# Sojourn times in a multiclass processor sharing queue

*Please check the document version of this publication:*

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

A.P. Zwart

*Eindhoven University of Technology*
*Department of Mathematics and Computing Science*
*P.O. Box 513, 5600 MB Eindhoven*
*phone: +31 40 2475478*
*fax: +31 40 2465995*
*email: zwart@win.tue.nl*

ABSTRACT

We consider a processor sharing queue with several customer classes. For an arbitrary customer of class $i$ we show that the sojourn time distribution is regularly varying of index $-\nu_i$ iff the service time distribution is regularly varying of index $-\nu_i$, and derive an explicit asymptotic formula. Furthermore, the tail of the sojourn time distribution of customer class $i$ is shown to be unaffected by the tails of the service time distributions of other customer classes, even if some of the latter tails are heavier. This result implies that, when the sojourn time of a customer is large, this is not due to long service requirements of other customer types. In particular, short-range dependent traffic does not suffer from longe-range dependent traffic if processor sharing is used as a service discipline.

## 1. INTRODUCTION

In this study, we investigate the tail behaviour of the sojourn time distribution in the M/G/1 queue with the processor sharing (PS) service discipline. Under the PS regime, each customer is being served with rate $1/X$ when there are $X$ customers in the system. An extensive overview on processor sharing queues can be found in the surveys [28, 29].

In particular, we are interested in the case of multiple customer classes, each with its own service time distribution. We assume that at least one arrival stream $i$ has heavy-tailed service requirements, i.e. the service time distribution $B_i(.)$ of at least one arrival stream $i$ satisfies

$$1 - B_i(x) \sim h_{\nu_i} x^{-\nu_i}, \tag{1.1}$$

if $x \to \infty$ (with $f(x) \sim g(x)$ we mean $f(x)/g(x) \to 1$), and $1 < \nu_i < 2$. The reason for this is the following. Recent traffic measurements for traffic in Ethernet Local Area Networks [26], Wide Area Networks [22], and VBR video [4] show that these systems exhibit phenomena like self-similarity and long-range dependence – phenomena that can be explained by the occurrence of service requirements in a queue or fluid queue like in (1.1). Furthermore, these communication networks are often used by different types of customer classes having both short-range and long-range dependent characteristics.

We believe the contribution of the paper to be twofold. Firstly, we give new asymptotic results for the sojourn time distribution in the M/G/1 processor sharing queue. The main result of this

study is Theorem 3.1. It is shown that the sojourn time distribution of customer class $i$ is regularly varying of index $-\nu_i$ iff the service time distribution of class $i$ has the same property. A simple and explicit characterisation of the tail of the sojourn time distribution is given in terms of the tail of the service time distribution. Asymptotic upper bounds are provided for general service time distributions. A secondary purpose of this paper is to shed some light on the relationship between service disciplines and the tail behaviour of sojourn time distributions, and to show that short-range dependent traffic does not suffer under long-range dependent traffic under the processor sharing discipline. As an illustration, the service disciplines Fist-Come-First-Served (FCFS), Last-Come-First-Served Preemptive Resume (LCFS-PR) and Processor Sharing (PS) are compared. A related study on the interplay between service disciplines and tail behaviour is [1].

Theorem 3.1 is a generalisation of Theorem 4.1 in [31]. The latter states, for a single class of customers, that the sojourn time distribution of an arbitrary customer is regularly varying of index $-\nu$ ($\nu$ being non-integer) iff the service time distribution is regularly varying of index $-\nu$. This result exhibits a crucial property of the processor sharing service discipline. For the GI/G/1 FCFS queue, a result of Cohen [11] implies that the sojourn time distribution is regularly varying of index $1-\nu$ iff the service time is regularly varying of index $-\nu$. Hence, a heavy-tailed service time distribution gives rise to an even heavier tail of the sojourn time distribution. This is not the case in the M/G/1 processor sharing queue.

The implications of Theorem 3.1 go one step further, namely that the tail of the sojourn time distribution of a particular customer is completely determined by its own service time distribution. Only the mean service times of other customer classes appear in the asymptotic formula, as a (harmless) constant. Hence, a heavy-tailed service time distribution of a particular class of customers has no effect on the tail of the sojourn time of customers with a tail which is lighter. If a job has a long sojourn time, Theorem 3.1 suggests that this is due to the fact that its own service time is long, so the delay is not caused by extremely long service times of other jobs. In this respect, PS differs from LCFS-PR. The sojourn time of an arbitrary customer in the LCFS-PR case has (up to a constant) the same tail behaviour as in the PS-case, but this tail behaviour is the same for all types of customers. This makes the LCFS discipline unattractive for systems with multiple arrival streams.

In order to prove Theorem 3.1, we use the expression for the transform of the sojourn time distribution derived in [31] and the Tauberian theorem of Bingham and Doney [6], see also Theorem 8.1.6 in [7]. The structure of the proof is similar to that of Theorem 4.1 in [31], taking into account the presence of other types of customers.

The paper is organised as follows. Model characteristics and preliminary results are given in Section 2. Section 3 presents the main results and discusses their implications. The proof of Theorem 3.1 is given in Section 4. Concluding remarks can be found in Section 5.

## 2. Model description

In this section, we introduce some notation and mention results obtained in [31] which will be used in the remainder of this study. We consider an M/G/1 PS queue with multiple arrival streams. Since we want to study one type of customer in isolation, it suffices to consider only two streams, the second stream possibly being the aggregate of a number of arrival streams.

For $i = 1, 2$ we define the following arrival and service time characteristics. Customers of type $i$ enter the system according to a Poisson process with rate $\lambda_i > 0$. The service time of a customer of type $i$ is denoted by $B_i$, with distribution function $B_i(x)$, $B_i(0+) = 0$. The moments (if finite) and Laplace-Stieltjes transforms (LST's) of the service times are given by $\beta_{i,k}, k \geq 1$, (with $\beta_{i,1} > 0$) and $\beta_i(s)$, respectively. The workload offered by class $i$ is given by $\rho_i = \lambda_i \beta_{i,1}$. We also consider

the aggregated interarrival and service times. Therefore we define $\rho := \rho_1 + \rho_2$, $\lambda := \lambda_1 + \lambda_2$, and

$$
\begin{aligned}
B(x) &:= \frac{\lambda_1}{\lambda} B_1(x) + \frac{\lambda_2}{\lambda} B_2(x), & x &\geq 0, \\
\beta_k &:= \frac{\lambda_1}{\lambda} \beta_{1,k} + \frac{\lambda_2}{\lambda} \beta_{2,k}, & k &\geq 1, \\
\beta(s) &:= \frac{\lambda_1}{\lambda} \beta_1(s) + \frac{\lambda_2}{\lambda} \beta_2(s), & \operatorname{Re} s &\geq 0.
\end{aligned}
$$

We will denote a random variable with distribution function $B(x)$ by $B$ and assume that the system is stable, i.e. $\rho < 1$. The excess distribution of the service time and its LST are defined by

$$
\begin{aligned}
\widetilde{B}(t) &:= \frac{1}{\beta_1} \int_0^t (1 - B(x)) \mathrm{d}x, & t &\geq 0, \\
\widetilde{\beta}(s) &:= \int_0^\infty \mathrm{e}^{-st} \mathrm{d}\widetilde{B}(t) = \frac{1 - \beta(s)}{\beta_1 s}, & \operatorname{Re} s &\geq 0.
\end{aligned}
$$

A similar definition holds for $\widetilde{B}_i(t)$ and $\widetilde{\beta}_i(s)$. The $n-$fold convolution of a distribution function $F$ with itself is defined recursively by, for $x \geq 0$,

$$
\begin{aligned}
F^{0*}(x) &= 1, \\
F^{n*}(x) &= \int_0^x F^{(n-1)*}(x - u) \mathrm{d}F(u), & n = 1, 2, \dots \, .
\end{aligned}
$$

We now give a short review of the queue length and sojourn time distribution. A well known result, due to Sakata et al. [23] (see also [16]), is that the steady state distribution $(P_n)_{n \geq 0}$ of the number of customers in the system is geometrically distributed and only depends on the service time distribution through its mean:

$$
P_n = (1 - \rho)\rho^n.
$$

In the multiclass case, we have for the steady state distibution $(P_{i,j})_{i,j \geq 0}$ of customers of type 1 and 2, cf. [3, 15, 2, 12],

$$
P_{i,j} = (1 - \rho) \binom{i + j}{i} \rho_1^i \rho_2^j.
$$

The *sojourn time* of a customer (the time that a customer spends in the system) of type $i$ is denoted by $V_i$ with LST $v_i(s)$. Of special interest is the conditional sojourn time $V(\tau)$, defined as the sojourn time of a customer having processing time $\tau$. It is not difficult to see that this random variable has the same distribution for all types of customers, so we can omit the subscripts. Let $v(s, \tau)$ be the LST of $V(\tau)$. Obviously, we have the identity

$$
v_i(s) = \int_0^\infty v(s, \tau) \mathrm{d}B_i(\tau), \qquad i = 1, 2. \tag{2.1}
$$

Contrasting with the queue length distribution, the sojourn time distribution of an arbitrary customer has a fairly complex form. Yashkov has derived an expression for $v(s, \tau)$ by writing the sojourn time as a functional on a branching process, see [27]. The analysis in [27] has been extended by Ott [21]. Different approaches for obtaining an expression for $v(s, \tau)$ are undertaken in

[5, 24]. A drawback of the results in [21, 27] is that these expressions contain intricate contour integrals. In [31] a series representation for $v(s,\tau)^{-1}$ is derived which does not contain any contour integrals and is suitable for the asymptotic analysis of $v(s,\tau)$ and $v(s)$ in the next sections. This expression is given by (cf. Theorem 3.1 in [31]), for Re $s \geq 0$, $\tau \geq 0$:

$$v(s,\tau) = \left[\sum_{k=0}^{\infty} \frac{s^k}{k!}\alpha_k(\tau)\right]^{-1}. \tag{2.2}$$

The constants $\alpha_k(\tau)$ are given by $\alpha_0(\tau) := 1, \alpha_1(\tau) := \frac{\tau}{1-\rho}$, and for $k \geq 2$,

$$\alpha_k(\tau) := \frac{k}{(1-\rho)^k}\int_{x=0}^{\tau}(\tau-x)^{k-1}R^{(k-1)*}(x)\mathrm{d}x. \tag{2.3}$$

$R(x)$ is the distribution function of the waiting time in the M/G/1 FCFS queue, i.e. for Re $s \geq 0$,

$$\omega(s) := \int_{0-}^{\infty}\mathrm{e}^{-st}\mathrm{d}R(t) = \frac{1-\rho}{1-\rho\widetilde{\beta}(s)}. \tag{2.4}$$

It turns out to be fruitful to write

$$\alpha_k(\tau) = \left(\frac{\tau}{1-\rho}\right)^k - \delta_k(\tau), \tag{2.5}$$

with $\delta_0(\tau) = \delta_1(\tau) := 0$, and

$$\delta_k(\tau) := \frac{k}{(1-\rho)^k}\int_0^{\tau}(\tau-x)^{k-1}(1-R^{(k-1)*}(x))\mathrm{d}x, \qquad k = 2,3,\dots. \tag{2.6}$$

We close the section by mentioning some properties for the moments of $V(\tau)$ and $V_i$, which can be derived from (2.2). Since it follows immediately from (2.2) that $v(s,\tau)$ is analytic in $s = 0$, all moments of $V(\tau)$, denoted by $\bar{v}_k(\tau)$, are finite. They can be calculated recursively by (see [31]) $\bar{v}_0(\tau) := 1$ and

$$\bar{v}_k(\tau) = -\sum_{j=1}^{k}\binom{k}{j}\bar{v}_{k-j}(\tau)\alpha_j(\tau)(-1)^j. \tag{2.7}$$

In particular, we have the well known results (see [17, 27]),

$$\bar{v}_1(\tau) = \frac{\tau}{1-\rho}, \qquad \mathbb{V}\mathrm{ar}[V(\tau)] = \delta_2(\tau).$$

The first result can be used to obtain the mean sojourn time of a customer of type $i$. This mean is given by

$$\mathbb{E}[V_i] = \frac{\beta_{i,1}}{1-\rho}.$$

Note that this mean depends only on the workload $\rho$ and the (first moment of the) service time of type $i$. Using (2.7), it is not difficult to obtain a necessary and sufficient condition for finiteness of the moments of $\mathbb{E}[V_i^k]$, $k \geq 1$. The proof is similar to that of the single traffic stream case.

**Proposition 2.1** *For integer $k \geq 1$ and $i = 1, 2$,*

$$\mathbb{E}[V_i^k] < \infty \qquad \Leftrightarrow \qquad \beta_{i,k} < \infty.$$

**Proof** '$\Rightarrow$' is trivial. To prove '$\Leftarrow$', it suffices to show that the moments $\bar{v}_k(\tau)$ of $V(\tau)$ satisfy the inequality

$$\bar{v}_k(\tau) \leq C_k \frac{\tau^k}{(1 - \rho)^k}.$$

This follows directly from (2.7), with $C_0 := 1$ and for $k \geq 1$, $C_k := \sum_{j=0}^{k-1} \binom{k}{j} C_j$, cf. [31]. ■

Proposition 2.1 indicates that the tail behaviour of the service time distribution and the sojourn time distribution of a particular customer are similar. Furthermore, we conclude that finiteness of the moments of the sojourn time of a particular class of customers is not affected by the other class(es). The results presented in the next section strengthen these observations.

## 3. Asymptotic results and their implications

In this section we present asymptotic results for the class$-i$ sojourn time tail $\mathbb{P}(V_i > x)$ for large $x$. For the case of a single arrival stream and regularly varying service time distribution, explicit asymptotics are given in Theorem 4.1 of [31], which states for the sojourn time $V$ of an arbitrary customer that the following asymptotic formula holds:

$$\mathbb{P}(V > x) \sim \mathbb{P}(B > (1 - \rho)x), \qquad x \to \infty, \qquad (3.1)$$

when either $V$ or $B$ has a regularly varying distribution with non-integer index. In particular, the index of the tail of the sojourn time distribution is *not* larger than that of the service time distribution, contrary to the FCFS case. In this section we go one step further and show that the index of the sojourn time distribution will be the same as that of the service time distribution even if another customer class possesses a service time distribution with a heavier tail.

The main theorem is the following. With $L(.)$ we denote a slowly varying function, cf. [7].

**Theorem 3.1** *If there exists a $\mu > 1$ such that $\mathbb{E}[B^\mu] < \infty$, then the following are equivalent for non-integer $\nu > 1$,*

    **(i)**                $\mathbb{P}(B_1 > x) \sim x^{-\nu} L(x), \qquad x \to \infty,$

    **(ii)**            $\mathbb{P}(V_1 > x) \sim (1 - \rho)^{-\nu} x^{-\nu} L(x), \qquad x \to \infty.$

*so that in case of regular variation,*

$$\mathbb{P}(V_1 > x) \sim \mathbb{P}(B_1 > (1 - \rho)x), \qquad x \to \infty.$$

The condition $\mathbb{E}[B^\mu] < \infty$ in Theorem 3.1 is made for technical reasons (for which we refer to the proof in the next section); it is weak enough for all practical purposes. In particular, Theorem 3.1 provides explicit asymptotics for the following case. Suppose that we have $N$ types of customers, with service time $B_i$ and stationary sojourn time $V_i$. Using Theorem 3.1, we immediately obtain the following result (choose $\mu \in (1, \min_i \nu_i)$ in Theorem 3.1):

**Corollary 3.1** *For $i = 1, ..., N$, and non-integer $\nu_i > 1$, $\mathbb{P}[B_i > x]$ is regularly varying of index $-\nu_i$ if and only if $\mathbb{P}[V_i > x]$ is regularly varying of index $-\nu_i$. Both imply that*

$$\mathbb{P}(V_i > x) \sim \mathbb{P}(B_i > (1 - \rho)x), \qquad x \to \infty. \tag{3.2}$$

To appreciate the implications of Theorem 3.1 and Corollary 3.1, we compare the multiclass M/G/1 PS queue with other service disciplines. Suppose we have a stable M/G/1 queue with $N$ types of customers and suppose that the service time of customers of type $i$ is regularly varying of non-integer index $-\nu_i$, with $1 < \nu_1 < \nu_2 < \cdots < \nu_N$. Note that the service time of an arbitrary customer is regularly varying of index $-\nu_1$. We are interested in the tail behaviour of the sojourn time distribution of a customer under the service disciplines FCFS, LCFS-PR and PS.

For a customer of type $i$, the following holds. In the FCFS case, the tail of the customer of type 1 dominates all other types, which leads to a regularly varying sojourn time distribution of index $1 - \nu_1$ for all types. The index is increased by 1 since an arbitrary customer has to wait with positive probability for a residual service time period of a customer of type 1. In [1] it has been shown that this is the case for all non-preemptive service disciplines where at most one customer is being served at the same time.

The situation under the LCFS-PR regime is slightly better; in this case the sojourn time of an arbitrary customer is regularly varying of index $-\nu_1$, like in the PS case, see [10]. However, the customers of type 1 still dominate the sojourn time of a customer of type $i$. With positive probability, a customer of type 1 enters the system when a customer of type $i$ is being served, so customers of type 1 dominate the tail of the sojourn time distribution of type $i$.

Theorem 3.1 and Corollary 3.1 show that under the PS regime, the tail of the sojourn time distribution of a customer of type $i$ is *not* dominated by a heavier tail of a customer of another type, so that in this case, the sojourn time distribution is regularly varying of index $-\nu_i$. In particular, long-range dependent traffic ($1 < \nu_1 < 2$) has no influence on short-range dependent traffic, see also the results for general service time distributions in Remark 3.3 below.

**Remark 3.1** A natural question is whether Theorem 3.1 can be extended to a larger class of service time distributions in the following way: Is the tail behaviour of the sojourn time distribution of a customer completely determined by (the tail behaviour of) its own service time distribution? The answer to this is completely open, since no other results concerning the tail behaviour of the sojourn time distribution are known, to the best of the author's knowledge. Even for the (single-class) M/M/1 PS queue, no asymptotic formula seems to be available for $\mathbb{P}(V > x)$, although an integral representation for this probability is derived in [19]. □

**Remark 3.2** Theorem 3.1 and Corollary 3.1 show that the tail behaviour of customer class $i$ is (in case of regular variation) the same as in the M/G/1 PS queue where (only) customers of class $i$ enter and where the server works at speed $\rho_i/\rho$. A question related to the one in the previous remark is whether this holds for a more general class of service time distributions. □

**Remark 3.3** Although no other explicit asymptotic results (other than Theorem 3.1 and Theorem 4.1 of [31]) are available, it is possible to give *upper bounds* for both $\mathbb{P}(V(\tau) > x)$ and $\mathbb{P}(V_i > x)$ if the service time $B_i$ has an arbitrary distribution.

From (2.7), it is possible to show that the moments $\bar{v}_k(\tau)$ satisfy the inequality

$$\bar{v}_k(\tau) \leq k! \left( \frac{(\mathrm{e} - 1)\tau}{1 - \rho} \right)^k. \tag{3.3}$$

This implies that $v(s, \tau)$ can be extended to Re $s \geq -\frac{1-\rho}{\gamma\tau}$, $\gamma > \mathrm{e} - 1$. Hence, we have the following bound for the tail of the distribution of $V(\tau)$:

$$\mathbb{P}(V(\tau) > x) = \mathrm{o}\big(\mathrm{e}^{-\frac{1-\rho}{\gamma\tau}x}\big), \qquad\qquad x \to \infty. \tag{3.4}$$

So when the service time of a customer is bounded, its sojourn time distribution has an exponential tail. Unfortunately, (3.4) does not imply that $\mathbb{P}(V_i > x)$ has an exponential tail if this is the case for $\mathbb{P}(B_i > x)$. An application of the Laplace method shows that the upper bound for $\mathbb{P}(V > x)$ is Weibullian, i.e. $\mathbb{P}(V > x) \leq C_1 \mathrm{e}^{-C_2\sqrt{x}}$ for some positive constants $C_1$ and $C_2$.

A bound for $\mathbb{P}(V_i > x)$ in the spirit of Theorem 3.1 can be given if $\mathbb{P}(B_i > x)$ has a tail which is lighter than a power tail, i.e. if it satisfies $\mathbb{P}(B_i > x) = \mathrm{o}(x^{-T})$, $x \to \infty$, for all $T > 0$. Following the proof of Theorem 3.1 and applying Lemma 2.2 in [9] it is immediately seen that

$$\mathbb{P}(V_i > x) = \mathrm{o}(x^{-T}), \qquad\qquad x \to \infty, \qquad \forall T > 0. \tag{3.5}$$

The converse statement follows in a similar way. Hence, the sojourn time distribution has a lighter tail than any power tail iff the service time distribution has the same property, even if another customer class does have a heavy-tailed distribution. □

**Remark 3.4** Another characteristic of the sojourn time distribution which is completely determined by its own service time is the limiting distribution in heavy traffic. The following result holds:

$$(1 - \rho)V_i \Rightarrow XB_i, \qquad\qquad \rho \to 1, \tag{3.6}$$

where '$\Rightarrow$' means convergence in distribution and where $X$ is an exponentially distributed random variable having mean 1, which is independent of $B_i$. The derivation is the same as in the single arrival stream case, see [25, 30, 31] for more results and references. □

## 4. Proof of Theorem 3.1

In this section we give a proof of Theorem 3.1. The proof makes use of the Tauberian theorem of Bingham and Doney (Theorem 8.1.6 in [7]) and the expression (2.2) for the LST of the sojourn time distribution.

Before we give a proof of Theorem 3.1, we make some preparations in the following three lemmas.

**Lemma 4.1** *If* $\mathbb{E}[B^\mu] < \infty$ *for some* $\mu > 1$ *then there exists a* $\delta > 0$ *such that for every* $n \geq 1$,

$$\bar{v}_n(\tau) - \bar{v}_1^n(\tau) = o(\tau^{n-\delta}), \qquad\qquad \tau \to \infty, \tag{4.1}$$

$$\sqrt{\mathbb{V}ar[V^n(\tau)]} = o(\tau^{n-\delta}), \qquad\qquad \tau \to \infty. \tag{4.2}$$

**Proof** The lemma is proven by combining (2.4)–(2.7). From (2.4) and (2.6) we obtain for $k \geq 2$,

$$\int_0^\infty \mathrm{e}^{-s\tau}\mathrm{d}\delta_k(\tau) = \frac{1}{s^k}\frac{k!}{(1-\rho)^k}\left(1 - \omega^{k-1}(s)\right). \tag{4.3}$$

Since $\mathbb{E}[B^\mu] < \infty$ for a $\mu > 1$ it follows that, cf. p. 199 in [18],

$$\beta(s) = 1 - \beta_1 s + \mathrm{O}(|s^\mu|), \qquad\qquad s \downarrow 0. \tag{4.4}$$

This implies, using (2.4), for $\delta \in (0, \mu - 1)$ and $k \geq 2$,

$$\omega^{k-1}(s) = 1 - o(s^\delta), \qquad s \downarrow 0. \tag{4.5}$$

Hence, from (4.3) it follows for $k \geq 2$,

$$\int_0^\infty e^{-s\tau} d\delta_k(\tau) = o(s^{\delta-k}), \qquad s \downarrow 0. \tag{4.6}$$

Since the function $\delta_k(\tau)$ is non-decreasing in $\tau$ it follows from Karamata's Tauberian theorem (see Theorem 1.7.1 in [7]) that for $k \geq 2$,

$$\delta_k(\tau) = o(\tau^{k-\delta}), \qquad \tau \to \infty. \tag{4.7}$$

Equation (4.1) now follows by an inductive argument using (2.5), (2.7) and (4.7). To prove (4.2), we have by using (4.1) for both $\bar{v}_{2n}(\tau)$ and $\bar{v}_n(\tau)$,

$$\mathrm{Var}(V^n(\tau)) = \bar{v}_{2n}(\tau) - \bar{v}_n^2(\tau) = o(\tau^{2n-\delta}),$$

which proves (4.2) with $\delta$ replaced by $\frac{1}{2}\delta$. ∎

Define

$$f(s, \tau) := v(s, \tau) - e^{-\frac{s\tau}{1-\rho}}. \tag{4.8}$$

We have the following bound for $f(s, \tau)$.

**Lemma 4.2** *If* $\mathbb{E}[B^\mu] < \infty$ *for a* $\mu > 1$, *then, for* $\gamma \in (0, 1)$, $\gamma < \mu - 1$,

$$f(s, \tau) = o(s^\gamma), \qquad \tau = O\left(1/s\right), \qquad s \to 0.$$

**Proof** Without loss of generality, it can be assumed that $\mu < 2$. From (2.2) and (2.5) we get

$$f(s, \tau) = \frac{e^{-\frac{2s\tau}{1-\rho}} \sum_{k=2}^\infty \frac{s^k}{k!} \delta_k(\tau)}{1 - e^{-\frac{s\tau}{1-\rho}} \sum_{k=2}^\infty \frac{s^k}{k!} \delta_k(\tau)}. \tag{4.9}$$

It follows immediately from (4.9), using $\delta_k(\tau) \leq \frac{\tau^k}{(1-\rho)^k}$, that for real $s \geq 0$,

$$f(s, \tau) \leq \frac{1}{1 + \frac{s\tau}{1-\rho}} e^{-\frac{s\tau}{1-\rho}} \sum_{k=2}^\infty \frac{s^k}{k!} \delta_k(\tau) \leq e^{-\frac{s\tau}{1-\rho}} \sum_{k=2}^\infty \frac{s^k}{k!} \delta_k(\tau). \tag{4.10}$$

We now derive an upper bound for $\delta_k(\tau)$. In view of (2.6), we need an upper bound for $1 - R^{(k-1)*}(x)$. From (4.5) with $k = 2$ and Lemma 2.2 in [9], we obtain for $\varepsilon \in (0, \mu - 1)$

$$1 - R(x) = o(x^{-\varepsilon}), \qquad x \to \infty. \tag{4.11}$$

Let $(W_i)_{i \geq 1}$ be an i.i.d. sequence with distribution function $R(x)$. Then, we have

$$
\begin{aligned}
1 - R^{(k-1)*}(x) &= \mathbb{P}(W_1 + \cdots + W_{k-1} > x) \leq \mathbb{P}(\cup_{i=1}^{k-1} \{W_i > \frac{x}{k-1}\}) \\
&\leq (k-1)\mathbb{P}(W_1 > \frac{x}{k-1}).
\end{aligned}
$$

Hence, from this and (4.11) we get for $x \to \infty$,

$$1 - R^{(k-1)*}(x) \leq (k-1)^2 \mathrm{o}(x^{-\varepsilon}), \tag{4.12}$$

where $\mathrm{o}(x^{-\varepsilon})$ is independent of $k \geq 2$. This implies, for Re $s \geq 0$,

$$\delta_k(\tau) \leq k(k-1)^2 \left(\frac{\tau}{1-\rho}\right)^k \mathrm{o}(\tau^{-\varepsilon}), \tag{4.13}$$

where $\tau \to \infty$. Since $\tau = \mathrm{O}(1/s), s \to 0$, it follows that

$$
\begin{aligned}
f(s,\tau) &\leq \mathrm{e}^{-\frac{s\tau}{1-\rho}} \sum_{k=2}^{\infty} \frac{s^k}{k!}(k-1)^2 k \left(\frac{\tau}{1-\rho}\right)^k \mathrm{o}(\tau^{-\varepsilon}) \\
&\leq \mathrm{o}(\tau^{-\varepsilon})\mathrm{e}^{-\frac{s\tau}{1-\rho}} \left(\frac{s\tau}{1-\rho}\right)^2 \sum_{k=0}^{\infty} \frac{k+1}{k!} \left(\frac{s\tau}{1-\rho}\right)^k \\
&= \mathrm{o}(\tau^{-\varepsilon}) \left(\frac{s\tau}{1-\rho}\right)^2 \left[1 + \left(\frac{s\tau}{1-\rho}\right)\right] = \mathrm{o}(\tau^{-\varepsilon}) \left(\frac{s\tau}{1-\rho}\right)^2 (1 + \mathrm{O}(1)) = \mathrm{o}(s^{\varepsilon}).
\end{aligned}
$$

Since this result applies for all $\varepsilon \in (0, \mu - 1)$, the lemma is proven. ∎

The $n-$th derivative of $f(s, \tau)$ with respect to $s$ is defined by

$$f^{(n)}(s,\tau) := \frac{\partial^n}{\partial s^n} f(s,\tau). \tag{4.14}$$

The following upper bound for $f^{(n+1)}(s, \tau)$ will be useful.

**Lemma 4.3** *For $n \geq 1$, Re $s \geq 0$, $\tau \geq 0$,*

$$|f^{(n+1)}(s,\tau)| \leq e^{-s\tau}\sqrt{\mathbb{V}\mathrm{ar}(V^{n+1}(\tau))} + v(s,\tau)\left(\bar{v}_{n+1}(\tau) - \bar{v}_1^{n+1}(\tau)\right) + \left(\frac{\tau}{1-\rho}\right)^{n+1} f(s,\tau).$$

**Proof**  Using the probabilistic interpretation $f(s,\tau) = \mathbb{E}[\mathrm{e}^{-sV(\tau)}] - \mathrm{e}^{-\frac{s\tau}{1-\rho}}$, we obtain

$$|f^{(n+1)}(s,\tau)| = \left|\mathbb{E}[V^{n+1}(\tau)\mathrm{e}^{-sV(\tau)}] - \left(\frac{\tau}{1-\rho}\right)^{n+1}\mathrm{e}^{-\frac{s\tau}{1-\rho}}\right|$$

$$\leq \left|\mathbb{E}[V^{n+1}(\tau)\mathrm{e}^{-sV(\tau)}] - \mathbb{E}[V^{n+1}(\tau)]\mathbb{E}[\mathrm{e}^{-sV(\tau)}]\right| +$$

$$\left|\mathbb{E}[V^{n+1}(\tau)]\mathbb{E}[\mathrm{e}^{-sV(\tau)}] - \left(\frac{\tau}{1-\rho}\right)^{n+1}\mathbb{E}[\mathrm{e}^{-sV(\tau)}]\right| +$$

$$\left|\left(\frac{\tau}{1-\rho}\right)^{n+1}\mathbb{E}[\mathrm{e}^{-sV(\tau)}] - \left(\frac{\tau}{1-\rho}\right)^{n+1}\mathrm{e}^{-\frac{s\tau}{1-\rho}}\right|$$

$$= |\mathbb{Cov}[V^{n+1}(\tau), e^{-sV(\tau)}]| + v(s,\tau)\left(\bar{v}_{n+1}(\tau) - \left(\frac{\tau}{1-\rho}\right)^{n+1}\right) + \left(\frac{\tau}{1-\rho}\right)^{n+1} f(s,\tau).$$

The first term can be bounded by the inequality of Cauchy-Schwarz and noting that $\mathbb{Var}[e^{-sV(\tau)}] \leq \mathbb{E}[e^{-2sV(\tau)}] \leq e^{-2s\tau}$, since $V(\tau) \geq \tau$.                                    ∎

**Proof of Theorem 3.1**

Recall that $\mathbb{E}[B^\mu] < \infty$ for some $\mu > 1$. Let $\nu \in (n, n+1)$. Without loss of generality, we can assume that $\mu < \min(\nu, 2)$. By Theorem 8.1.6 in [7], it suffices to show that (i) or (ii) in Theorem 3.1 implies that for real $s \downarrow 0$,

$$v_1(s) - \beta_1\left(\frac{s}{1-\rho}\right) - \sum_{k=0}^{n} \frac{(-s)^k}{k!}\left(\mathbb{E}[V_1^k] - \frac{\beta_{1,k}}{(1-\rho)^k}\right) = o(s^\nu L(1/s)). \tag{4.15}$$

Write

$$v_1(s) - \beta_1\left(\frac{s}{1-\rho}\right) - \sum_{k=0}^{n} \frac{(-s)^k}{k!}\left(\mathbb{E}[V_1^k] - \frac{\beta_{1,k}}{(1-\rho)^k}\right) = \int_0^\infty R_n(s,\tau)\mathrm{d}B_1(\tau),$$

with $R_n(s,\tau)$ the residual term of the $n-$term Taylor expansion of $f(s,\tau)$ in $s = 0$, i.e.

$$R_n(s,\tau) = f(s,\tau) - \sum_{k=0}^{n} \frac{s^k}{k!}f^{(k)}(0,\tau). \tag{4.16}$$

Since $f(s,\tau)$ is analytic in $s = 0$, we can apply Taylor's theorem, which gives, for $s$ in a neighborhood of 0,

$$|R_n(s,\tau)| = \left|\int_0^s \frac{(s-u)^n}{n!}f^{(n+1)}(u,\tau)\mathrm{d}u\right| \leq s^n \int_0^s \left|f^{(n+1)}(u,\tau)\right|\mathrm{d}u. \tag{4.17}$$

Using Lemma 4.3 and (4.17) we obtain

$$\int_0^\infty |R_n(s,\tau)|\,\mathrm{d}B_1(\tau) \leq s^n \int_0^\infty \sqrt{\mathbb{Var}(V^{n+1}(\tau))} \int_0^s e^{-u\tau}\mathrm{d}u\mathrm{d}B_1(\tau) +$$

$$s^n \int_0^\infty \left(\bar{v}_{n+1}(\tau) - \bar{v}_1^{n+1}(\tau)\right) \int_0^s v(u,\tau)\mathrm{d}u\mathrm{d}B_1(\tau) + s^n \int_0^\infty \bar{v}_1^{n+1}(\tau) \int_0^s f(u,\tau)\mathrm{d}u\mathrm{d}B_1(\tau)$$

$$=: I + II + III.$$

This implies that the proof of Theorem 3.1 is complete if we have shown that all three integrals (I, II and III) on the right hand side are of $o(s^\nu L(1/s))$ for $s \downarrow 0$. Suppose now that (i) holds with equality (which is no restriction, since the class of regularly varying distribution is closed under tail equivalence).

**Part I** It is convenient to split the integral in two parts. Using part (ii) of Lemma 4.1 for $\tau$ large we get for a $\delta > 0$ and a finite constant $M$:

$$s^n \int_0^\infty \sqrt{\mathbb{V}\mathrm{ar}(V^{n+1}(\tau))} \int_0^s e^{-u\tau} \mathrm{d}u \mathrm{d}B_1(\tau) \leq$$

$$Ms^n \int_0^{s^{-1}} \tau^{n+1-\delta} \int_0^s e^{-u\tau} \mathrm{d}u \mathrm{d}B_1(\tau) + Ms^n \int_{s^{-1}}^\infty \tau^{n+1-\delta} \int_0^s e^{-u\tau} \mathrm{d}u \mathrm{d}B_1(\tau), \tag{4.18}$$

where $M$ is some finite constant. The first part of (4.18) can be bounded by using $e^{-u\tau} \leq 1$ and $\tau \leq s^{-1}$:

$$Ms^n \int_0^{s^{-1}} \tau^{n+1-\delta} \int_0^s e^{-u\tau} \mathrm{d}u \mathrm{d}B_1(\tau) \leq Ms^{n+1} \int_0^{s^{-1}} \tau^{n+1-\delta} \mathrm{d}B_1(\tau)$$

$$\leq Ms^{\nu+\frac{1}{2}\delta} \int_0^{s^{-1}} \tau^{\nu-\frac{1}{2}\delta} \mathrm{d}B_1(\tau) \leq M\mathbb{E}[B_1^{\nu-\frac{1}{2}\delta}]s^{\nu+\frac{1}{2}\delta}.$$

To bound the second integral in the right-hand side of (4.18), use $\int_0^s e^{-u\tau} \mathrm{d}u \leq \frac{1}{\tau}$ and apply partial integration:

$$Ms^n \int_{s^{-1}}^\infty \tau^{n+1-\delta} \int_0^s e^{-u\tau} \mathrm{d}u \mathrm{d}B_1(\tau) \leq -Ms^n \int_{s^{-1}}^\infty \tau^{n-\delta} \mathrm{d}(1 - B_1(\tau))$$

$$= M(1 - B_1(1/s))s^\delta + M(n-\delta)s^n \int_{s^{-1}}^\infty \tau^{n-1-\delta}(1 - B_1(\tau))\mathrm{d}\tau$$

$$= Ms^{\nu+\delta}L(1/s) + M(n-\delta)s^n \int_{s^{-1}}^\infty \tau^{n-1-\delta-\nu}L(\tau)\mathrm{d}\tau.$$

It follows from Karamata's theorem that the expression in the right hand side behaves proportionally to $s^{\nu+\delta}L(1/s)$ for real $s \downarrow 0$, which completes the proof of part I.

**Part II** Identical to Part I, using part (i) of Lemma 4.1, and $v(s, \tau) \leq e^{-s\tau}$.

**Part III** We split the leftmost integral in III again up in two parts, namely $0 \leq \tau \leq T/s$ for some finite $T$, and $\tau \geq T/s$. Using Lemma 4.2 and a similar calculation as in the first part of the proof of Part I, we can conclude that the first integral is $o(s^{\nu+\varepsilon})$, for an $\varepsilon > 0$ and $s \downarrow 0$. This result holds for each finite $T$. Bounding the second integral is more difficult than in Part I and II. It follows

from the first inequality in (4.10) and $\delta_k(\tau) \leq \left(\frac{\tau}{1-\rho}\right)^k$, $k \geq 2$, that $f(s,\tau) \leq 1/(1 + \frac{s\tau}{1-\rho})$. This implies that

$$\int_0^s f(u,\tau)\mathrm{d}u \leq \int_0^s \frac{1}{1 + \frac{u\tau}{1-\rho}}\mathrm{d}u = \frac{1-\rho}{\tau}\ln\left(1 + \frac{s\tau}{1-\rho}\right).$$

Note that for each $\gamma > 0$ we have for large $T$ that $\ln\left(1 + \frac{s\tau}{1-\rho}\right) \leq (s\tau)^\gamma$ if $\tau \geq T/s$. The second integral can now be handled by using partial integration, with $\gamma \in (0, \nu - n)$:

$$s^n \int_{T/s}^\infty \left(\frac{\tau}{1-\rho}\right)^{n+1} \int_0^s f(u,\tau)\mathrm{d}u\mathrm{d}B_1(\tau) \leq s^{n+\gamma} \int_{T/s}^\infty \tau^{n+\gamma}\mathrm{d}B_1(\tau)$$

$$= s^{n+\gamma}(T/s)^{n+\gamma}(1 - B_1(T/s)) + (n+\gamma)s^{n+\gamma}\int_{T/s}^\infty (1 - B_1(\tau))\tau^{n+\gamma-1}\mathrm{d}\tau$$

$$= s^\nu L(1/s))T^{n+\gamma-\nu} + (n+\gamma)s^{n+\gamma}\int_{T/s}^\infty \tau^{n+\gamma-1-\nu}\mathrm{d}\tau$$

$$\sim s^\nu L(1/s)T^{n+\gamma-\nu}\left(1 + \frac{n+\gamma}{\nu - n - \gamma}\right). \tag{4.19}$$

The last result holds for $s \downarrow 0$ for each $T > 0$ by an application of Karamata's theorem. Finally, the proof of Part III follows from (4.19) and the bound for the first part of the integral in III by choosing $T$ arbitrarily large.

We conclude that (4.15) holds, which implies (ii). If (ii) holds with equality, the proof of I, II, and III (and hence that of (i)) follows similarly, using the stochastic dominance $1 - B_1(\tau) \leq \mathbb{P}(V_1 > \tau) = (1-\rho)^{-\nu_i}\tau^{-\nu_i}L(\tau)$; we omit the details. ∎

## 5. CONCLUDING REMARKS

In this paper, the tail behaviour of the sojourn time distributions in a multiclass processor sharing queue has been studied. Both explicit asymptotics (Theorem 3.1 and Corollary 3.1) and bounds (Remark 3.3) for the tail of the sojourn time distributions have been derived. The main result is that the sojourn time distribution of a particular customer class $i$ is regularly varying of index $-\nu_i$ iff the service time distribution is regularly varying of index $-\nu_i$, with $\nu_i$ non-integer. Both imply that the asymptotic relation $\mathbb{P}(V_i > x) \sim \mathbb{P}(B_i > x(1-\rho))$ holds for $x \to \infty$.

The results in Section 3 indicate that extremely long jobs have limited influence on the sojourn times of other jobs when compared to service disciplines like FCFS and LCFS-PR. In particular, short-range dependent traffic does not suffer from long-range dependent traffic. This property is of relevance for networks with different types of traffic where some of the traffic streams exhibit heavy-tailedness.

A related service discipline which receives quite some attention is *generalised processor sharing*, see e.g. [8] and references therein. An interesting subject for further research is to examine under which conditions Theorem 3.1 can be extended to this model.

Another topic for further research is the extension of Theorem 3.1 to a larger class of (heavy-tailed) distributions. An extension to regular variation with integer index might be possible for a particular class of slowly varying functions, using the techniques in [20]. A generalisation to subexponential distribution functions seems difficult to prove along the lines of this paper, since no characterisation of a subexponential distribution function by its LST is available.

**Acknowledgements**
The author is indebted to Onno Boxma and Sindo Núñez Queija for interesting discussions and useful remarks.

# References

1. V. Anantharam. Scheduling strategies and long-range dependence. Report, Dept. of Electrical Engineering & Computer Sciences, Univ. of California at Berkeley, 1997.

2. A.D. Barbour. Networks of queues and the method of stages. *Advances in Applied Probability* 8:584-591, 1976.

3. F. Baskett, K.M. Chandy, R.R. Muntz, and F.G. Palacios. Open, closed, and mixed networks of queues with different classes of customers. *Journal of the ACM*, 22:248-260, 1975.

4. J. Beran, R. Sherman, M.S. Taqqu, and W. Willinger. Long-range dependence in variable-bit-rate video. *IEEE Transactions on Communications*, 43:1566-1579, 1995.

5. J.L. van den Berg and O.J. Boxma. The M/G/1 queue with processor sharing and its relation to a feedback queue. *Queueing Systems,* 9:365-401, 1991.

6. N.H. Bingham and R.A. Doney. Asymptotic properties of supercritical branching processes I: The Galton-Watson process. *Advances in Applied Probability*, 6:711-731, 1974.

7. N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation.* Cambridge University Press, Cambridge, 1987.

8. S.C. Borst, P. Jelenković, and O.J. Boxma. Generalized processor sharing with subexponential traffic sources. Report, 1998.

9. O.J. Boxma and V. Dumas. The busy period in the fluid queue. *Performance Evalution Review,* 26:100-110, 1998.

10. O.J. Boxma and J.W. Cohen. The single server queue: heavy tails and heavy traffic. To appear in *Self-similar Network Traffic and Performance Evaluation,* eds. K. Park and W. Willinger, Wiley, 1998.

11. J.W. Cohen. Some results on regular variation in queueing and fluctuation theory. *Journal of Applied Probability,* 10:343-353, 1973.

12. J.W. Cohen. The multiple phase service network with generalized processor sharing. *Acta Informatica,* 12:245-284, 1979.

13. P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events.* Springer Verlag, Berlin, 1997.

14. W. Feller. *An Introduction to Probability Theory and its Applications,* Volume II. Second edition, Wiley, New York, 1971.

15. F.P. Kelly. Networks of queues. *Advances in Applied Probability* 8, 416-432, 1976.

16. F.P. Kelly. *Reversibility and Stochastic Networks.* Wiley, New York, 1979.

17. L. Kleinrock. *Queueing Systems, Volume II: Computer Applications.* Wiley, New York, 1976.

18. M. Loève. *Probability Theory.* Second edition, Van Nostrand, New York, 1960.

19. J.A. Morrison. Response-time distribution for a processor sharing system. *SIAM Journal of Applied Mathematics* 45:152-167, 1985.

20. A. de Meyer and J.L. Teugels. On the asymptotic behaviour of the distributions of the busy period and service time in M/G/1. *Journal of Applied Probability,* 17:802-813, 1980.

21. T.J. Ott. The sojourn-time distribution in the M/G/1 queue with processor sharing. *Journal of Applied Probability,* 21:360-378, 1984.

22. V. Paxson and S. Floyd. Wide area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking,* 3:226-244, 1995.

23. M. Sakata, S. Noguchi, and J. Oizumi. Analysis of a processor shared queueing model for time sharing systems. *Proceedings of the second Hawaii International Conference on System Sciences,* 625-628, 1969.

24. R. Schassberger. A new approach to the M/G/1 processor sharing queue. *Advances in Applied Probability,* 16:802-813, 1984.

25. B. Sengupta. An approximation for the sojourn-time distribution for the GI/G/1 processor-sharing queue. *Communications in Statistics - Stochastic Models,* 8:35-57, 1992.

26. W. Willinger, M.S. Taqqu, W.E. Leland, and D.V. Wilson. Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements. *Statistical Science,* 10:67-85, 1995.

27. S.F. Yashkov. A derivation of response time distribution for a M/G/1 processor sharing queue. *Problems of Control and Information Theory* 12:133-148, 1983.

28. S.F. Yashkov. Processor-sharing queues: Some progress in analysis. *Queueing Systems,* 2:1-17, 1987.

29. S.F. Yashkov. Mathematical problems in the theory of shared-processor systems. *Journal of Soviet Mathematics,* 58:101-147, 1992.

30. S.F. Yashkov. On a heavy traffic limit theorem for the M/G/1 processor sharing queue. *Communications in Statistics - Stochastic Models,* 9:467-471, 1993.

31. A.P. Zwart and O.J. Boxma. Sojourn time asymptotics in the M/G/1 processor sharing queue. CWI Report PNA-R9802, CWI, Amsterdam, 1998.