times in the $M/G/1$ queue with deterministic feedback, Stochastic Models 5 (1989) 115–129.
[7] O.J. Boxma and J.W. Cohen, The $M/G/1$ queue with permanent customers, IEEE J. Sel. Areas Commun. SAC-9 (1991) 179–184.
[8] O.J. Boxma and H. Daduna, Sojourn times in queueing networks, in: *Stochastic Analysis of Computer and Communication Systems*, ed. H. Takagi (North-Holland, Amsterdam, 1990) pp. 401–450.
[9] E. Çinlar, *Introduction to Stochastic Processes* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
[10] E.G. Coffman, R.R. Muntz and H. Trotter, Waiting time distributions for processor-sharing systems, J. ACM 17 (1970) 123–130.
[11] J.W. Cohen, The multiple phase service network with generalized processor sharing, Acta Informatica 12 (1979) 245–284.
[12] R.L. Disney, A note on sojourn times in $M/G/1$ queues with instantaneous, Bernoulli feedback, Naval Res. Log. Quarterly 28 (1981) 679–684.
[13] R.L. Disney and P.C. Kiessler, *Traffic Processes in Queueing Networks – A Markov Renewal Approach* (The Johns Hopkins University Press, Baltimore, 1987).
[14] R.L. Disney and D. König, Queueing networks: a survey of their random processes, SIAM Rev. 27 (1985) 335–403.
[15] B.T. Doshi and J.S. Kaufman, Sojourn time in an $M/G/1$ queue with Bernoulli feedback, in: *Queueing Theory and its Applications – Liber Amicorum for J.W. Cohen*, eds. O.J. Boxma and R. Syski (North-Holland, Amsterdam, 1988) pp. 207–233.
[16] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1 (Wiley, New York, 1950).
[17] R.D. Foley and G.A. Klutke, Stationary increments in the accumulated work process in processor-sharing queues, J. Appl. Prob. 26 (1989) 671–677.
[18] J.J. Hunter, Sojourn time problems in feedback queues, Queueing Systems 5 (1989) 55–76.
[19] L. Kleinrock, *Queueing Systems*, Vol. 1 (Wiley, New York, 1975).
[20] L. Kleinrock, *Queueing Systems*, Vol. 2 (Wiley, New York, 1976).

[21] G.A. Klutke, P.C. Kiessler and R.L. Disney, Interoutput times in processor sharing queues with feedback, Queueing Systems 3 (1988) 363–376.
[22] S.S. Lam and A.U. Shankar, A derivation of response time distributions for a multi-class feedback queueing system, Perform. Eval. 1 (1981) 48–61.
[23] T.J. Ott, The sojourn-time distribution in the $M/G/1$ queue with processor sharing, J. Appl. Prob. 21 (1984) 360–378.
[24] J.A.C. Resing, G. Hooghiemstra and M.S. Keane, The $M/G/1$ processor sharing queue as the almost sure limit of feedback queues, J. Appl. Prob. 27 (1990) 913–918.
[25] R. Schassberger, A new approach to the $M/G/1$ processor-sharing queue, Adv. Appl. Prob. 16 (1984) 202–213.
[26] L. Takács, A single-server queue with feedback, Bell Syst. Tech. J. 42 (1963) 505–519.
[27] H.C. Tijms, *Stochastic Modeling and Analysis* (Wiley, New York, 1986).
[28] D. Voelker and G. Doetsch, *Die Zweidimensionale Laplace-Transformation* (Birkhäuser, Basel, 1950).
[29] P.R. de Waal, An approximation method for a processor sharing queue with controlled arrivals and a waitbuffer, in: *Proc. 28th IEEE Conf. on Decision and Control*, Tampa, FL (1989).
[30] J. Walrand, *An Introduction to Queueing Networks* (Prentice-Hall, Englewood Cliffs, NJ, 1988).
[31] S.F. Yashkov, A derivation of response time distribution for a $M/G/1$ processor sharing queue, Problems Contr. Info. Theory 12 (1983) 133–148.
[32] S.F. Yashkov, Processor-sharing queues: some progress in analysis, Queueing Systems 2 (1987) 1–17.