Информационные процессы, Том 3, № 2, 2003, стр. 128–137.

MATHEMATICAL MODELS

DISTRIBUTION OF THE VIRTUAL SOJOURN TIME IN THE M/G/1 PROCESSOR SHARING QUEUE[1]

A.S.Yashkova*, **S.F.Yashkov****

**Municipal Institute of Zhukovsky, Chair Appl. Informatics in Economics, 15, Mayakovsky Street, 140180 Zhukovsky, Moscow region, Russia*
***Institute for Information Transmission Problems, 19, Bolshoy Karetny Lane, 101447 Moscow GSP–4, Russia*
E-mail: yashkov@iitp.ru
Received September 2, 2003

Abstract—Processor sharing queuing systems are well–known and very attractive models in applied probability and modern applications of queueing theory. This paper continues the investigations of the M/G/1 queue with Egalitarian Processor Sharing (EPS) discipline. We give, in the form of triple transforms, the time–dependent distribution of virtual sojourn time of a job, arriving at time t with the service demand (length) u in this queueing system which is empty at initial instant $t = 0$. Further we study main important special cases as the direct consequences of our general result. We show also how extend the main theorem to the case when the M/G/1—EPS queue is modified by having $K \geq 0$ extra permanent jobs with infinite lengths.

1. INTRODUCTION

Processor sharing queueing models, made popular by the works of L.Kleinrock [1] and S.F.Yashkov [2], [3], [4], [5], were originally proposed to analyze the performance of time sharing scheduling algorithms in computer systems. During the past two decades, the processor sharing paradigm has emerged as a powerful concept for modeling of Web servers, in particular, the flow–level performance of bandwidth–sharing protocols in the nodes of modern computer–communication networks. The Transmission Control Protocol (TCP) in adaptive window mechanism can be considered as an example.

This paper continues the investigations of the M/G/1 queue with the Egalitarian Processor Sharing (EPS) discipline from [3], [5], [6], [7], [8].

An idea of EPS is due to Kleinrock [9]. Under the EPS discipline, the processor (server) is shared equally by all jobs in the system. To put more concretely, when $1 \leq n < \infty$ jobs are present in the system, each job receives service at rate $1/n$. In other words, all these jobs receive $1/n$ times the rate of service which a solitary job in the processor would receive. Jumps of the service rate occur at the instants of arrivals and departures from the system. Therefore, the rate of service received by a specific job fluctuates with time and, importantly, its sojourn time depends not only on the jobs in the processor at its time of arrival there, but also on subsequent arrivals shorter of which can overtake a specific job. This makes the EPS system intrinsically harder to analyze than, say, the First Come — First Served (FCFS) queue.

The problem of the determination of the stationary sojourn time distribution in the M/G/1—EPS queue was open a long time. This problem was first solved, after puzzling researchers for 15 years, by Yashkov in the late 1970s (see his papers of that period, cited in [4], [5], or [3], [10] with complete proofs) in terms of double Laplace tranforms (LT) (see also [5]). The time–dependent queue–length distributions in

[1] The research of the second author was supported in part by Grant–in–Aid for Leading Scientific Schools of President of Russia for the Promotion of Science under Grant no. Sci.Sch.–1532.2003.1 (Head N.A.Kuznetsov) and Grant no. Sci.Sch.–934.2003.1 (Head R.A.Minlos).

this and related (for example, FBPS) M/G/1 models have been derived in increasing level of generality in [11], [5], [12], [13]. These exact solutions have been extended further to obtain the time–dependent distribution of the virtual sojourn time of a job arriving at time t with the length u into the M/G/1 — EPS queue (in terms of triple transforms). Almost all available at present analytic solutions of the M/G/1—EPS system (and also many new) can be derived as special cases of this generalized result, as a rule, by means of Tauber's and Abel's theorems. The main purposes of this paper are to present our generalized theorems and to study basic consequences from theirs.

The rest of this paper is divided into four sections. In section 2, the mathematical model is described and the main theorem is presented. In section 3, we derive a number of new propositions as corollaries of theorem 2.1. In section 4, we give an extension of theorem 2.1 to the case of K permanent jobs (the time–dependent solution). Two new ways for obtaining the steady–state solution are also considered. As a by–product, we show that the determination of the steady–state sojourn time distribution in the queue M/G/1—EPS with K permanent jobs is simple extension of the results from [3], [10]. Finally, in section 5, we provide short concluding remark.

2. MAIN RESULT

We consider an M/G/1 queue where the jobs arrive according to Poisson process with the rate λ. Let $B(x)$ be the distribution function of the service times (better: the lengths) of jobs $(B(0) = 0, B(\infty) = 1)$ with the mean $\beta_1 < \infty$ and the Laplace–Stieltjes transform (LST) $\beta(s)$. The service discipline is the EPS: when there are n jobs in the system, each receive service at a rate $1/n$. Let $L(t)$ and $V(t, u)$ be the number of jobs at time t and the sojourn time of a tagged job (of the length u) arriving at time t, respectively. We assume that the system is empty at time $t = 0$. Let $\zeta = \inf(t > 0 : L(t) = 0)$ and $\pi(s) = \mathsf{E}[\mathrm{e}^{-s\zeta}]$ is the LST of the busy period distribution, i.e., it is the positive root of the functional equation

$$\pi(s) = \beta(s + \lambda - \lambda\pi(s)) \tag{2.1}$$

with the smallest absolute value. Define

$$\tilde{v}_0(r, s, u) \doteq \int_0^\infty \mathrm{e}^{-st}\mathsf{E}\left[\mathrm{e}^{-rV(t,u)} \mid L(0) = 0\right] dt \quad (\mathrm{Re}\, s, r > 0), \tag{2.2}$$

$\mathsf{P}_{00}(t) \doteq \mathsf{P}(L(t) = 0|L(0) = 0)$ and let $\tilde{p}_{00}(s)$ be the LT of $\mathsf{P}_{00}(t)$.

Theorem 2.1. *For any positive $\rho = \lambda\beta_1$, the exact solution for the function $\tilde{v}_0$ is:*

$$\tilde{v}_0(r, s, u) = \tilde{p}_{00}(s)\frac{\delta(r, u)}{1 - \tilde{a}(r, s, u)/\psi(r, u)}, \tag{2.3}$$

where

$$\tilde{p}_{00}(s) = [s + \lambda - \lambda\pi(s)]^{-1}, \tag{2.4}$$

$$\delta(r, u) = \mathrm{e}^{-u(r+\lambda)} / \psi(r, u) \ , \tag{2.5}$$

$\psi(r, u)$ is given by its LT with respect to (w.r.t.) u (argument q):

$$\tilde{\psi}(r, q) = \frac{q + r + \lambda\beta(q + r + \lambda)}{(q + r + \lambda)(q + \lambda\beta(q + r + \lambda))} \quad (r \geq 0,\ q > -\lambda\pi(r)), \tag{2.6}$$

$$\tilde{a}(r, s, u) = \int_0^\infty \mathrm{e}^{\lambda\pi(s)t}\alpha(r, s, t, u)\, dt \tag{2.7}$$

and $\alpha(r,s,t,u)$ is given by its double LT w.r.t. t and u (arguments q and c, respectively):

$$\tilde{\alpha}(r,s,q,c) = \frac{\lambda}{c+r-q-s}\Bigg\{\tilde{\psi}(r,c)[\beta(q+s+\lambda)-\beta(c+r+\lambda)] + \frac{1}{q+s+\lambda}(1-\beta(q+s+\lambda)) - \frac{1}{c+r+\lambda}(1-\beta(c+r+\lambda))\Bigg\}. \quad (2.8)$$

Proof. Due to lack of space, we give below only main comments to the proof.

The proof is based on some extensions of arguments from [3], [5], [12] and it also uses an amalgam of techniques such as terminating (sub)busy periods, the random time change, and also some facts from the renewal theory and branching processes. The special case as $t \to \infty$ and $\rho = \lambda\beta_1 < 1$ is proved in [3], [10]. The time–dependent distributions of the jobs at the time t is investigated in [11], [5], [6], [12], [7] and [8]. These works contain also many important things which are the corner–stones of our promising method of analysis. To renew the omitted steps of our proof, it is useful to exploit the aforementioned proofs as intermediate phases. Briefly, we have derived the expression for $\tilde{v}_0$ by writing the sojourn time and the number of jobs at time t as some generalized functional on a (non-trivial) branching process (like the processes by Crump–Mode–Jagers). Using the structure of the branching process, we found and solved a system of partial differential equations (of the first order) determining $\tilde{v}_0$. The solution contains the Bromwich countour integrals, i.e., the Laplace transform inversion operators $\mathcal{L}^{-1}$. For example, $\psi(r,u) = \mathcal{L}^{-1}(\tilde{\psi}(r,q))(r,u)$ in (2.6).

Some additional comments. The one of the main ideas of the proof is to study an additive functional describing the accumulation of the attained service of some tagged job in time

$$X(t) = \int_0^t \frac{1}{1+L(y)}\, dy$$

with $L(y)$ being the number of jobs at time y (cf. the equation (2.34) in [6]). (It can be shown that the process $\{L(\cdot)\}$ is positive recurrent and Harris ergodic when $\rho < 1$. That will not pose a problem as $\rho > 1$, this and related processes will be transient in such case.) If we define for $u \geq 0$

$$V(0,u) = \inf\,(t \geq 0 : \, X(t) \geq u),$$

then $V(0,u)$ being the sojourn time of the tagged job arriving at time 0. From another side, the sojourn time $V(t,u)$ satisfies to the equality

$$\int_t^{t+V(t,u)} \frac{1}{1+L(y)}\, dy = u.$$

The study of such processes on the (standard) first busy period enables us to decompose $V(t,u)$ on some terminating (sub)busy periods (introduced in [3], [10]) and to obtain corresponding equations describing an evolution of these terminating (sub)busy periods. The matter is made easier if to introduce the new time scale by means of a random time change. In the new time scale, the system M/G/1—EPS is transformed into the system $\tilde{M}/G/\infty$ in which the rate of input is equal to $\lambda L(t)$ (and to λ if $L(t) = 0$). After the solution of the corresponding equations (the generalized versions of (3.10) and (3.13) from [3]), the standard arguments from the renewal theory can be used to connect $\mathsf{E}[\mathrm{e}^{-sV(t,u)}\mathbf{1}_{(\zeta>t)}]$ on the first busy period with $\mathsf{E}[\mathrm{e}^{-sV(t,u)}]$ on all positive half–axis. □

Remark 2.1. Theorem 2.1 remains in force not only for the stability condition $\rho < 1$ but even for $\rho \geq 1$.

Remark 2.2. Theorem 2.1 can be represented in several equivalent forms. Here we have chosen the form which resembles the formula (5.2) from [4] (see also [13] for another appearance of the result announced as Theorem 3).

The equation (2.3) in the theorem 2.1 gives us also the unconditional distribution of the time–dependent sojourn time after removing the condition on u (averaging on $B(\cdot)$)

$$\int_0^\infty e^{-st}\mathsf{E}\left[e^{-rV(t)}|L(0)=0\right]\,dt = \int_0^\infty \frac{\delta(r,u)}{1-\tilde{a}(r,s,u)/\psi(r,u)}\,dB(u), \tag{2.9}$$

where $\psi(r,u)=\mathcal{L}^{-1}(\tilde{\psi}(r,q))(r,u)$ and $\tilde{\psi}(r,q)$ is given by (2.6).

The formula (2.9) is some analog of the well–known result of Takacs for the time–dependent virtual waiting time distribution in the M/G/1—FCFS queue (see [14]):

$$\int_0^\infty e^{-st}\mathsf{E}\left[e^{-rW(t)}|L(0)=0\right]\,dt = \frac{s+\lambda-\lambda\pi(s)-r}{(s+\lambda-\lambda\beta(r)-r)[s+\lambda-\lambda\pi(s)]}. \tag{2.10}$$

Takacs obtained (2.10) by means of derivation and solution of his integro–differential equation for the process of the unfinished work in the system M/G/1—FCFS [14]. However, his method does not work for obtaining (2.9), so we derive (2.9) by means of our principally new method which introduced in [3], [10].

Remark 2.3. It is not difficult to show that the functional equations (2.1) and (2.4) are equivalent [15]. To do this, we can reformulate (2.1) and (2.4) to the form

$$\tilde{p}_{00}(s) = [s+\lambda-\lambda\beta(1/\tilde{p}_{00}(s))]^{-1}.$$

3. MAIN CONSEQUENCES

We proceed to some important corollaries from Theorem 2.1. The corollaries of this theorem will be formulated in the form of propositions.

Proposition 3.1. *For any positive ρ, $\mathsf{E}[V(t,u)]$, the mean conditional sojourn time of the job (of the length u) arriving at time t, is given by its LT w.r.t. t (argument s)*

$$v_1(s,u) = \frac{\delta_1(u)}{s^2\tilde{p}_{00}(s)} - \frac{\lambda}{s^2\tilde{p}_{00}(s)}\int_0^u \delta_1(u-x)\left[e^{x/\tilde{p}_{00}(s)}\int_x^\infty e^{-y/\tilde{p}_{00}(s)}dB(y)\right]dx,$$

where

$$\delta_1(u) = u + \int_0^u \sum_{n=1}^\infty \rho^n F^{n*}(x)dx$$

for $0<\rho<\infty$. Here $\tilde{p}_{00}(s)$ is given by the equation (2.4) (it is the LT of $\mathsf{P}\{L(t)=0\}$) and $F^{n*}(x)$ is n–fold convolution of $F(x)=\beta_1^{-1}\int_0^x(1-B(y))dy$ (the excess of $B(\cdot)$) with itself ($F^{0*}(x)=\mathbf{1}(x)$, the Heaviside function, $F^{1*}(x)=F(x)$).

Proof. From (2.3) via $v_1(s,u)=-\lim_{r\downarrow 0}\partial\tilde{v}_0(r,s,u)/\partial r$. □

Remark 3.1. When $\rho<1$, the well–known result for the (conditional) steady–state mean sojourn time $\mathsf{E}[V(u)]=u/(1-\rho)$ follows from the proposition 3.1 as $t\to\infty$ by means of classical Tauber's theorem.

Proposition 3.2. *For $\rho<1$, the distribution of $V(u)$, the stationary sojourn time of the job with the length u, is given by*

$$v(r,u) \doteq \mathsf{E}[e^{-rV(u)}] = \frac{(1-\rho)\delta(r,u)}{1-\tilde{a}(r,0,u)/\psi(r,u)}\,, \tag{3.1}$$

where

$$\tilde{a}(r,0,u) = \lambda\int_0^u \psi(r,u-x)e^{-x(r+\lambda)}(1-B(x))dx + \lambda e^{-u(r+\lambda)}\int_u^\infty (1-B(x))dx \tag{3.2}$$

and the LT of $\psi(r,u)$ w.r.t. u is given by (2.6).

Proof. For $\rho < 1$, we have $\pi(0) = 1$. Hence, the result follows from (2.6) by the classical Tauber's theorem as

$$\lim_{s\downarrow 0} s\tilde{v}_0(r,s,u) = \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathsf{E}[\mathrm{e}^{-rV(t,u)}]dt = v(r,u)$$

taking also into account L'Hospitale's rule. □

At first this result had been proved by Yashkov in the late of seventies (see the references in [4] and also [3], [10] for the improved proof with details). Like the theorem 2.1, this solution contains the contour integrals. Equivalent form of (3.1) (without contour integrals) is given in the following proposition.

Proposition 3.3. *Equivalent form of (3.1) (without the Bromwich countour integrals) is given by*

$$\frac{1}{v(r,u)} = \sum_{n=0}^{\infty} \frac{r^n}{n!}\xi_n(u), \tag{3.3}$$

where

$$\xi_0(u) = 1, \quad \xi_n(u) = \frac{n}{(1-\rho)^n} u^{n-1} * W^{(n-1)*}(u), \quad n = 1, 2, \dots \tag{3.4}$$

Here $W^{(n-1)*}(u)$ is $(n-1)$–fold convolution of the steady–state waiting time distribution $W(u)$ in the familiar M/G/1—FCFS system with itself ($W^{0*}(u) = \mathbf{1}(u)$, $W^{1*}(u) = W(u)$), the LST of $W(u)$ is given by the well–known Pollaczek–Khintchine formula

$$w(q) = \frac{1-\rho}{1-\rho f(q)}, \tag{3.5}$$

where $f(q) = (1-\beta(q))/(q\beta_1)$ is the LST of the excess $F(x)$, introduced in proposition 3.1.

Proof. We rewrite (3.1) in the form of the equation (5.5) from [4]. To do this, it is used the LT of $1/\delta(r,u)$ w.r.t. u (argument q), which is given by (2.6) as $\tilde{\psi}(r, q-r-\lambda), \quad r \geq 0,\ q > r+\lambda-\lambda\pi(r)$ (see also the third line in [4, p. 8]). We obtain after simple algebra

$$\int_0^\infty \mathrm{e}^{-qu}\frac{1}{v(r,u)}du = \frac{1}{q}\left[1 + \frac{1}{1-\rho}\frac{r}{q}\frac{1}{1-\frac{1}{1-\rho}\frac{r}{q}w(q)}\right] = \frac{1}{q}\left[1+\sum_{n=1}^{\infty}\left(\frac{1}{1-\rho}\frac{r}{q}\right)^n w(q)^{n-1}\right], \tag{3.6}$$

where $w(q)$ is given by (3.5). We note that $\left|\frac{rw(q)}{(1-\rho)q}\right| < 1$ as $q > r+\lambda-\lambda\pi(r),\ \rho < 1$. Now it is easily to invert analytically (on argument q) each term of the power series in r (3.6). The result is given by (3.4) whence the assertion (3.3) follows, the right–hand side of which is the power series in r with coefficients $\xi_n(u)/n!$ □

Remark 3.2. The idea of such approach has been known at least since the book of O.Heaviside [16, Ch. 5]. This result is also obtained in [17] as Theorem 3.1. However, the starting point of the proof in [17] is the formula which is the other (less convinient) form of (3.1), namely, (5.6) in [4]; moreover, our proof does not use the knowledge of $\mathsf{Var}[V(u)]$ [3], [10]

$$\mathsf{Var}[V(u)] = \frac{2}{(1-\rho)^2}\int_0^u (u-x)(1-W(x))dx. \tag{3.7}$$

In fact, the form of $\mathsf{Var}[V(u)]$ stimulates a guess about the possibility of such expansion.

Remark 3.3. We note that the by–product of our analysis is the distribution function $W(x)$ whose LST is given by (3.5). However, the analysis of EPS queue gives the other quantity (corresponding to a non–probability measure) $W^\circ(x) = W(x)/(1-\rho)$. The form of the LST of $W^\circ(x)$ is well–known: $w^\circ(q) = \sum_{n=0}^\infty \rho^n f^n(q)$. Unlike $W(x)$, $W^\circ(x)$ is well defined for all $\rho > 0$ and $x > 0$. It can be shown that $W^\circ(x) < \infty$ for all $\rho > 0$, $x > 0$ and for any $B(\cdot)$ (despite on the fact that, for $\rho \geq 1$, $W^\circ(x) \to \infty$ as $x \to \infty$).

Proposition 3.4. *Let* $v_n(u) = \mathsf{E}[V(u)^n]$, $n = 1, 2, \ldots$ *Then it holds the following recursive formula*

$$v_n(u) = \sum_{i=1}^{n} \binom{n}{i} v_{n-i}(u)\xi_i(u)(-1)^{i+1}. \tag{3.8}$$

Proof. It is well–known the Tailor series expansion of $v(r,u)$ for small $r > 0$

$$v(r,u) = 1 - \frac{r}{1!}v_1(u) + \frac{r^2}{2!}v_2(u) - \frac{r^3}{3!}v_3(u) + \ldots \tag{3.9}$$

The product of (3.9) and (3.3) gives

$$\begin{aligned} -\frac{r}{1!}[v_1(u) - \xi_1(u)] + \frac{r^2}{2!}[v_2(u) - 2v_1(u)\xi_1(u) + \xi_2(u)] \\ - \frac{r^3}{3!}[v_3(u) - 3v_2(u)\xi_1(u) - 3v_1(u)\xi_2(u) - \xi_3(u)] + \ldots = 0 \end{aligned}$$

and it leads to the assertion (3.8) after differentiating n times w.r.t. r and setting $r = 0$. □

Remark 3.4. Proposition 3.4 is important for applications, because the LST (3.1) is very difficult to differentiate at zero more than once (practically almost impossible matter) due to a highly complex form of such constituents of (3.1) as $\tilde{a}$ and ψ.

Remark 3.5. The formula (3.8) was derived by Yashkov in 1979 by solving a system of equations describing the decomposition of $\mathsf{E}[V(u)^2]$ (see, for example, [3] and the references herein) for $n = 2$. The proof above is contained in [8] (cf. the related proof in [17]) and the result holds for arbitrary integer n. The formula (3.8) can be very useful for asymptotic expansion of $v_n(u)$ for small and large u in the spirit of such expansion for $\mathsf{Var}[V(u)]$, which made at first in [18], [10]. In particular, the rough asymptotic of $\mathsf{E}[V(u)^2]$ from [17, remark 3.3] coincides with one from [18, remark 2]. However, [18], [10] contain more exact asymptotics of $\mathsf{Var}[V(u)^2]$, including the estimate of the rate of the convergence (see [18, theorem 2]).

4. THE M/G/1—EPS QUEUE WITH K PERMANENT JOBS

We give at first the theorem for time–dependent distribution of $V(t,u)$ when the M/G/1—EPS queue is modified by having $K \geq 0$ extra permanent jobs with infinite lengths. Almost all main definitions and notations have been introduced in section 2; some additional ones are the following. Let us define

$$\tilde{v}_K^B(r,s,u) \doteq \int_0^\infty \mathrm{e}^{-st}\mathsf{E}\left[\mathrm{e}^{-rV(t,u)}\mathbf{1}_{(\zeta>t)} \,|\, K\right] dt \quad (\mathrm{Re}\, s, r > 0) \tag{4.1}$$

on the first busy period and

$$\tilde{v}_K(r,s,u) \doteq \int_0^\infty \mathrm{e}^{-st}\mathsf{E}\left[\mathrm{e}^{-rV(t,u)} \,|\, K\right] dt \quad (\mathrm{Re}\, s, r > 0) \tag{4.2}$$

on all positive half–axis.

The notation $L(t)$ remains in force for the standard jobs, the same is true for the standard busy period (2.1) for such jobs. Now $\tilde{p}_{0K}(s)$ being the LT of $\mathsf{P}_{0K}(t) \doteq \mathsf{P}(L(t) = 0|K)$.

Theorem 4.1. *For any positive $\rho = \lambda\beta_1$, the exact solution for the function $\tilde{v}_K$ is:*

$$\tilde{v}_K(r,s,u) = \tilde{p}_{0K}(s)[\delta(r,u) + \lambda\tilde{v}_1^B(r,s,u)] + \tilde{v}_K^B(r,s,u), \tag{4.3}$$

where

$$\tilde{p}_{0K}(s) = \pi^K(s)\tilde{p}_{00}(s), \tag{4.4}$$

$\pi(s)$, $\tilde{p}_{00}(s)$ and $\delta(r,u)$ are given by (2.1), (2.4) and (2.5), respectively,

$$\tilde{v}_K^B(r,s,u) = \frac{1}{\lambda^K}\delta(r,u)\int_0^\infty \left[-\frac{\partial}{\partial t}\ln\psi^V(r,s,t,u)\right]^K dt, \tag{4.5}$$

the function $\psi^V(r,s,t,u)$ is given by its LT w.r.t. t (argument q):

$$\tilde{\psi}^V(r,s,q,u) = \frac{\psi(r,u) - \tilde{\alpha}(r,s,q,u)}{q + \lambda\beta(s+\lambda+q)}. \tag{4.6}$$

Here $\psi(r,u) = \mathcal{L}^{-1}(\tilde{\psi}(r,q))(r,u)$ in (2.6) and $\tilde{\alpha}(r,s,q,u) = \mathcal{L}^{-1}(\tilde{\alpha}(r,s,q,c))(r,s,q,u)$ in (2.8). It remains to add that for $K = 1$, (4.5) is reduced to (in fact, it is very non–trivial task)

$$\tilde{v}_1^B(r,s,u) = \frac{\mathrm{e}^{-u(r+\lambda)}a(r,s,u)}{\lambda[\psi(r,u) - a(r,s,u)]}, \tag{4.7}$$

where $a(r,s,u)$ is found by inversions of (2.8) (on arguments c and q) and then from (2.7).

Proof. Omitted. However, the comments to the proof of theorem 2.1 remain in force with the evident slight modifications. In general, we deal with the investigation of the same processes as in the case $K = 0$ but for other initial conditions. □

Now we show how to obtain one of the corollaries of theorem 4.1. In fact, it is an extension of proposition 3.2 to this case.

Proposition 4.1. *For $\rho < 1$,*

$$\mathsf{E}[\mathrm{e}^{-rV(u)}] = v(r,u)^{K+1}, \tag{4.8}$$

where $v(r,u)$ is given by (3.1), and $V(u)$ is the steady–state sojourn time of the standard job of the length u.

Proof. (Sketch.) The first way: to apply the classical Tauber's theorem as it was done in the proof of proposition 3.2. It needs some tedious algebra for this. Details are omitted. □

Proof. The second way. For $K = 0$, this proposition has been proved in [3] as theorem 4 (see also [10] and [5, p.63–70]) . The partial differential equations in the proof of aforementioned theorem rely on a decomposition of the sojourn time of the job of the length u that arrives to the EPS queue when n standard jobs are present with remaining service demands $x_1, \ldots x_n$ (a key ingredient of analysis). Denoting this conditional sojourn time by $V_n(u; x_1, \ldots x_n)$, it holds

$$V_n(u; x_1, \ldots x_n) = D(u) + \sum_{i=1}^{n} \Phi(x_i, u). \tag{4.9}$$

The random variable $D(u)$ constitutes a "main" component of the sojourn time: it has the distribution of the sojourn time of a job with the length u that enters into a empty system. When the system is not empty, the i–th job (among the jobs which are sharing the capacity of the processor), having remaining length x_i, "adds" a delay $\Phi(x_i, u) = \Phi(x_i \wedge u, u)$ to the new job's sojourn time. Note that $D(u) = \Phi(x_i, u)$ for $x_i \geq u$;

the expectation of $D(u)$ is given by $\delta_1(u)$ (see proposition 3.1). In the case $K \geq 0$, the decomposition in (4.9) will be

$$V_n(u; x_1, \dots x_n | K) = (K+1)D(u) + \sum_{i=1}^{n} \Phi(x_i, u).$$

Then the same chain of arguments as in [3] can be used to derive (4.8). □

Remark 4.1. This result was obtained in [19] by means of another approach (see also [20]). For other applications of our method to the various variants of the EPS queue (as a rule, in steady state) see, for example, [5], [21].

Corollary 4.1. *In the setting above,*

$$\mathsf{E}[V(u)|K] = (K+1)u/(1-\rho),$$

$$\mathsf{Var}[V(u)|K] = 2(K+1)/(1-\rho)^2 \int_0^u (u-x)(1-W(x))dx,$$

where $W(u)$ is the steady–state waiting time distribution in the M/G/1—FCFS queue, whose LST is given by (3.5).

Remark 4.2. Using proposition 3.4 (or corollaries 5 and 6 from [8]), we can easily obtain all other moments of $V(u)$ in M/G/1—EPS queue with K permanent jobs.

Note that unconditional (only on $B(x)$) sojourn time distribution of an arbitrary job is

$$\mathsf{P}(V \leq x|K) = \int_0^\infty \mathsf{P}(V(u) \leq x|K) dB(u)$$

.

Assume that $\mathsf{P}(B > x) \sim x^{-\alpha}\ell(x)$, $x \to \infty$, where $\alpha > 1$, α not an integer, and ℓ is a slowly varying (at infinity) function. Then it holds

Proposition 4.2. *In the setting above (M/G/1—EPS queue with K permanent jobs, $B(x)$ has a regularly varying tail), it holds*

$$\mathsf{P}(V > x|K) \sim \mathsf{P}\left(B > (K+1)(1-\rho)x\right), \quad x \to \infty. \tag{4.10}$$

Proof. We give here only few comments to the proof.

Modify the proof of theorem 4.1 in [17] to this case, assuming $K > 0$ and using our corollary 4.1 and remark 4.2. □

So far, the assertion (4.10) has been apparently proved for $K = 0$. Using the new asymptotic technique (somewhat intricate) from [22], proposition 4.2 can be extended on more wide subclasses of subexponential distributions $B(\cdot)$ in the queue M/G/1—EPS (the case $K = 0$ is completely investigated in [22]) and also on other variants of processor sharing models. Such distributions arise often in the modern communication networks; asymptotic behaviour of simpler model M/G/1—FCFS is well studied in [23] under subexponential $B(\cdot)$.

The relation (4.10) is called Reduced Load Equivalence. The first guess about the possibility of similar equivalence (in weaker form) apparently is due to Kleinrock (see [1, p. 175]). In essence, up to now Kleinrock's explanation of this is used for comments to Reduced Load Equivalence in a number of recent mathematical papers.

5. CONCLUSION

Theorems 2.1, 4.1 and their main corollaries provide the extremely powerful tools for the deep study of time–dependent behaviour of the queue M/G/1 under egalitarian processor sharing discipline. The decomposition in (4.9) has played a key role in the exact and asymptotic analysis of processor sharing queues. Our results give also the ways for investigation of the transient ($\rho > 1$) and null recurrent ($\rho = 1$) behaviour of this queueing system with K permanent jobs. Such queue represents the basic model for predicting delays in the host–servers of computer–communication networks.

REFERENCES

1. Kleinrock L. *Queueing Systems*. New-York: Wiley, 1976, vol. 2. Russian edition: Kleinrock L. *Computer systems with Queues*. Moscow: Mir, 1979.
2. Lipayev V.V., Yashkov S.F. *Efficiency of Methods of Organizing Computation Process in Automatic Control Systems*. Moscow: Statistika, 1975 (in Russian).
3. Yashkov S.F. A derivation of response time distribution for an M/G/1 processor-sharing queue. *Probl. of Contr. and Info. Theory*, 1983, vol. 12, no. 2, pp. 133–148.
4. Yashkov S.F. Processor–sharing queues: some progress in analysis. *Queueing Systems*, 1987, vol. 2, no. 1, pp. 1–17.
5. Yashkov S.F. *Queueing Analysis for Computers*. Moscow: Radio i Svyaz, 1989 (in Russian).
6. Yashkov S.F. Mathematical problems in the theory of shared-processor systems. In: *Itogi Nauki i Tekhniki. Ser.: Probability Theory*. Moscow: VINITI, 1990, vol. 29, pp. 3–82 (in Russian). Engl. edition: *J. of Soviet Mathematics*, 1992, vol. 58, no. 2 (Jan.), pp. 101–147.
7. Yashkov S.F., Yashkova A.S. Processor sharing queue: transient solutions. In: *Computer Sci. and Info. Technologies (CSIT'99). Proc. 2nd Int. Conf.* (Yerevan, Aug. 17–22, 1999). Yerevan: National Acad. of Sci. of Armenia, 1999, pp. 99–103.
8. Yashkov S.F., Yashkova A.S. Processor sharing queue: additional results. In: *Distributed Comput. Commun. Networks. Proc. 3rd Int. Conf.* (Tel–Aviv, Nov. 9–13, 1999), Moscow: Inst. for Info. Transm. Probl., 1999, pp. 216–221.
9. Kleinrock L. Time–shared systems: a theoretical treatment. *J. Assoc. Comput. Mach.*, 1967, vol. 14, no. 2, pp. 242-251.
10. Yashkov S.F. Some results of analyzing a probabilistic model of remote processing systems. *Avtom. i Vychisl. Tekhn.*, 1981, no. 4, pp. 3-11 (in Russian). Engl. edition: *Autom. Control and Computer Sci.*, 1981, vol. 15, no. 4, pp. 1-8.
11. Yashkov S.F. The non–stationary distribution of numbers of calls in the M/G/1 processor–sharing queue. In: *Systems Analysis and Simulation 1988. Proc. 3rd Int. Symp.* (Berlin, Sept. 12–16, 1988). Eds. A. Sydow et al. Berlin: Akademie, 1988, vol. 2 (Ser.: Math. Research, Bd. 47), pp. 158–162. Reprinted in: *Advances in Simulation*. Eds. P.A.Lukar and B.Schmidt. Berlin: Springer, 1988, vol. 2, pp. 158–162.
12. Yashkov S.F. Time–dependent analysis of processsor–sharing queue. In: *Queueing, Performance and Control in ATM. Proc. 13th Int. Teletraffic Congress.* (Copenhagen, June 19–26, 1991). Eds. J.W.Cohen and C.D.Pack. Amsterdam: Elsevier, 1991, pp. 199-204.
13. Yashkov S.F., Yashkova A.S. The M/G/1 processor–sharing system: transient solutions. In: *Distributed Comput. Commun. Networks. Proc. 2nd Int. Conf.* (Tel–Aviv, Nov. 4-8, 1997). Moscow: Inst. for Info. Transm. Probl., 1997, pp. 261–272.
14. Takacs L. *Introduction to the Theory of Queues*. New–York: Oxford Univ. Press, 1962.
15. Abate J., Whitt W. Transient behavior of the M/G/1 workload process. *Operations Research*, 1994, vol. 42, no. 4, pp. 750–764.
16. Heaviside O. *Electromagnetic Theory*. London: The Electrician Co., 1893, vol. 1; 1899, vol. 2; 1912, vol. 3.

17. Zwart A.P., Boxma O.J. Sojourn time asymptotics in the M/G/1 processor sharing queue. *Queueing Systems*, 2000, vol. 35, pp. 141–166.

18. Yashkov S.F. A note on asymptotic estimates of the sojourn time variance in the M/G/1 queue with processor–sharing. *Systems Analysis. Modelling. Simulation*, 1986, vol. 3, no. 3, pp. 267–269.

19. van den Berg J.L., Boxma O.J. M/G/1 queue with processor–sharing and its relation to a feedback queue. *Queueing Systems*, 1991, vol. 9, no. 4, pp. 365-401.

20. Whitt W. The M/G/1 processor–sharing queue with long and short jobs. *Unpublished manuscript*, 1998 (Sept.).

21. Nunez–Queija R.N. Sojourn times in non–homogeneous QBD processes with processor–sharing. *Stochastic Models*, 2001, vol. 17, no. 1, pp. 61–92.

22. Jelenkovic P., Momeilovic P. Resource sharing with subexponential distributions. In: *Proc.IEEE Infocom'2002* (June 2002, New York). New York, 2002.

23. Abate J., Choudhuri G.L., Whitt W. Waiting–time probabilities in queues with long–tail service–time distributions. *Queueing Systems*, 1994, vol. 16, no. 3–4, pp. 311–338.

This paper was recommended for publication by V.I.Venets, a member of the Editorial Board