Memorandum No. 1766

**Insensitive bounds for the moments
of the sojourn time distribution
in the M/G/1 processor-sharing queue**

S.K. Cheung, J.L. van den Berg[1]
and R.J. Boucherie

June, 2005

[1]TNO Information and Communication Technology, PO Box 5050, 2600 GB Delft, The Netherlands

Insensitive bounds for the moments of the sojourn time distribution in the M/G/1 processor-sharing queue

Sing-Kong Cheung [a], Hans van den Berg [b,c], and Richard J. Boucherie [a]

[a] University of Twente, Department of Applied Mathematics
Stochastic Operations Research Group
P.O. Box 217, 7500 AE Enschede, The Netherlands

[b] TNO Information and Communication Technology
P.O. Box 5050, 2600 GB Delft, The Netherlands

[c] University of Twente, Department of Computer Science
Design and Analysis of Communication Systems
P.O. Box 217, 7500 AE Enschede, The Netherlands

E-mail: {S.K.Cheung, R.J.Boucherie}@utwente.nl, J.L.vandenBerg@telecom.tno.nl

Abstract: This paper studies the M/G/1 processor-sharing (PS) queue and the sojourn time distribution conditioned on the initial job size. Although several expressions for the Laplace-Stieltjes transform (LST) are known, these expressions are not applicable for computational purposes. This paper derives readily applicable expressions for insensitive bounds of *all* moments of the conditional sojourn time distribution. The *instantaneous* sojourn time, the sojourn time of a very small job, leads to insensitive upper bounds with special structure requiring only knowledge of the traffic load and the initial job size. Interestingly, the special form of the upper bounds involves polynomials with so-called Eulerian numbers as coefficients. In addition, stochastic ordering and moment ordering results for the sojourn time distribution are obtained.

Keywords: M/G/1 PS, conditional sojourn time, moments, insensitive bounds, instantaneous sojourn time, Euler's number triangle, Laplace transform- and moment ordering.

AMS Subject Classifications: 60K25, 60E15

1   Introduction

With the introduction of time-sharing computing in the nineteen sixties, people became interested in the processor-sharing (PS) discipline as the idealization of time-sharing queueing models. Nowadays, the PS service discipline is of considerable interest in many application areas in which different users receive a share of a scarce common system resource. In particular, in the field of the performance evaluation of computer and communication systems, the PS discipline has been widely adopted as a convenient paradigm for modelling bandwidth sharing.

In this paper we study the sojourn time of the classical M/G/1 queue with egalitarian PS service discipline, conditioned on the initial service requirement (job size) of the customer. In the

egalitarian PS service discipline, every customer simultaneously gets an equal share of $1/n$ of the service capacity when $n > 0$ customers (jobs) are in the system. An extensive body of literature on processor-sharing queues was initiated by Kleinrock [9, 10], who studied the M/M/1 case. In particular he showed that the mean sojourn time conditional on the service requirement is proportional to the service requirement. For an extensive overview of the literature on processor-sharing queues we refer to Yashkov's survey papers [23, 24].

The exact determination of the (stationary) sojourn time distribution in the M/G/1 processor-sharing queue was an open problem for a long time. After puzzling researchers for 15 years, Yashkov [22] found an analytic solution in terms of double Laplace transforms. Schassberger [16] provided another approach to the exact solution by considering the PS discipline as a limit of the round-robin discipline. Other solutions (via different methods) and additional contributions were also made by for example Ott [14], Van den Berg [1], Whitt [20], and Núñez-Queija [13]. Recently, Zwart and Boxma [26] derived a new expression for the Laplace-Stieltjes transform (LST) of the sojourn time distribution, that avoids the complex contour integrals of most of the previous results.

Despite the availability of several exact expressions for the LST of the conditional sojourn time distribution, these expressions are generally of complex form and simple explicit expressions for the distribution function and general moments seem not to be available. A recursive expression for the moments has been derived in [26]. However, the exact and explicit expression for the k-th moment of the conditional sojourn time distribution increases significantly in complexity and in length, as k grows larger. Even for the fourth moment in the M/M/1 PS case, the explicit closed-form expression is very lengthy. In this paper, we use this recursive formula to obtain new properties for and insights in the sojourn time distribution. In particular, we are interested in bounds for the moments of the (conditional) sojourn time distribution in the M/G/1 PS queue.

The first moment of the conditional sojourn time distribution is well-known to be a linear function of the initial service requirement (see e.g. [11]) and is independent of the service time distribution apart from its mean (see e.g. [5]). A customer with x times larger service requirement will have x times larger mean sojourn time. Therefore, the processor-sharing discipline is often referred to a *fair* policy. Also in PS queues, the influence of long jobs on the sojourn time of the short jobs is limited, since short jobs can overtake long jobs. By studying the bounds of higher moments we strengthen the fact that PS is indeed a very fair service discipline.

For the second moment of the conditional sojourn time distribution, Van den Berg [1] obtained an upper (and lower) bound that only depends on the workload ρ and the initial service requirement τ. In addition, the bounds have the attractive property of insensitivity to the service time distribution; and the difference between the upper and lower bound is small, particularly for small and moderate values of the workload ρ.

Motivated by [1], we generalize these second moment bounds into higher moments of the sojourn time distribution. In particular, in this paper we will obtain tight upper and lower bounds for *all* moments of the conditional sojourn time distribution and we give related stochastic ordering results. The main result is that there exists an upper bound for the k-th moment of the sojourn time distribution, for all $k = 1, 2, ...$, and conditioned on the initial service requirement $\tau > 0$, that only depends on the workload ρ and the initial service requirement τ; it does *not* depend on the service time distribution, except for its mean. This shows that the impact of excessive behavior of other customer is always limited for the sojourn time of a tagged customer.

The upper bound for the k-th moment involves a polynomial in ρ of degree $k - 1$, which has so-called Eulerian numbers as coefficients. Eulerian numbers and the related Euler's number triangle have many combinatorial interpretations and applications, see e.g. [3, 6, 19]. The poly-

nomial structure of the upper bound is obtained from the recursion of [26], and a generalization of results from [1]; see Section 3.

The coefficients in the upper bounds also arise by means of an *instantaneous* sojourn time analysis, where we analyze the sojourn time of a very small job ($\tau \to 0$); see Section 4. Surprisingly, the instantaneous sojourn time analysis provides us coefficients of the polynomial that turns out to be Eulerian numbers. A formal proof of the upper bound containing Eulerian numbers is given in Section 5, together with a more general stochastic ordering- and moment ordering results. The general moment ordering result is stated for two different but related processor-sharing queues, both with a random number of *permanent* customers. Finally, we illustrate the quality of the bounds in Section 6 and summarize our findings in Section 7.

2 Preliminaries

In this section we introduce the notation used in the paper and give a short review of the M/G/1 PS queue. Customers arrive at a single server queue according to a Poisson process with rate $\lambda > 0$. Their required service times are i.i.d. random variables with a general distribution $B(x)$ and $B(0+) = 0$. Let β_k denote the k-th moment of the service time distribution. The server shares its fixed capacity among all customers present in the system. Thus, every customer is being served with rate $1/n$, when $n > 0$ customers are present in the system. Assume that the workload is less than one, i.e., the system is in steady state or $\rho := \lambda\beta_1 < 1$.

The steady state queue length distribution $(\pi_n)_{n \in \mathbb{N}_0}$ is geometrically distributed and only depends on the first moment of the service time distribution (see [15]):

$$\pi_n = (1 - \rho)\rho^n, \quad n \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}. \tag{2.1}$$

We shall let $V(\tau)$ denote the (conditional) sojourn time of a customer entering the system in steady state having a required service time equal to τ upon arrival. Define the k-th moment by $v_k(\tau) = \mathbb{E}V(\tau)^k$. The first moment of $V(\tau)$ is given by $v_1(\tau) = \tau/(1-\rho)$, i.e., the mean sojourn time $v_1(\tau)$ is proportional to τ and is independent of the initial service time distribution $B(x)$ for fixed β_1 (see e.g. [11, 5]). In this paper, we analyze higher moments $v_k(\tau)$ of the sojourn time $V(\tau)$ for all $k \in \mathbb{N}$.

Define the Laplace-Stieltjes transform (LST) of $V(\tau)$ by $v(s, \tau) = \mathbb{E}\left[e^{-sV(\tau)}\right]$, for $\text{Re}\, s \geq 0$ and $\tau \geq 0$. Yashkov [22] derived an expression for $v(s, \tau)$ by writing the sojourn time as a functional on a branching process. Using this structure of the branching process, Yashkov found and solved a system of differential equations determining $v(s, \tau)$. The result is

$$v(s, \tau) = \frac{(1 - \rho)e^{-(s+\lambda)\tau}}{\psi(s, \tau) - \lambda \int_0^\tau e^{-(s+\lambda)x}\psi(s, \tau - x)\overline{B}(x)dx - \lambda e^{-(s+\lambda)\tau}\int_\tau^\infty \overline{B}(x)dx},$$

where $\psi(s, \tau)$ is the LST of some unknown function which in turn has a Laplace transform $\widetilde{\psi}(s, q)$ (with argument q):

$$\widetilde{\psi}(s, q) = \frac{q + s + \lambda\beta(q + s + \lambda)}{(q + s + \lambda)[q + \lambda\beta(q + s + \lambda)]},$$

$$\psi(s, \tau) = \frac{1}{2\pi i} \int_{-i\infty+0}^{+i\infty+0} \widetilde{\psi}(s, q)e^{q\tau}dq,$$

with $\beta(s) = \int_0^\infty e^{-sx}dB(x)$ as the LST of the service time distribution and $\overline{B}(x) = 1 - B(x)$.

3

Via different approaches, similar results for $v(s, \tau)$ are obtained in [14, 16]. These alternative expressions (and Yashkov's result) all contain contour integrals which are inversion formulas of Laplace transforms. The expression for $v(s, \tau)$ obtained by Zwart and Boxma [26] that avoids the contour integrals, is the most suitable one for our purposes. They showed that $v(s, \tau)^{-1}$ can be written as an exponential generating function of the sequence $\alpha_n(\tau), n = 0, 1, ...,$ i.e.,

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

The coefficients $\alpha_n(\tau)$ are related to the waiting time distribution in the equivalent M/G/1 queue with First Come First Serve (FCFS) discipline: $\alpha_0(\tau) := 1$, and for $n \geq 1$,

$$\alpha_n(\tau) = \frac{n}{(1-\rho)^n} \int_{x=0}^{\tau} (\tau - x)^{n-1} R^{(n-1)*}(x) dx, \tag{2.3}$$

where R^{n*} is the n-fold (Stieltjes) convolution of the waiting time distribution $R(x)$ in the M/G/1 FCFS queue: $R^{0*}(x) = 1$, and

$$R^{n*}(x) = \int_0^x R^{(n-1)*}(x - u) dR(u), \quad n \in \mathbb{N}. \tag{2.4}$$

The waiting time distribution $R(x)$ in the M/G/1 FCFS queue is given by the Pollaczek-Khintchine formula:

$$w(s) := \int_0^{\infty} e^{-sx} dR(x) = \frac{1 - \rho}{1 - \rho \left(\frac{1 - \beta(s)}{\beta_1 s} \right)},$$

and by inversion of $w(s)^n$ it can be shown that (cf. [26])

$$R^{n*}(x) = (1-\rho)^n \sum_{m=0}^{\infty} \binom{m+n-1}{n-1} \rho^m \widetilde{B}^{m*}(x), \tag{2.5}$$

where $\widetilde{B}^{m*}(x)$ is the m-fold convolution of the integrated tail- or excess service time distribution $\widetilde{B}(x) = \frac{1}{\beta_1} \int_0^x (1 - B(u)) du$.

By differentiating both sides of the identity $v(s, \tau) \sum_{n=0}^{\infty} \frac{s^n}{n!} \alpha_n(\tau) = 1$, with respect to s and putting $s = 0$, it is shown in [26] that the moments $v_k(\tau)$ can be calculated recursively, as $v_0(\tau) := 1$ and for $k \geq 1$,

$$v_k(\tau) = -\sum_{j=1}^{k} \binom{k}{j} v_{k-j}(\tau) \alpha_j(\tau) (-1)^j. \tag{2.6}$$

In particular, it holds that $v_1(\tau) = \alpha_1(\tau) = \tau/(1-\rho)$, and $v_2(\tau) = 2\tau^2/(1-\rho)^2 - \alpha_2(\tau)$.

3 Upper and lower bounds for the conditional sojourn time

In this section, we establish the bounds for the moments of the conditional sojourn time distribution, which have the form $1 \leq (1-\rho)^k v_k(\tau)/\tau^k \leq \phi_{k-1}(\rho)$, where $\phi_{k-1}(\rho)$ is a polynomial in ρ of (at most) degree $k - 1$ and with non-negative coefficients. These bounds are insensitive to the service time distribution for all $k \geq 1$, and existence is guaranteed by $\rho < 1$, regardless

of the higher moments of the service time distribution. Van den Berg [1] obtained the simple bounds for the second moment of the sojourn time in the M/G/1 PS queue:

$$z_2(\tau) := \frac{1}{(1-\rho)^2}\tau^2 \le v_2(\tau) \le \frac{1+\rho}{(1-\rho)^2}\tau^2 =: w_2(\tau). \tag{3.1}$$

The lower bound $z_2(\tau)$ follows from Schwartz' inequality and the upper bound $w_2(\tau)$ can be obtained by noting that $R(0) = 1 - \rho > 0$, which is the probability that an arriving customer in the corresponding M/G/1 FCFS queue does not need to wait. By property of Poisson Arrivals See Time Averages (PASTA), this probability equals the long run fraction of time the server is idle, thus $R(0) = 1 - \rho$. From (3.1) we note that the upper bound for the second moment is $100\rho\%$ larger than the lower bound, and these bounds only depend on the mean service time and do not depend on the second moment of the service time distribution.

By using the recursive formula (2.6) for $v_k(\tau)$ and 'ignoring' the alternating term $(-1)^j$, the following crude upper bound for all moments can be given:

$$v_k(\tau) \le k! \left(\frac{(e-1)\tau}{1-\rho} \right)^k, \tag{3.2}$$

also see Zwart [27]. As a consequence of this bound, it holds that the tail probability $\mathbb{P}(V(\tau) > x)$ is of order $o(e^{-\frac{1-\rho}{\gamma\tau}})$, with $\gamma > e - 1$, and the sojourn time $V(\tau)$ is always light-tailed conditional upon its service requirement. Intuitively it supports the conjecture that a large sojourn time is not due to excessive behavior of other customers present in the system.

The crude bound (3.2) for the second moment ($k = 2$) is always worse than the upper bound $w_2(\tau)$ given in (3.1), since $1 + \rho < 2 < 2!(e-1)^2$. Furthermore, for $\rho \to 0$, we have the attractive property that: $w_2(\tau) - z_2(\tau) \downarrow 0$. We generalize this result (3.1) for *all* moments, by using the results in [26] and using a generalized idea from [1] to obtain (tight) bounds. It turns out that these bounds have a similar and special structure.

Theorem 3.1 For all $k \ge 2$, there exist non-negative constants $c_i^k \in \mathbb{N}_0$ such that $v_k(\tau)$ is bounded by

$$\frac{1}{(1-\rho)^k}\tau^k \le v_k(\tau) \le \frac{\phi_{k-1}(\rho)}{(1-\rho)^k}\tau^k, \tag{3.3}$$

where $\phi_{k-1}(\rho) = \sum_{i=0}^{k-1} c_i^k \rho^i$ is a polynomial in ρ of degree $k - 1$ (if k even) or $k - 2$ (if k odd) and $c_0^k = 1$.

Proof. The lower bound in (3.3) follows by applying Jensen's inequality, since the map $f(x) = x^k$ is a convex function on $[0, \infty)$ for all $k \ge 1$. By noting that $1 - \rho \le R(x) \le 1$ and using (2.4), it is readily verified that $(1-\rho)^n \le R^{n*}(x) \le 1$. Therefore, by using (2.3) we obtain bounds for $\alpha_n(\tau)$:

$$\frac{\tau^n}{1-\rho} \le \alpha_n(\tau) \le \frac{\tau^n}{(1-\rho)^n}, \quad \text{for } n \ge 1, \text{ and } \alpha_0(\tau) := 1. \tag{3.4}$$

By doing simple algebraic manipulations using the recursive formula (2.6), the structure for the upper bound in (3.3) can be obtained. To this end, rewrite $\bar{v}_k(\tau) := (1-\rho)^k v_k(\tau)/\tau^k$, $\bar{\alpha}_j(\tau) := (1-\rho)^j \alpha_j(\tau)/\tau^j$, and the recursive formula (2.6) as $\bar{v}_k(\tau) = -\sum_{j=1}^{k} \binom{k}{j} \bar{v}_{k-j}(\tau)\bar{\alpha}_j(\tau)(-1)^j$. The bounds for $\bar{\alpha}_j(\tau)$ are given by

$$(1-\rho)^{j-1} \le \bar{\alpha}_j(\tau) \le 1, \quad \text{for all } j \ge 1,$$

5

and assume by induction hypothesis that the following bounds holds for $\overline{v}_{k-1}(\tau), \overline{v}_{k-2}(\tau), \dots$:

$$1 \leq \overline{v}_{k-j}(\tau) \leq \phi_{k-j-1}(\rho), \quad \text{for all } j = 1, \dots, k,$$

hence

$$(1 - \rho)^{j-1} \leq \overline{v}_{k-j}(\tau)\overline{\alpha}_j(\tau) \leq \phi_{k-j-1}(\rho), \quad \text{for all } j = 1, \dots, k.$$

We will apply induction and take into account the alternating term $(-1)^j$, hence we need both upper and lower bounds for the product $\overline{v}_{k-j}(\tau)\overline{\alpha}_j(\tau)$. Also, use the fact that

$$(1 - \rho)^{j-1} = 1 + \sum_{i=1}^{j-1} (-1)^i \binom{j-1}{i} \rho^i,$$

by Newton's binomial theorem. Summarizing, use the bounds for $\overline{v}_{k-1}(\tau)\overline{\alpha}_1(\tau), \overline{v}_{k-2}(\tau)\overline{\alpha}_2(\tau), \dots$:

$$1 + \sum_{i=1}^{j-1} (-1)^i \binom{j-1}{i} \rho^i \leq \overline{v}_{k-j}(\tau)\overline{\alpha}_j(\tau) \leq 1 + \sum_{i=1}^{k-j-1} c_i^{k-j} \rho^i, \quad \text{for all } j = 1, \dots, k,$$

by induction hypothesis, to obtain the upper bound for $\overline{v}_k(\tau)$, for $k \geq 2$. Bounds for $k = 1$ are trivial and exact, and by applying induction in (3.5) we obtain:

$$
\begin{aligned}
\overline{v}_k(\tau) &= \sum_{\substack{j=1 \\ j:\,\mathrm{odd}}}^{k} \binom{k}{j} \overline{v}_{k-j}(\tau)\overline{\alpha}_j(\tau) - \sum_{\substack{j=2 \\ j:\,\mathrm{even}}}^{k} \binom{k}{j} \overline{v}_{k-j}(\tau)\overline{\alpha}_j(\tau) \\
&\leq \sum_{\substack{j=1 \\ j:\,\mathrm{odd}}}^{k} \binom{k}{j} \left\{ 1 + \sum_{i=1}^{k-j-1} c_i^{k-j} \rho^i \right\} - \sum_{\substack{j=2 \\ j:\,\mathrm{even}}}^{k} \binom{k}{j} \left\{ 1 + \sum_{i=1}^{j-1} (-1)^i \binom{j-1}{i} \rho^i \right\} \qquad (3.5) \\
&= 1 + \sum_{\substack{j=1 \\ j:\,\mathrm{odd}}}^{k} \binom{k}{j} \left\{ \sum_{i=1}^{k-j-1} c_i^{k-j} \rho^i \right\} - \sum_{\substack{j=2 \\ j:\,\mathrm{even}}}^{k} \binom{k}{j} \left\{ \sum_{i=1}^{j-1} (-1)^i \binom{j-1}{i} \rho^i \right\} \equiv \sum_{i=0}^{k-1} c_i^k \rho^i, \quad (3.6)
\end{aligned}
$$

since $\sum_{\substack{j=1 \\ j:\,\mathrm{odd}}}^{k} \binom{k}{j} - \sum_{\substack{j=2 \\ j:\,\mathrm{even}}}^{k} \binom{k}{j} = \sum_{j=1}^{k} \binom{k}{j}(-1)^{j+1} = 1$, hence $c_0^k = 1$; and (3.6) holds by definition of the coefficients c_i^k. Also note that

$$\sum_{\substack{j=1 \\ j:\,\mathrm{odd}}}^{k} \left\{ \sum_{i=1}^{k-j-1} c_i^{k-j} \rho^i \right\} \text{ is a polynomial in } \rho \text{ of (at most) degree } k - 2, \text{ and}$$

$$\sum_{\substack{j=2 \\ j:\,\mathrm{even}}}^{k} \left\{ \sum_{i=1}^{j-1} (-1)^i \binom{j-1}{i} \rho^i \right\} \text{ is of degree } k - 1 \text{ (if } k \text{ even), and of degree } k - 2 \text{ (if } k \text{ odd).}$$

By comparing the terms in (3.6) it is also not difficult to see that $c_{k-1}^k = 1$ if k even, and $c_{k-1}^k = 0$ if k odd. Furthermore, it can be shown that $c_i^k \geq 0$. However, for the existence of an upper bound of the described structure, it is not necessary since if $c_i^k < 0$ we may simply set c_i^k to an arbitrary non-negative number. $\blacksquare$

Remark 3.2 We refer to the 'best possible' (insensitive) upper bound for $v_k(\tau)$ of the form given in (3.3) with only knowledge of τ and ρ, as the upper bound that has the polynomial $\phi^*_{k-1}(\rho)$ with the 'smallest possible' coefficients c_i^k, such that $(1-\rho)^k v_k(\tau)/\tau^k \leq \phi^*_{k-1}(\rho)$ holds for all $\tau \geq 0$ and for all $\rho < 1$.

Remark 3.3 In principle, we can apply the 'alternating' procedure to obtain a lower bound as well. However, the resulting lower bound is always worse than the Jensen's lower bound.

For the second moment, we obtain $c_1^2 = 1$ since $k = 2$ is even, and the result is the same as in [1], see (3.1). As a consequence of Theorem 3.1 we have the following Corollary 3.4.

Corollary 3.4 For all $k \geq 1$, and $\tau > 0$,

$$v_k(\tau) < \infty \Leftrightarrow \rho < 1.$$

Proof. The '$\Rightarrow$' is trivial. Since the sojourn time of a customer is always at least its service requirement, then if $v_k(\tau)$ is finite for arbitrary $k \geq 1$, then the unconditional moment $\mathbb{E}V^k$ is also finite, and hence $\beta_k < \infty$ and $\rho < 1$. The '$\Leftarrow$' follows from Theorem 3.1. ∎

The distribution of $V(\tau)$ is not insensitive to the service time distribution, since the expression for the second moment $v_2(\tau) = 2\tau^2/(1-\rho)^2 - \alpha_2(\tau)$ depends on the service time distribution through $\alpha_2(\tau)$ and the waiting time distribution $R(x)$ in the M/G/1 FCFS queue. However, Corollary 3.4 implies that the sensitivity of the service time distribution on $V(\tau)$ is always limited whenever $\rho < 1$, confirming that PS can be considered as a *fair* policy. Similar results holds for unconditional moments of the sojourn time, whenever they exist.

Corollary 3.5 The k-th moment of the unconditional sojourn time V is upper bounded by:

$$\mathbb{E}V^k = \int_0^\infty v_k(\tau) dB(\tau) \leq \frac{\phi_{k-1}(\rho)}{(1-\rho)^k} \int_0^\infty \tau^k dB(\tau) = \frac{\phi_{k-1}(\rho)}{(1-\rho)^k} \beta_k.$$

This results implies that

$$\mathbb{E}V^k < \infty \Leftrightarrow \beta_k < \infty,$$

which indicates that the tail behavior of the unconditional sojourn time distribution and the service time distribution are similar. This tail equivalence result has been obtained in [26].

4 The instantaneous sojourn time

In this section we study the behavior of $V(\tau)$ and $v_k(\tau)$ as $\tau \to 0$. The key idea is as follows. A customer with a very small job size arrives at the system in steady state, say at time t_0. The tagged customer sees n other customers upon arrival with probability p_n, and by the PASTA property: $p_n = \pi_n = (1-\rho)\rho^n$. Let the remaining service requirements of the n other customers at time t_0 be x_i, $i = 1, ..., n$. The key assumption of the instantaneous sojourn time analysis is that at time point t_0, the tagged customer requires a very small job size $\tau \ll \min_{i=1,...,n} x_i$. Furthermore, we assume that τ is small enough such that no other customers arrive during time interval $[t_0, t_0 + (n+1)\tau)$.

Under these assumptions, the next event after time point t_0 is the service completion of the same short job (as if the tagged customer arrived at a system with n *permanent* customers with probability π_n and with no other arriving customers). The tagged customer receives service at

rate $1/(n+1)$ during its whole stay in the system. We define the *instantaneous* sojourn time as $\widehat{V}(\tau) = (N+1)\tau$, when the tagged customer sees N other customers upon arrival, with N distributed as $\mathbb{P}(N = n) = \pi_n$. The k-th moment of the true sojourn time can be approximated with the k-th moment of the *instantaneous* sojourn time $\widehat{v}_k(\tau)$ if $\tau \to 0$:

$$v_k(\tau) \approx \widehat{v}_k(\tau) := \mathbb{E}\widehat{V}(\tau)^k = \sum_{n=0}^{\infty} \pi_n \left\{(n+1)\tau\right\}^k, \quad k \in \mathbb{N}.$$

4.1 The moments of the instantaneous sojourn time

The next theorem shows that $(1-\rho)^k \widehat{v}_k(\tau)/\tau^k$ can be expressed as a polynomial in ρ of (exact) degree $k-1$, where the coefficients are so-called Eulerian numbers $\left\langle {k \atop j} \right\rangle$, which counts the total number of permutations of the elements in the set $\{1, ..., k\}$ having j permutation ascents; see Appendix for more definitions and facts about Eulerian numbers.

Theorem 4.1 The k-moment of the instantaneous sojourn time equals

$$\widehat{v}_k(\tau) = \frac{\tau^k}{(1-\rho)^k} \sum_{j=0}^{k-1} \left\langle {k \atop j} \right\rangle \rho^j, \quad \text{for } k = 1, 2, ...,$$

where the coefficients $\left\langle {k \atop j} \right\rangle$ are Eulerian numbers.

    Proof. Using the identity (see Appendix)

$$\sum_{k=1}^{\infty} k^n r^k \equiv \frac{1}{(1-r)^{n+1}} \sum_{i=0}^{n} \left\langle {n \atop i} \right\rangle r^{n-i} = \frac{r}{(1-r)^{n+1}} \sum_{i=0}^{n-1} \left\langle {n \atop i} \right\rangle r^{n-i-1},$$

where $\left\langle {n \atop i} \right\rangle$ are Eulerian numbers for $i = 0, 1, ..., n-1$ (note that: $\left\langle {n \atop n} \right\rangle = 0$), we readily derive:

$$\widehat{v}_k(\tau)/\tau^k = \sum_{n=0}^{\infty} \pi_n (n+1)^k = \frac{1-\rho}{\rho} \sum_{m=1}^{\infty} m^k \rho^m = \frac{\sum_{i=0}^{k-1} \left\langle {k \atop i} \right\rangle \rho^{k-i-1}}{(1-\rho)^k} = \frac{\sum_{j=0}^{k-1} \left\langle {k \atop j} \right\rangle \rho^j}{(1-\rho)^k}, \qquad (4.1)$$

where the last equality sign in (4.1) uses the fact $\left\langle {n \atop m} \right\rangle = \left\langle {n \atop n-m-1} \right\rangle$, for $0 \leq m \leq n-1$, by symmetry of the Euler's number triangle; see Appendix. ■

4.2 The instantaneous sojourn time and Euler's number triangle

For $\tau \to 0$ we obtained the approximations $v_k(\tau) \approx \widehat{v}_k(\tau)$. In particular, from the values in the Euler's number triangle we obtain the approximations for the first four moments:

$$v_1(\tau) \approx \widehat{v}_1(\tau) = \frac{\tau}{1-\rho}, \quad v_2(\tau) \approx \widehat{v}_2(\tau) = \frac{1+\rho}{(1-\rho)^2}\tau^2,$$

$$v_3(\tau) \approx \widehat{v}_3(\tau) = \frac{1+4\rho+\rho^2}{(1-\rho)^3}\tau^3, \quad v_4(\tau) \approx \widehat{v}_4(\tau) = \frac{1+11\rho+11\rho^2+\rho^3}{(1-\rho)^4}\tau^4.$$

    Surprisingly, the instantaneous sojourn time approximation yields an exact expression for the first moment, and the approximation for the second moment yields an upper bound. This might suggest that the moments of the instantaneous sojourn time $\widehat{v}_k(\tau)$ are upper bounds for

$v_k(\tau)$ as well for all $k \geq 3$. In Section 5, we will formally prove that indeed $v_k(\tau) \leq \widehat{v}_k(\tau)$ for all $\tau \geq 0$ and $k \in \mathbb{N}$; see Theorem 5.14 and Theorem 5.10. An intuitive explanation is given in the next Remark 4.2.

Remark 4.2 With the instantaneous sojourn time analysis ($\tau \to 0$) we assume that during an interval of length $(n+1)\tau$ there is no other activity in the system. This is not very likely in the real system, especially when n is large upon arrival. The instantaneous sojourn time overestimates the true sojourn time when n is large upon arrival of the tagged customer, and underestimates the true sojourn time when n is small upon arrival. Apparently, for the first moment: over- and under estimation outweigh each other (weighted with probability π_n). For higher moments: overestimation is weighted more heavily than underestimation, since the queue length process $N(t)$ shows a negative drift for a large initial value of the number of customers present in the system. To this end, note that we assume $\rho < 1$.

In [26] the following heavy traffic result has been derived: $\lim_{\rho \to 1} \mathbb{E}\left[(1-\rho)V(\tau)\right]^k = k!\tau^k$. For the instantaneous sojourn time we arrive at the following heavy traffic result for the moments.

Corollary 4.3
$$\lim_{\rho \to 1}(1-\rho)^k\widehat{v}_k(\tau) = k!\tau^k. \tag{4.2}$$

Proof. Use the fact that $\sum_{j=0}^{k-1}\left\langle{k \atop j}\right\rangle = k!$, i.e., the total number of permutations of the elements in the set $\{1,...,k\}$ equals $k!$. ∎

For $\rho = 0$ we have the equality $\widehat{v}_k(\tau) = v_k(\tau) = \tau^k$ (deterministic), and for $\rho \to 1$ we have that (4.2) is similar to the heavy-traffic result in [26]. This qualitative result and Remark 4.2 suggest that $\widehat{v}_k(\tau)$ are tight upper bounds of a simple and special form, only depending on τ and on the first moment of the service time (through ρ). Also note that $k!\tau^k/(1-\rho)^k$, $k = 1, 2, ...$, is the moment sequence of an exponentially distributed random variable with mean $\tau/(1-\rho)$. From [25] we know that for fixed $\tau > 0$:

$$\mathbb{P}(V(\tau)(1-\rho)/\tau \leq x) \to 1 - e^{-x}, \quad \text{as } \rho \uparrow 1, \ x \geq 0, \tag{4.3}$$

i.e., $V(\tau)$ converges in distribution to an exponentially distributed random variable, properly scaled and as $\rho \uparrow 1$. These observations suggest that $V(\tau)$ is stochastic ordered in *convex order*. Informally stated: $V(\tau)$ is always less variable than an exponential random variable with the same mean $\tau/(1-\rho)$.

In the next section we obtain more precise stochastic ordering results together with the formal proof that the instantaneous sojourn time moments are upper bounds for $v_k(\tau)$, for all $\tau \geq 0$ and $k \in \mathbb{N}$, with Eulerian numbers as coefficients for the polynomials.

5 Stochastic ordering

In this section we give some new results for the distribution of $V(\tau)$ in relation with stochastic ordering theory. For stochastic ordering theory we refer to Stoyan [18], and Shaked & Shanthikumar [17]. The main goal of this Section 5 is to prove that the moments of the instantaneous sojourn time serve as upper bound for the moments of the conditional sojourn time.

In Section 5.1 we first establish a Laplace transform ordering between $V(\tau)$ and the instantaneous sojourn time $\widehat{V}(\tau)$. In addition, in Section 5.1 a characterization that the distribution $V(\tau)$ belongs to the so-called $\mathcal{L}$-class of life time distributions will be derived, which is related to

the Laplace transform ordering. In Section 5.2 we finally arrive at the moment ordering result between $V(\tau)$ and $\widehat{V}(\tau)$, i.e., $\mathbb{E}V(\tau)^k \leq \mathbb{E}\widehat{V}(\tau)^k$ for all $\tau \geq 0$ and $k \in \mathbb{N}$.

5.1 Laplace transform ordering

The distribution of $V(\tau)$ can be stochastically ordered in Laplace transforms as a consequence of Theorem 5.1.

Theorem 5.1 For all $\mathrm{Re}\, s \geq 0$, the LST $v(s, \tau)$ of the random variable $V(\tau)$ is bounded by

$$z(s, \tau) \leq v(s, \tau) \leq w(s, \tau), \tag{5.1}$$

with $z(s, \tau) := e^{-s\tau/(1-\rho)}$, and $w(s, \tau) := \frac{1-\rho}{e^{s\tau}-\rho}$.

    Proof. Use the expression for $v(s, \tau)$ given by (2.2), and the bounds for $\alpha_n(\tau)$ given in (3.4). From $v(s, \tau)^{-1} = \sum_{n=0}^{\infty} \frac{s^n}{n!}\alpha_n(\tau)$, $\alpha_0(\tau) = 1$, and $\frac{\tau^n}{1-\rho} \leq \alpha_n(\tau) \leq \frac{\tau^n}{(1-\rho)^n}$ for $n \geq 1$, we get

$$w(s, \tau)^{-1} = 1 + \sum_{n=1}^{\infty} \frac{s^n}{n!}\frac{\tau^n}{1-\rho} \leq v(s, \tau)^{-1} \leq \sum_{n=0}^{\infty} \frac{s^n}{n!}\frac{\tau^n}{(1-\rho)^n} = z(s, \tau)^{-1}.$$

∎

    Note that the lower bound $z(s, \tau)$ is the LST of the constant value $\tau/(1 - \rho)$. This bound has a form similar to Jensen's lower bound $v_k(\tau) \geq v_1(\tau)^k$. The upper bound $w(s, \tau)$ is the LST of the instantaneous sojourn time $\widehat{V}(\tau) = (N + 1)\tau$, since for $\mathrm{Re}\, s \geq 0$ and $\tau \geq 0$

$$\mathbb{E}\left[e^{-s(N+1)\tau}\right] = \sum_{n=0}^{\infty} e^{-s(n+1)\tau}(1-\rho)\rho^n = \frac{1-\rho}{e^{s\tau}-\rho} = w(s, \tau), \text{ if } \rho e^{-s\tau} \leq \rho < 1.$$

    This observation suggests that the distribution of $V(\tau)$ is in certain sense 'in between' a deterministic and an exponential random variable with the same means. A more precise characterization follows in Theorem 5.4. However, the stochastic ordering in Laplace transform is in general a weak relation. From the Laplace transforms ordering: $z(s, \tau) \leq v(s, \tau) \leq w(s, \tau)$, denoted as $\frac{\tau}{1-\rho} \geq_{Lt} V(\tau) \geq_{Lt} \widehat{V}(\tau)$, the only implication is the ordering of the first moments $\mathbb{E}(\frac{\tau}{1-\rho}) \geq \mathbb{E}V(\tau) \geq \mathbb{E}\widehat{V}(\tau)$, cf. [18]. Implications for higher moments cannot be made in general for stochastic ordering in Laplace transforms. If in addition, equality in means is given, we obtain the following second moment ordering result.

Theorem 5.2 For any random variables Z, V, W with respectively the Laplace-Stieltjes transforms $z(s), v(s), w(s)$, such that $z(s) \leq v(s) \leq w(s)$, for all $\mathrm{Re}\, s \geq 0$, and if $\mathbb{E}Z = \mathbb{E}V = \mathbb{E}W$, then it holds that

$$\mathbb{E}Z^2 \leq \mathbb{E}V^2 \leq \mathbb{E}W^2.$$

    Proof. The first moment $\mathbb{E}V$ equals $-\lim_{s\to 0} \frac{d}{ds}v(s)$, i.e., the slope (with a minus sign) of the tangent line of the LST $v(s)$ at $s = 0$, and a LST is equivalent to a completely monotonic function with value 1 at the function argument 0 (Bernstein's Theorem, cf. [21]). Observe that the tangent line of $v(s)$ at $s = 0$ is equal to the tangent line of $w(s)$ at $s = 0$, by $\mathbb{E}V = \mathbb{E}W$. Then, by convexity and analyticity of completely monotonic functions, and the ordering $v(s) \leq w(s)$, it is readily seen that $\frac{d^2}{ds^2}v(s) \leq \frac{d^2}{ds^2}w(s)$ for s in a neighborhood of 0, and thus: $\lim_{s\to 0} \frac{d^2}{ds^2}v(s) = \mathbb{E}V^2 \leq \mathbb{E}W^2$. ∎

For the M/G/1 PS case, all corresponding means in (5.1) equal $\tau/(1-\rho)$, hence we have again that $z_2(\tau) \leq v_2(\tau) \leq w_2(\tau) = (1+\rho)\tau^2/(1-\rho)^2$. We have even a stronger stochastic ordering result (see Theorem 5.4), which states that $V(\tau)$ belongs to the so-called $\mathcal{L}$-class of life time distributions (see Klefsjö [8]).

Definition 5.3 It is said that $V(\tau)$ belongs to the $\mathcal{L}$-class of life time distributions if for some *exponentially* distributed random variable $X(\tau)$, with $\mathbb{E}X(\tau) = \mathbb{E}V(\tau)$, that the Laplace transform ordering $V(\tau) \geq_{Lt} X(\tau)$ holds, i.e., for all $\mathrm{Re}\, s \geq 0 : v(s,\tau) \leq x(s,\tau)$ where $x(s,\tau)$ is the LST of $X(\tau)$.

Theorem 5.4 $V(\tau) \in \mathcal{L}$ if $\rho < 1$.
    Proof. From Theorem 5.1, we can further upper bound the LST by

$$v(s,\tau) \leq \frac{1-\rho}{e^{s\tau}-\rho} \leq \frac{1-\rho}{1+s\tau-\rho} = \frac{1}{1+s\tau/(1-\rho)} =: x(s,\tau).$$

Clearly, $x(s,\tau)$ is the LST of an exponential distribution with mean $\tau/(1-\rho)$. ∎

Distributions belonging to the $\mathcal{L}$-class of life time distributions always have a finite second moment, and the coefficient of variation is not greater than one; see e.g. [2, 12]. For $V(\tau)$ in the M/G/1 PS queue, we have in fact that the coefficient of variation is less or equal to $\sqrt{\rho}$, cf. (3.1). More interestingly, although the $\mathcal{L}$-class is a wide class of distributions, Klar [7] obtained explicit and sharp 'reliability bounds' for any $\mathcal{L}$-class distribution. Applying these reliability bounds (Theorem 4.1 from [7]), for the sojourn time distribution in the M/G/1 PS queue, we obtain simple explicit and insensitive bounds for the tail probabilities $\mathbb{P}(V(\tau) > x)$.

Corollary 5.5 For $x \leq \tau/(1-\rho)$ we have the insensitive lower bound

$$\mathbb{P}(V(\tau) > x) \geq 1 - \frac{1}{(x(1-\rho)/\tau)^2 - 2x(1-\rho)/\tau + 2},$$

and this lower bound is sharp for $2 - \sqrt{2} \leq x(1-\rho)/\tau \leq 1$.
And for $x > \tau/(1-\rho)$ we have the insensitive upper bound

$$\mathbb{P}(V(\tau) > x) \leq \frac{1}{(x(1-\rho)/\tau)^2 - 2x(1-\rho)/\tau + 2},$$

which is sharp if $x(1-\rho)/\tau \geq 2 + \sqrt{2}$.
    Proof. Direct application of Theorem 4.1 from [7]. ∎

Remark 5.6 Stronger results for the reliability bounds exist for life time distributions belonging to subclasses of the $\mathcal{L}$-class.

In the remainder of this paper, our main goal is to provide the 'best possible' upper bounds for $v_k(\tau)$ in the sense of Remark 3.2. By Theorem 3.1, existence of an upper bound of the form $(1-\rho)^k v_k(\tau)/\tau^k \leq 1 + \sum_{i=1}^{k-1} c_i^k \rho^i$, is guaranteed for some constants $c_i^k \in \mathbb{N}_0$. As in the proof of Theorem 3.1, we can calculate the coefficients c_i^k by bounding the product $v_{k-j}(\tau)\alpha_j(\tau)$ in the recursive formula (2.6) term by term, for $j = 1, ..., k$, in a recursive manner and by alternated use of lower and upper bounds for $v_{k-j}(\tau)$ and $\alpha_j(\tau)$. However, due to the alternated term $(-1)^j$ in the recursive formula, this procedure of term by term bounding does not provide us the 'best possible' upper bound.

The coefficients $\alpha_j(\tau)$ for $j = 2, 3, \ldots$ are not independent of each other. For example, if $\alpha_j(\tau)$ is close to its lower bound $\tau^j/(1-\rho)$, then $\alpha_{j+1}(\tau)$ is generally not close to its upper bound $\tau^{j+1}/(1-\rho)^{j+1}$. In fact, if $\alpha_j(\tau) = \tau^j/(1-\rho)$ for some $j \geq 2$, then necessarily $\alpha_k(\tau) = \tau^k/(1-\rho)$ for all $k \geq 2$. In addition, if it holds that $\alpha_j(\tau) = \tau^j/(1-\rho)^j$ for some $j \geq 2$, then necessarily $\alpha_k(\tau) = \tau^k/(1-\rho)^k$ for all $k \geq 2$. In other words: the 'alternating' procedure of term by term bounding in a recursive manner is too conservative.

It is not immediately clear that the moments $\widehat{v}_k(\tau)$ of the instantaneous sojourn time are upper bounds for the moments $v_k(\tau)$. The main difficulty in proving (upper) bounds is caused by the alternating term $(-1)^j$ in the recursive formula. On the other hand, it is also due to the presence of the alternating term $(-1)^j$ that the obtained bounds are tight.

Via a different construction, we will formally prove that the moments of the instantaneous sojourn time are upper bounds for the moments of the conditional sojourn time. This will be done by comparing two different but related PS queues, and both queues are constructed with a random number of *permanent* customers. The general moment ordering result, that compares the sojourn times of the two related PS queues, will be established in the next Section 5.2.

5.2 Moment ordering

In this section, we will prove our main result that $v_k(\tau) \leq (1 + \sum_{i=1}^{k-1} \left\langle {k \atop i} \right\rangle \rho^i)/[(1-\rho)/\tau]^k$, which is identical to the moment ordering $\mathbb{E}V(\tau)^k \leq \mathbb{E}\widehat{V}(\tau)^k$ for all $k \geq 1$; see Theorem 5.14. This moment ordering result follows from a more general moment ordering result between the random variables $V_{\lambda_1}(\tau)^{(N_2+1)*}$ and $V_{\lambda_2}(\tau)^{(N_1+1)*}$. These random variables will be defined for PS queues with a random number of *permanent* customers. The general moment ordering result between these new defined random variables: $\mathbb{E}(V_{\lambda_1}(\tau)^{(N_2+1)*})^k \leq \mathbb{E}(V_{\lambda_2}(\tau)^{(N_1+1)*})^k$, for all $k \geq 1$, is only valid for a certain range of values of λ_1 and λ_2; see Theorem 5.10.

Definition 5.7 For random variables X and Y, X is said to be smaller than Y in *moments* order, denoted by $X \leq_{\mathrm{mom}} Y$, if and only if $\mathbb{E}\phi(X) \leq \mathbb{E}\phi(Y)$, for all polynomial ϕ with non-negative coefficients; see [17].

We construct two independent processor-sharing queues as follows. Both M/G/1 PS queues have the same service time distribution, but with different arrival rates only, respectively with λ_1 and λ_2. Let $V_{\lambda_i}(\tau)$ denote the sojourn time in a $M_{\lambda_i}/G/1$ PS queue with arrival rate λ_i, and service time distribution $B(x)$ with mean $\mathbb{E}X$, for $i = 1, 2$.

Next, we construct two other related processor-sharing queues as follows. We define the random variable $V_{\lambda_1}(\tau)^{(j+1)*}$ as a random variable whose distribution is the $(j+1)$-fold convolution of the distribution of $V_{\lambda_1}(\tau)$. This can be interpreted as the sojourn time in the same $M_{\lambda_1}/G/1$ PS model with j *permanent* customers; see e.g. [1].

Next, the random variable $V_{\lambda_1}(\tau)^{(N_2+1)*}$ is the random variable that is in distribution equal to $V_{\lambda_1}(\tau)^{(j+1)*}$ with probability $\mathbb{P}(N_2 = j)$, for $j \in \mathbb{N}_0$. Hence, $V_{\lambda_1}(\tau)^{(N_2+1)*}$ can be interpreted as the sojourn time in the $M_{\lambda_1}/G/1$ PS model with an additional random number of permanent customers in the system, with the number of permanent customers distributed as N_2.

Finally, the random variable $V_{\lambda_2}(\tau)^{(N_1+1)*}$ is defined analogously with the same service time distribution, and the distributions of N_1 and N_2 are defined by the geometric probabilities:

$$\mathbb{P}(N_1 = i) = \frac{1-\rho}{1-\rho_2}\left(\frac{\rho_1}{1-\rho_2}\right)^i, \text{ and } \mathbb{P}(N_2 = j) = \frac{1-\rho}{1-\rho_1}\left(\frac{\rho_2}{1-\rho_1}\right)^j, \tag{5.2}$$

for $i \in \mathbb{N}_0$, $j \in \mathbb{N}_0$, $\rho_1 = \lambda_1 \mathbb{E}X$, $\rho_2 = \lambda_2 \mathbb{E}X$, and $\rho := \rho_1 + \rho_2 < 1$. The probabilities $\mathbb{P}(N_1 = i)$ and $\mathbb{P}(N_2 = j)$ are in fact the marginal queue length probabilities of a two-class egalitarian processor-sharing queue (with two types of customer classes); see e.g. [4].

It is not difficult to show that $\mathbb{E}V_{\lambda_2}(\tau)^{(N_1+1)*} = \mathbb{E}V_{\lambda_1}(\tau)^{(N_2+1)*} = \tau/(1-\rho)$. In general, we will prove that the following moment ordering holds: $V_{\lambda_1}(\tau)^{(N_2+1)*} \leq_{\text{mom}} V_{\lambda_2}(\tau)^{(N_1+1)*}$, if the values for λ_1 and λ_2 satisfies $\lambda_1 \geq \lambda_2$; see Theorem 5.10. By special choice of the value λ_2, this result will enable us to prove that $\mathbb{E}V(\tau)^k \leq \mathbb{E}\widehat{V}(\tau)^k$ for all $k \geq 1$ and $\tau \geq 0$; see Theorem 5.14. First we derive the LST's of the random variables $V_{\lambda_1}(\tau)^{(N_2+1)*}$ and $V_{\lambda_2}(\tau)^{(N_1+1)*}$.

Lemma 5.8 The LST's for the random variables $V_{\lambda_1}(\tau)^{(N_2+1)*}$ and $V_{\lambda_2}(\tau)^{(N_1+1)*}$ are given by

$$\widetilde{v}(s,\tau) := \mathbb{E}\left(e^{-sV_{\lambda_1}(\tau)^{(N_2+1)*}}\right) = \left(\sum_{n=0}^{\infty} \frac{s^n}{n!}\widetilde{\alpha}_n(\tau)\right)^{-1}, \quad \mathrm{Re}\, s \geq 0, \tag{5.3}$$

$$\widetilde{w}(s,\tau) := \mathbb{E}\left(e^{-sV_{\lambda_2}(\tau)^{(N_1+1)*}}\right) = \left(\sum_{n=0}^{\infty} \frac{s^n}{n!}\widetilde{\beta}_n(\tau)\right)^{-1}, \quad \mathrm{Re}\, s \geq 0, \tag{5.4}$$

where $\widetilde{\alpha}_0(\tau) = \widetilde{\beta}_0(\tau) = 1$, $\widetilde{\alpha}_1(\tau) = \widetilde{\beta}_1(\tau) = \tau/(1-\rho)$, and for $n \geq 2$:

$$\widetilde{\alpha}_n(\tau) = \frac{n}{1-\rho}\sum_{m=0}^{\infty}\binom{m+n-2}{n-2}\rho_1^m\int_{x=0}^{\tau}(\tau-x)^{n-1}\widetilde{B}^{m*}(x)dx, \tag{5.5}$$

$$\widetilde{\beta}_n(\tau) = \frac{n}{1-\rho}\sum_{m=0}^{\infty}\binom{m+n-2}{n-2}\rho_2^m\int_{x=0}^{\tau}(\tau-x)^{n-1}\widetilde{B}^{m*}(x)dx. \tag{5.6}$$

Note that the only difference between the expressions in $\widetilde{\alpha}_n(\tau)$ and $\widetilde{\beta}_n(\tau)$ is the term ρ_1 or ρ_2.

Proof. The LST's are readily expressed by

$$\widetilde{v}(s,\tau) = \sum_{j=0}^{\infty}\{v_1(s,\tau)\}^{j+1}\mathbb{P}(N_2=j) = \frac{(1-\rho)\,v_1(s,\tau)}{1-\rho_1-\rho_2 v_1(s,\tau)},$$

$$\widetilde{w}(s,\tau) = \sum_{i=0}^{\infty}\{v_2(s,\tau)\}^{i+1}\mathbb{P}(N_1=i) = \frac{(1-\rho)\,v_2(s,\tau)}{1-\rho_1 v_2(s,\tau)-\rho_2},$$

where $v_i(s,\tau)$ is the LST of $V_{\lambda_i}(\tau)$ of an ordinary $M_{\lambda_i}/G/1$ PS queue, i.e., $v_1(s,\tau)^{-1} = \sum_{n=0}^{\infty}\frac{s^n}{n!}\alpha_n(\tau)$, and $v_2(s,\tau)^{-1} = \sum_{n=0}^{\infty}\frac{s^n}{n!}\beta_n(\tau)$, where the coefficients $\alpha_n(\tau)$ and $\beta_n(\tau)$ are similar to (2.3). Alternatively, we can rewrite the LST's as

$$\widetilde{v}(s,\tau) = \left(\frac{1-\rho_1}{1-\rho}v_1(s,\tau)^{-1} - \frac{\rho_2}{1-\rho}\right)^{-1} = \left(\sum_{n=0}^{\infty}\frac{s^n}{n!}\widetilde{\alpha}_n(\tau)\right)^{-1},$$

$$\widetilde{w}(s,\tau) = \left(\frac{1-\rho_2}{1-\rho}v_2(s,\tau)^{-1} - \frac{\rho_1}{1-\rho}\right)^{-1} = \left(\sum_{n=0}^{\infty}\frac{s^n}{n!}\widetilde{\beta}_n(\tau)\right)^{-1},$$

where $\widetilde{\alpha}_0(\tau) = 1, \widetilde{\beta}_0(\tau) = 1$, and for $n \geq 1$:

$$\widetilde{\alpha}_n(\tau) = \frac{1-\rho_1}{1-\rho}\alpha_n(\tau), \; \widetilde{\beta}_n(\tau) = \frac{1-\rho_2}{1-\rho}\beta_n(\tau).$$

More explicitly, the coefficients can be expressed as $\widetilde{\alpha}_1(\tau) = \widetilde{\beta}_1(\tau) = \tau/(1-\rho)$, since $\alpha_1(\tau) = \tau/(1-\rho_1)$, $\beta_1(\tau) = \tau/(1-\rho_2)$, and for $n \geq 2$: $\widetilde{\alpha}_n(\tau)$ and $\widetilde{\beta}_n(\tau)$ can be expressed as (5.5) and (5.6), which follows by combining the expressions given by (2.3) and (2.5) for an ordinary $M_{\lambda_i}/G/1$ PS queue. $\blacksquare$

As a direct consequence of Lemma 5.8, the moments of the random variables $V_{\lambda_1}(\tau)^{(N_2+1)*}$ and $V_{\lambda_2}(\tau)^{(N_1+1)*}$ satisfy a similar recursion as for an ordinary M/G/1 PS queue.

Corollary 5.9 For all $\tau \geq 0$, the moments of $V_{\lambda_1}(\tau)^{(N_2+1)*}$ and $V_{\lambda_2}(\tau)^{(N_1+1)*}$, defined by respectively $\widetilde{v}_k(\tau) = \mathbb{E}\left\{V_{\lambda_1}(\tau)^{(N_2+1)*}\right\}^k$ and $\widetilde{w}_k(\tau) = \mathbb{E}\left\{V_{\lambda_2}(\tau)^{(N_1+1)*}\right\}^k$, are recursively given by $\widetilde{v}_0(\tau) = \widetilde{w}_0(\tau) = 1$, and for $k \geq 1$:

$$\widetilde{v}_k(\tau) = -\sum_{j=1}^k \binom{k}{j} \widetilde{v}_{k-j}(\tau) \widetilde{\alpha}_j(\tau)(-1)^j,$$

$$\widetilde{w}_k(\tau) = -\sum_{j=1}^k \binom{k}{j} \widetilde{w}_{k-j}(\tau) \widetilde{\beta}_j(\tau)(-1)^j.$$

Theorem 5.10 For $\tau > 0$, and $\rho := \rho_1 + \rho_2 < 1$, the moment ordering

$$V_{\lambda_1}(\tau)^{(N_2+1)*} \leq_{\mathrm{mom}} V_{\lambda_2}(\tau)^{(N_1+1)*}$$

holds, if and only if $\rho_1 \geq \rho_2$ (equivalently $\lambda_1 \geq \lambda_2$).

Proof. First, we have the following observations and implications. If $\rho_1 = \rho_2$, then obviously it holds that $V_{\lambda_1}(\tau)^{(N_2+1)*} \stackrel{d}{=} V_{\lambda_2}(\tau)^{(N_1+1)*}$ and hence $\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau)$, for all $k \geq 1$ and all $\tau \geq 0$. It always holds that $\widetilde{v}_1(\tau) = \widetilde{w}_1(\tau) = \tau/(1-\rho)$ for all $\tau \geq 0$, irrespective of ρ_1 and ρ_2. Also note that the following statements are equivalent:

(i)   There exists a $\tau > 0$ such that $\widetilde{\alpha}_n(\tau) = \widetilde{\beta}_n(\tau)$, for some $n \geq 2$
(ii) $\rho_1 = \rho_2$
(iii) For all $\tau \geq 0$ and for all $n \geq 1$: $\widetilde{\alpha}_n(\tau) = \widetilde{\beta}_n(\tau)$
(iv)   For all $\tau \geq 0$: $\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau)$ for all $k \geq 1$

The equivalent statements can be seen as follows: (i) $\Rightarrow$ (ii) follows by comparing (5.5) and (5.6). These expressions are the same infinite polynomials (power series, generating function), $\mathbb{R}_+$-valued and with the same strictly positive coefficients for every $n \geq 2$ and for every $\tau > 0$, only evaluated at a different point, respectively at ρ_1 and ρ_2. A polynomial $\sum_{m=0}^{\infty} a_m x^m$, with strictly positive coefficients is increasing convex for $x \geq 0$ and strictly increasing. Thus, if it holds that $\sum_{m=0}^{\infty} a_m \rho_1^m = \sum_{m=0}^{\infty} a_m \rho_2^m$, then necessarily $\rho_1 = \rho_2$ by the fact of strict increasing. The implication (ii) $\Rightarrow$ (iii) is trivial, and (iii) $\Rightarrow$ (iv) follows by the corresponding recursive formula for the moments; see Corollary 5.9. Finally, (iv) $\Rightarrow$ (i) is also trivial since (i) holds in particular for $n = 2$ for some $\tau > 0$. In fact, the latter must hold for all $\tau > 0$.

To complete the proof, we will first prove that the following equivalence holds for all $k \geq 2$ and for all $\tau > 0$:

$$\text{``}\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau) \Leftrightarrow \rho_1 = \rho_2\text{''}. \tag{5.7}$$

We stress that the above equivalence is not the same as:

$$\text{``for all } k \geq 2 \text{ and for all } \tau > 0 : \widetilde{v}_k(\tau) = \widetilde{w}_k(\tau)\text{''} \Leftrightarrow \rho_1 = \rho_2. \tag{5.8}$$

In fact, the latter equivalence (5.8) is obviously true. The statement (5.7) reads that we have to prove the equivalence for any $k \geq 2$ fixed and for any $\tau > 0$ fixed. The '$\Leftarrow$' implication is trivial. The difficult implication in (5.7) is the '$\Rightarrow$', i.e., prove for arbitrary $\tau > 0$ fixed that the following implications hold

$$\widetilde{v}_2(\tau) = \widetilde{w}_2(\tau) \Rightarrow \rho_1 = \rho_2 \tag{5.9}$$

$$\widetilde{v}_3(\tau) = \widetilde{w}_3(\tau) \Rightarrow \rho_1 = \rho_2 \tag{5.10}$$

$$\vdots \qquad \qquad \vdots$$

$$\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau) \Rightarrow \rho_1 = \rho_2 \tag{5.11}$$

for all $k \geq 2$. Equivalently, we have to prove that if $\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau)$ for some $k \geq 2$ and for some $\tau > 0$, then necessarily $\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau)$ for all $k \geq 1$ and for all $\tau \geq 0$.

The proof of the claim is as follows. Fix arbitrary $\tau > 0$. Then, the implication (5.9) is straightforward, since $\widetilde{v}_2(\tau) = \widetilde{w}_2(\tau)$ is equivalent to

$$2\tau^2/(1-\rho)^2 - \widetilde{\alpha}_2(\tau) = 2\tau^2/(1-\rho)^2 - \widetilde{\beta}_2(\tau),$$

and thus $\widetilde{\alpha}_2(\tau) = \widetilde{\beta}_2(\tau)$, and hence necessarily $\rho_1 = \rho_2$. The implications (5.11) for $k \geq 3$ are not straightforward due to the alternating term $(-1)^j$ in the recursive formula for the moments. For example, if $\widetilde{v}_3(\tau) = \widetilde{w}_3(\tau)$ holds, then it is not immediately clear that $\widetilde{v}_2(\tau) \neq \widetilde{w}_2(\tau)$ cannot be true. We will prove now that if $\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau)$, for some $k \geq 3$, then necessarily $\widetilde{v}_2(\tau) = \widetilde{w}_2(\tau)$ and $\rho_1 = \rho_2$. To this end, we first need the following strong equivalences:

$$\text{For some } n \geq 2 : \widetilde{\alpha}_n(\tau) = \widetilde{\beta}_n(\tau) \Leftrightarrow \text{For all } n \geq 2 : \widetilde{\alpha}_n(\tau) = \widetilde{\beta}_n(\tau) \Leftrightarrow \rho_1 = \rho_2, \tag{5.12}$$

$$\text{For some } n \geq 2 : \widetilde{\alpha}_n(\tau) \neq \widetilde{\beta}_n(\tau) \Leftrightarrow \text{For all } n \geq 2 : \widetilde{\alpha}_n(\tau) \neq \widetilde{\beta}_n(\tau) \Leftrightarrow \rho_1 \neq \rho_2, \tag{5.13}$$

i.e., either $\widetilde{\alpha}_n(\tau) = \widetilde{\beta}_n(\tau)$ for all $n \geq 2$, or $\widetilde{\alpha}_n(\tau) \neq \widetilde{\beta}_n(\tau)$ for all $n \geq 2$, must be true. The strong equivalences (5.12) and (5.13) follow directly from the similar $\mathbb{R}_+$-valued polynomial structure of (5.5) and (5.6) with the same strictly positive coefficients if $\tau > 0$, for every $n \geq 2$; also see the equivalences (i)-(ii)-(iii)-(iv).

Now we are able to prove the implication (5.11) for arbitrary $k \geq 3$. So, suppose if $\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau)$, then we have two mutual exclusive possibilities:

(a) $\widetilde{\alpha}_k(\tau) = \widetilde{\beta}_k(\tau)$
(b) $\widetilde{\alpha}_k(\tau) \neq \widetilde{\beta}_k(\tau)$

Exactly one possibility must hold; $\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau)$ implies either (a) or (b). Moreover, if (a) holds, then necessarily $\widetilde{\alpha}_2(\tau) = \widetilde{\beta}_2(\tau)$, $\widetilde{v}_2(\tau) = \widetilde{w}_2(\tau)$ and $\rho_1 = \rho_2$ by (5.12). If (b) holds, then necessarily $\widetilde{\alpha}_2(\tau) \neq \widetilde{\beta}_2(\tau)$, $\widetilde{v}_2(\tau) \neq \widetilde{w}_2(\tau)$ and $\rho_1 \neq \rho_2$ by (5.13). We have proven the implication (5.11) if (a) holds. To this end, we will show that (b) cannot occur, by contradiction. So, suppose (b) is true and necessarily (a) false, i.e., if (b) is true then $\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau)$ implies (b), which is equivalent with the statement

$$\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau) \Rightarrow \rho_1 \neq \rho_2, \tag{5.14}$$

which is not true, since the negation of (5.14), i.e.,

$$\rho_1 = \rho_2 \Rightarrow \widetilde{v}_k(\tau) \neq \widetilde{w}_k(\tau),$$

is clearly not the case. Hence, we conclude that the assumption of possibility (b) true, is false.

Thus, necessarily possibility (a) is true, i.e.,

$$\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau) \Rightarrow \rho_1 = \rho_2,$$

which completes the proof of the sequence of implications (5.9), (5.10),... for all $k \geq 2$.

Since $\tau > 0$ is arbitrary, it also holds that the equivalence $\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau) \Leftrightarrow \rho_1 = \rho_2$ is true for all $k \geq 2$ and for all $\tau > 0$. In particular, if for some $\tau^* > 0$ and some $k \geq 2 : \widetilde{v}_k(\tau^*) = \widetilde{w}_k(\tau^*)$ then necessarily $\rho_1 = \rho_2$, and thus $\widetilde{v}_k(\tau) = \widetilde{w}_k(\tau)$ for all $\tau \geq 0$ and for all $k \geq 1$.

We conclude for $\rho_1 > \rho_2$ that either $\widetilde{v}_k(\tau) < \widetilde{w}_k(\tau)$ must be true, or $\widetilde{v}_k(\tau) > \widetilde{w}_k(\tau)$ must be true, for all $k \geq 2$ and for all $\tau > 0$, since $\widetilde{v}_k(\tau)$ and $\widetilde{w}_k(\tau)$ are continuous functions in all arguments τ, ρ_1 and ρ_2. This can be seen from (5.5), (5.6) and its corresponding recursive formulas for the moments, and the fact that $\widetilde{B}(x)$ is a proper distribution function. Note from the above that not necessarily $\widetilde{v}_j(\tau) < \widetilde{w}_j(\tau)$ for all $j \geq 2$ must be true, or not necessarily $\widetilde{v}_j(\tau) > \widetilde{w}_j(\tau)$ for all $j \geq 2$ must be true. But we will also show that this is the case.

To this end, the proof is completed, if we can find a $\tau^* > 0$, such that $\widetilde{v}_j(\tau^*) < \widetilde{w}_j(\tau^*) \Leftrightarrow \rho_1 > \rho_2$, for all $j \geq 2$. This can be done by choosing τ^* large enough, since $V_{\lambda_1}(\tau)/\tau \xrightarrow{\mathbb{P}} 1/(1-\rho_1)$ and $V_{\lambda_2}(\tau)/\tau \xrightarrow{\mathbb{P}} 1/(1-\rho_2)$, as $\tau \to \infty$, cf. [26]. And for $\tau \to \infty$ we have that

$$\frac{V_{\lambda_1}(\tau)^{(N_2+1)*}}{\tau} \xrightarrow{d} \frac{N_2+1}{1-\rho_1}, \text{ and } \frac{V_{\lambda_2}(\tau)^{(N_1+1)*}}{\tau} \xrightarrow{d} \frac{N_1+1}{1-\rho_2}.$$

It is readily verified that (cf. proof of Theorem 4.1, and (5.2)):

$$\mathbb{E}\left(\frac{N_2+1}{1-\rho_1}\right)^k = \frac{\sum_{i=0}^{k-1} \left\langle{k \atop i}\right\rangle \left(\frac{\rho_2}{1-\rho_1}\right)^i}{(1-\rho)^k}, \text{ and } \mathbb{E}\left(\frac{N_1+1}{1-\rho_2}\right)^k = \frac{\sum_{i=0}^{k-1} \left\langle{k \atop i}\right\rangle \left(\frac{\rho_1}{1-\rho_2}\right)^i}{(1-\rho)^k},$$

hence for $k = 1$ we have the equality

$$\mathbb{E}\left(\frac{N_2+1}{1-\rho_1}\right) = \mathbb{E}\left(\frac{N_1+1}{1-\rho_2}\right) = \frac{1}{1-\rho},$$

irrespective of ρ_1 and ρ_2, provided that $\rho := \rho_1 + \rho_2 < 1$. For higher moments we have the moment ordering:

$$\mathbb{E}\left(\frac{N_2+1}{1-\rho_1}\right)^j < \mathbb{E}\left(\frac{N_1+1}{1-\rho_2}\right)^j, \text{ for all } j \geq 2, \text{ iff } \rho_1 > \rho_2, \qquad (5.15)$$

and with equality in (5.15) iff $\rho_1 = \rho_2$, and hence $\widetilde{v}_k(\tau) \leq \widetilde{w}_k(\tau)$ iff $\rho_1 \geq \rho_2$. $\blacksquare$

Remark 5.11 The moment ordering (5.15) can be intuitively explained as follows. Both random variables have the same means, irrespective of ρ_1 and ρ_2 (provided $\rho < 1$). However, if $\rho_1 > \rho_2$, then N_1 is stochastically larger than N_2, while $1/(1-\rho_1)$ and $1/(1-\rho_2)$ are fixed numbers. Hence, we expect that $\frac{N_2+1}{1-\rho_1}$ is less variable than $\frac{N_1+1}{1-\rho_2}$ if $\rho_1 > \rho_2$. Note that $\frac{N_1+1}{1-\rho_2}$ is not stochastically larger than $\frac{N_2+1}{1-\rho_1}$. This can be seen from the fact that $X \leq_{st} Y$ and $\mathbb{E}X = \mathbb{E}Y$ implies $X \stackrel{d}{=} Y$, for arbitrary (non-negative) random variables X and Y.

Theorem 5.10 can be interpreted as follows. For a fixed $\tau > 0$, if the sojourn time $V_{\lambda_2}(\tau)^{(N_1+1)*}$ is very large, then this is more likely due to the presence of many permanent customers in the

system (large λ_1) instead of a large arrival intensity of non-permanent customers (large λ_2).

Remark 5.12 We conjecture that $V_{\lambda_1}(\tau)^{(N_2+1)*} \leq_{cx} V_{\lambda_2}(\tau)^{(N_1+1)*}$ holds if $\lambda_1 \geq \lambda_2$, i.e., the random variables are ordered in *convex stochastic ordering* (see [18, 17]). Then, it is said that the random variable $V_{\lambda_1}(\tau)^{(N_2+1)*}$ is less variable than the variable $V_{\lambda_2}(\tau)^{(N_1+1)*}$. The first moments are necessarily equal. This can also be interpreted as: $V_{\lambda_2}(\tau)^{(N_1+1)*}$ is more likely to take 'extreme' values than $V_{\lambda_1}(\tau)^{(N_2+1)*}$, if $\lambda_1 \geq \lambda_2$. A sufficient condition for convex stochastic ordering is the so-called Karlin & Novikoff cut-criterion, cf. [18], which states that two r.v's X and Y are convex stochastic ordered if the means are equal and the corresponding distribution functions crosses each other once and exactly once. The difficulty to verify the cut-criterion is that we do not have the distribution functions explicitly. We note that the cut-criterion and the intuition for the conjecture given in the instantaneous sojourn time analysis, are similar (see Remark 4.2).

Remark 5.13 If $\lambda_1 \geq \lambda_2$, then it follows immediately from (5.5)-(5.6) that $\widetilde{\alpha}_n(\tau) \geq \widetilde{\beta}_n(\tau)$, and from (5.3)-(5.4) that $\widetilde{v}(s,\tau) \leq \widetilde{w}(s,\tau)$ for all $\mathrm{Re}\, s \geq 0$. Since $\widetilde{v}_1(\tau) = \widetilde{w}_1(\tau) = \tau/(1-\rho)$, it follows by Theorem 5.2 that $\widetilde{v}_2(\tau) \leq \widetilde{w}_2(\tau)$. And for all λ_1 and λ_2 provided $\rho < 1$, it also holds that $V_{\lambda_1}(\tau)^{(N_2+1)*} \in \mathcal{L}$ and $V_{\lambda_2}(\tau)^{(N_1+1)*} \in \mathcal{L}$.

We arrive at our final result that the instantaneous sojourn time moments are upper bounds for the moments of the conditional sojourn time distribution in the M/G/1 PS queue.

Theorem 5.14 In the M/G/1 PS queue, if $\rho < 1$, and for $\tau \geq 0$, $k \in \mathbb{N}$:

$$z_k(\tau) := \tau^k/(1-\rho)^k \leq v_k(\tau) \leq w_k(\tau) := \tau^k \mathbb{E}(N+1)^k,$$

where N is the steady state queue length distribution (2.1). The moments $w_k(\tau) = \tau^k \mathbb{E}(N+1)^k$ are given by the moments $\widehat{v}_k(\tau)$ of the instantaneous sojourn time, i.e.,

$$\widehat{v}_k(\tau) = \tau^k \mathbb{E}(N+1)^k = \frac{\tau^k}{(1-\rho)^k} \sum_{j=0}^{k-1} \left\langle {k \atop j} \right\rangle \rho^j.$$

Proof. The result is trivial for $\tau = 0$. For $\tau > 0$, we take $\rho_2 = 0$ in Theorem 5.10. Then, we have $N_2 = 0$ and $V_{\lambda_2}(\tau) = \tau$ (with probability 1), thus

$$V_{\lambda_1}(\tau)^{(N_2+1)*} \overset{d}{=} V_{\lambda_1}(\tau), \text{ and } V_{\lambda_2}(\tau)^{(N_1+1)*} \overset{d}{=} \tau\,(N_1+1)\,.$$

Hence, by Theorem 5.10, for all $\rho \equiv \rho_1 \geq \rho_2 = 0$, and $\rho < 1$,

$$V(\tau) \equiv V_{\lambda_1}(\tau) \leq_{\mathrm{mom}} \tau\,(N_1+1) \equiv \tau\,(N+1) = \widehat{V}(\tau),$$

and also $v_k(\tau) \equiv \widetilde{v}_k(\tau) \leq \widetilde{w}_k(\tau) \equiv w_k(\tau)$ for all $k \in \mathbb{N}$. By Theorem 4.1, we have that $w_k(\tau) = \widehat{v}_k(\tau) = \tau^k/(1-\rho)^k \sum_{j=0}^{k-1} \left\langle {k \atop j} \right\rangle \rho^j$. The lower bound $z_k(\tau)$ for $v_k(\tau)$ is trivial, by Jensen's inequality. ■

Remark 5.15 Note that taking $\rho_2 = 0$ in Theorem 5.10 and in the proof of Theorem 5.14, is essentially the same as the assumptions made in the instantaneous sojourn time analysis, $\tau \to 0$, as in Section 4. For $\rho_2 \to 0 : V_{\lambda_2}(\tau)^{(N_1+1)*} \to \widehat{V}(\tau) = (N_1+1)\tau \equiv (N+1)\tau$, as if the tagged customer arrived at a system with n_1 permanent customers with probability $\mathbb{P}(N_1 = n_1)$ and with no other arriving customers ($\rho_2 = 0$).
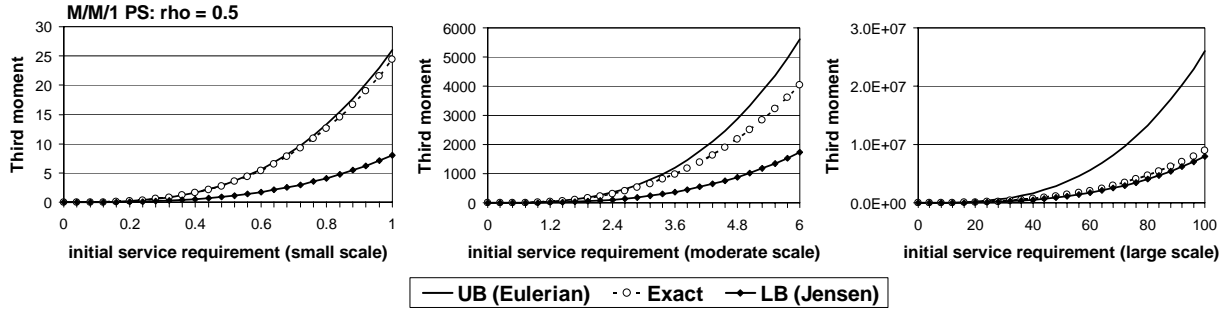
Figure 1: Third moment of the conditional sojourn time in the M/M/1 PS queue with $\rho = 0.5$ and mean service time $\beta_1 = 2$, as function of the initial service requirement $\tau > 0$.
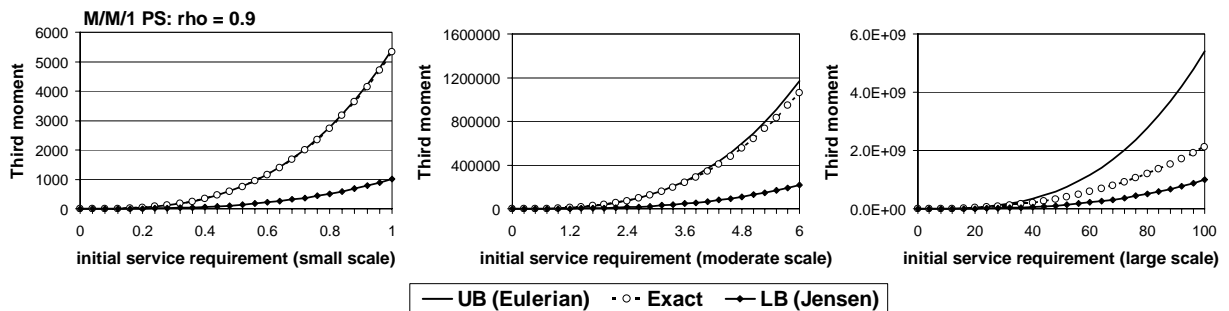

Figure 2: Third moment of the conditional sojourn time in the M/M/1 PS queue with $\rho = 0.9$ and mean service time $\beta_1 = 2$, as function of the initial service requirement $\tau > 0$.

Remark 5.16 The upper bound for the tail probabilities $\mathbb{P}(V(\tau) > x)$ as given in Corollary 5.5 can be improved by using $v_k(\tau) \leq (1 + \sum_{i=1}^{k-1} \langle {}_i^k \rangle \rho^i) / [(1-\rho)/\tau]^k$ and the Chebyshev-Markov inequalities, for *all* $k \geq 1$. The improvement is considerable, particularly for large x; the bounds are still simple, explicit and in closed-form expressions.

6 Numerical example

In this section we numerically illustrate the lower and upper bounds. Furthermore we examine the quality of the bounds for the M/M/1 PS queue. The service requirements are exponentially distributed with mean $\beta_1 = 2$. Figure 1 and Figure 2 depict the third moment $v_3(\tau)$, together with its Jensen's lower bound $z_3(\tau) = \tau^3/(1-\rho)^3$, and the third moment of the instantaneous sojourn time $w_3(\tau) = (1 + 4\rho + \rho^2)\tau^3/(1-\rho)^3$ as an upper bound for $v_3(\tau)$.

In Figure 1 the workload ρ equals 0.5, and in Figure 2 the workload ρ equals 0.9. In addition, both Figures are depicted on three different scales for the initial service requirement τ, respectively on a small ($0 \leq \tau \leq \frac{1}{2}\beta_1$), moderate ($0 \leq \tau \leq 3\beta_1$), and a large scale ($0 \leq \tau \leq 50\beta_1$). Graphs for other moments than $k = 3$ show similar behavior with respect to the Figures 1 and 2, except for $v_1(\tau)$. The only difference is a different scale and magnitude.

As expected from the instantaneous sojourn time analysis ($\tau \to 0$), it turns out that the upper bound is a good approximation for $v_k(\tau)$ when the initial service requirement τ is small ($\tau \to 0$).

The lower bound is generally a better approximation for very large τ, since $V(\tau)/\tau \xrightarrow{\mathbb{P}} 1/(1-\rho)$ as $\tau \to \infty$, for all $\rho < 1$ fixed. But obviously, it is intuitively clear that the 'speed of convergence' of $V(\tau)/\tau$ to $1/(1-\rho)$ as $\tau \to \infty$, is slower for a larger ρ.

In addition, for all $\tau > 0$ fixed we have that $v_k(\tau)/w_k(\tau) \to 1$ and $v_k(\tau)/z_k(\tau) \to 1$ as $\rho \to 0$, and $v_k(\tau)/w_k(\tau) \to 1$ and $v_k(\tau)/z_k(\tau) \to k!$ as $\rho \to 1$. These observations are not illustrated in the figures, but it can be seen from the results (4.2), (4.3), and Theorem 5.14.

7   Conclusion

In this study, we have investigated the sojourn time $V(\tau)$ conditional on the initial service requirement $\tau > 0$ of a tagged customer in the M/G/1 processor-sharing (PS) queue. In particular, we studied all moments of $V(\tau)$ and we obtained upper and lower bounds. Our main result (Theorem 5.14 and 5.10) is that there exists an upper bound for the moments of the conditional sojourn time distribution, which can be expressed in terms of the workload $\rho < 1$ and the initial service requirement τ only. The upper bound for $\mathbb{E}V(\tau)^k$ is given by $\left(1 + \sum_{i=1}^{k-1} \left\langle {k \atop i} \right\rangle \rho^i \right) \tau^k/(1-\rho)^k$, where $\left\langle {k \atop i} \right\rangle$ are Eulerian numbers. A lower bound follows easily from Jensen's inequality.

An attractive feature of the upper bound of the above structure is that it is independent of second- and higher moments of the service time distribution. Another attractive feature is that the upper bound converges to the Jensen's lower bound when $\rho \to 0$, for all τ fixed. And for $\rho \to 1$, the k-th moment of the true sojourn time and its upper bound of the above structure, converge to the same expression, after proper scaling.

The upper bound of the above structure with Eulerian numbers is related to a so-called *instantaneous* sojourn time analysis, where we studied the sojourn time of a customer with a very small initial service requirement ($\tau \to 0$). If the initial service requirement $\tau > 0$ is arbitrary (and not necessarily small), the instantaneous sojourn time analysis also corresponds to the situation of a certain PS model with a random number of permanent customers.

By studying the higher moments and providing insensitive upper bounds, we strengthen the fact that processor-sharing is a very fair service discipline. Under the stability condition $\rho < 1$, excessive behavior of other customers in the system always has a limited influence on the sojourn time of the tagged customer. Intuitively, from a tagged customer point-of-view, the influence of the service time requirements of other customers in the system on the sojourn time of the tagged customer, is nearly insensitive. Even when there is a customer with infinite service time requirement, the influence of this *permanent* customer on non-permanent customers is limited.

We conclude this paper with the remark that considerable attention has been paid in the literature to the exact analysis of the sojourn time in the M/G/1 PS queue. Relatively little work has been done on the investigation of the practical implications of the results. From a practical point-of-view, the discovery of simple bounds for *all* moments of the conditional sojourn time stimulates the investigation of simple but nevertheless good approximations for the distribution of $V(\tau)$, the moments and the tail probabilities. This remains a topic for further research.

References

[1] J.L. van den Berg. *Sojourn times in Feedback and Processor Sharing Queues*, Ph.D. thesis, Utrecht University, The Netherlands, 1990.

[2] A. Bhattacharjee & D. Sengupta. On the coefficient of variation of the $\mathcal{L}$- and $\overline{\mathcal{L}}$-classes. *Statistics & Probability Letters* 27 (1996) 177–180.

[3] L. Carlitz. Eulerian Numbers and Polynomials of Higher Order. *Duke Mathematical Journal* 27 (1960) 401-423.

[4] S.K. Cheung, J.L. van den Berg, and R.J. Boucherie. Decomposing the queue length distribution of processor-sharing models into queue lengths of permanent customer queues. To appear in: *Proceedings of IFIP's Performance 2005*, Oct. 3-7, Juan-les-Pins, France. Also available at http://www.math.utwente.nl/publications as Research Memorandum No. 1753, Department of Applied Mathematics, University of Twente, 2005.

[5] J.W. Cohen. The multiple phase service network with generalized processor sharing. *Acta Informatica* 12 (1979) 245-284.

[6] R.L. Graham, D.E. Knuth & O. Patashnik. *Eulerian Numbers, §6.2 in Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. Reading, MA: Addison-Wesley, 1994, pp. 267-272.

[7] B. Klar. A note on the $\mathcal{L}$-class of life distributions. *Journal of Applied Probability* 39 (2002), no. 1: 11-19.

[8] B. Klefsjö. A useful ageing property based on Laplace transform. *Journal of Applied Probability* 20 (1983) 615-626.

[9] L. Kleinrock. Analysis of a time-shared processor. *Naval Research Logistics Quarterly* 11 (1964) 59- 73.

[10] L. Kleinrock. Time-shared systems: A theoretical treatment. *Journal of the Association for Computing Machinery* 14 (1967) 242-261.

[11] L. Kleinrock. *Queueing Systems, Vol II: Computer Applications*, Wiley, New York, 1976.

[12] G. Lin. Characterizations of the $\mathcal{L}$-class of life distributions. *Statistics & Probability Letters* 40 (1998) 259–266.

[13] R. Núñez-Queija. *Processor Sharing Models for Integrated Services Networks*, Ph.D. thesis, Eindhoven University of Technology, The Netherlands, 2000.

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

[15] M. Sakata, S. Noguchi & J. Oizumi. Analysis of a processor-shared queueing model for time-sharing systems. *Proceedings 2nd Hawaii International Conference on System Sciences*, Jan. 1969, pp. 625-628.

[16] R. Schassberger. A new approach to the M/G/1 processor sharing queue. *Advances in Applied Probability* 16 (1984) 202-213.

[17] M. Shaked & J.G. Shanthikumar. *Stochastic orders and their applications*, Academic Press Inc., 1994.

[18] D. Stoyan. *Comparison Methods for Queues and Other Stochastic Models*, Wiley, Chichester, 1983.

[19] Weisstein, Eric W., "Eulerian Number". *From MathWorld–A Wolfram Web Resource.* http://mathworld.wolfram.com/EulerianNumber.html

[20] W. Whitt. The M/G/1 processor-sharing queue with long and short jobs. *Unpublished manuscript* (1998).

[21] D.V. Widder. *The Laplace Transform.* Princeton, NJ: Princeton University Press, 1941.

[22] S.F. Yashkov. A derivation of response time distribution for a M/G/1 processor-sharing queue. *Problems of Control and Information Theory* 12 (1983) 133-148.

[23] S.F. Yashkov. Processor-sharing queues: Some progress in analysis. *Queueing Systems* 2 (1987) 1-17.

[24] S.F. Yashkov. Mathematical problems in the theory of processor-sharing queueing systems. *Journal of Soviet Mathematics* 58 (1992) 101-147.

[25] S.F. Yashkov. On a heavy traffic limit theorem for the M/G/1 processor sharing queue. *Communications in Statististics - Stochastic Models* 9 (1993), no. 3: 467-471.

[26] A.P. Zwart & O.J. Boxma. Sojourn time asymptotics in the M/G/1 processor sharing queue. *Queueing Systems* 35 (2000) 141-166.

[27] A.P. Zwart. *Queueing Systems with Heavy Tails*, Ph.D. thesis, Eindhoven University of Technology, The Netherlands, 2001.

Appendix

Eulerian numbers

For $k \in \mathbb{N}$, and $j = 0, 1, ..., k$, the Eulerian numbers $\left\langle {k \atop j} \right\rangle$ are given by

$$\left\langle {k \atop 0} \right\rangle = \left\langle {k \atop k-1} \right\rangle = 1, \left\langle {k \atop k} \right\rangle = 0, \tag{A.1}$$

and for $j = 1, ..., k-1$, given by the recurrence relation

$$\left\langle {k \atop j} \right\rangle = (k-j)\left\langle {k-1 \atop j-1} \right\rangle + (j+1)\left\langle {k-1 \atop j} \right\rangle. \tag{A.2}$$

A slightly different definition of the Eulerian number can be used. The notation $A(k, j)$ is sometimes referred as the Eulerian number and it is defined *shifted*, i.e., $A(k, j) = \left\langle {k \atop j-1} \right\rangle$ or equivalently defined by: $A(k, 0) = 0$, $A(k, 1) = A(k, k) = 1$, and the recurrence relation: $A(k, j) = (k - j + 1) \cdot A(k-1, j-1) + j \cdot A(k-1, j)$. We use the notation and the definition $\left\langle {k \atop j} \right\rangle$ for the Eulerian number throughout this paper.

The Eulerian numbers $\left\langle {k \atop j} \right\rangle$ are given explicitly by the sum

$$\left\langle {k \atop j} \right\rangle = \sum_{i=0}^{j+1} (-1)^i \binom{k+1}{i} (j-i+1)^k. \tag{A.3}$$

The Eulerian numbers satisfy the sum identity

$$\sum_{j=0}^{k} \left\langle {k \atop j} \right\rangle = k!. \tag{A.4}$$

The arrangement of the numbers $\left\langle {k \atop j} \right\rangle$ for $k \geq 1$ and $j = 0, ..., k-1$ gives the Euler's number triangle. The defining recurrence relation weighs the sum of neighbors by their row and column numbers, respectively.

Euler's number triangle

```
                              1
                       1             1
                 1            4            1
           1          11           11          1
      1         26          66          26         1
  1       57         302         302         57        1
1     120       1191        2416       1191       120      1
1   247     4293       15619       15619      4293      247    1
```

The Eulerian numbers have many combinatorial applications, and they represent a sort of generalization of the binomial coefficients. It is also interesting to note that the rows of Eulerian numbers approach a normal density for $k \to \infty$, just as do the rows of the binomial coefficients (Pascal's number triangle). The Eulerian number $\left\langle {k \atop j} \right\rangle$ gives the number of permutations of the elements in the set $\{1, ..., k\}$ having j permutation ascents (or descents). For example, the six possible permutations of the set $\{1, 2, 3\}$, along with the number of 'rises' (i.e., the number of times it goes from a lower to a higher number, reading left to right) are shown below:

permutation	1-2-3	1-3-2	2-1-3	2-3-1	3-1-2	3-2-1
number of rises	2	1	1	1	1	0

The numbers of permutations having exactly 0, 1, and 2 rises (or falls) are 1, 4, and 1 respectively, and these numbers comprise the 3rd row of the Euler's number triangle, i.e., $\left\langle {3 \atop 0} \right\rangle = 1$, $\left\langle {3 \atop 1} \right\rangle = 4$, $\left\langle {3 \atop 2} \right\rangle = 1$. The number $\left\langle {3 \atop 3} \right\rangle$ is by definition 0.

Interestingly, the Eulerian numbers arises in polylogarithm functions with negative integer index of the form (see e.g., [19])

$$\sum_{k=1}^{\infty} k^n r^k = Li_{-n}(r) \equiv \frac{1}{(1-r)^{n+1}} \sum_{i=0}^{n} \left\langle {n \atop i} \right\rangle r^{n-i}, \tag{A.5}$$

where the $\mathcal{Z}$-transform of the sequence $\{n^k\}_{n=1}^{\infty}$ is equal to $\mathcal{Z}\left[\{n^k\}_{n=1}^{\infty}\right](z) = Li_{-k}\left(\frac{1}{z}\right)$ and the polylogarithm function is defined as $Li_n(z) \equiv \sum_{k=1}^{\infty} \frac{z^k}{k^n}$. For more on Eulerian numbers we refer to [19, 3, 6] and references therein.