Информационные процессы, Том 6, № 3, 2006, стр. 250–255.

TWO SPECIAL CASES OF THE M/G/1—EPS QUEUE [1]

S.F.Yashkov

Institute for Information Transmission Problems, 19, Bolshoi Karetny lane, 127994 Moscow GSP–4, Russia. E-mail: yashkov@iitp.ru

Received June 20, 2006

Abstract—Starting from the well–known results of analysis of the M/G/1 queue with egalitarian processor sharing (EPS), we consider two special cases: the M/M/1—EPS and the M/D/1—EPS, and show how to obtain the (conditional) sojourn time distribution for these special cases from more general results.

1. INTRODUCTION

Last years processor sharing models continue to find new applications, for example, for predicting delays in WEB servers, in performance evaluation of the main variants of the Transmission Control Protocols in INTERNET, for provisioning servers for E–commerce systems, etc. The central role is played by the simpler variants of the M/G/1 queue with egalitarian processor sharing (EPS), for instance, by the M/M/1—EPS queueing system. It is instructive to show how some special results can be obtained from the well–known results of analysis of the M/G/1—EPS queue. Seemingly, two main seminal papers concerning the determination of the stationary sojourn time distribution in such model (in terms of double Laplace transforms (LT)) are [1] (1983) and [2] (1984). (It seems that all subsequent papers in this direction are, in fact, some "derivatives" of the papers cited above. For example, the Ott's paper (1984) has, in essence, a loan character despite a slight generalized result.) Starting from the results of [1,2], here we consider two special cases: the M/M/1—EPS and the M/D/1—EPS, and show how to obtain the (conditional) sojourn time distribution for these special cases from more general results. We also rely (partially) on the begin of the section 2.6 in [3, pp. 73–75] (1989) (this research monograph is not sufficiently available for English–language readers because it was published in Russian).

2. PRELIMINARIES

In this section we give a short representation about the main results of the determination of the stationary sojourn time distribution (in terms of double Laplace transform) for the M/G/1—EPS queue. These results were derived in the begin of eighties of twenty centure by Yashkov [1] (1983) and Schassberger [2] (1984) independently from each other by means of totally different new analytical methods . The papers [1, 2] also give some representation about the previous works of these authors in this directory, that enable them to get eventually this result the significance of which is at least commensurable with the Pollaczek–Khinchine formula.

Jobs arrive to the single processor (server) according to a Poisson process with the rate $\lambda > 0$. Their sizes (reguired service times) are i.i.d. random variables with a general distribution function $B(x)$ $((B(0) = 0,\ B(\infty) = 1))$ with the mean $\beta_1 < \infty$ and the Laplace–Stieltjes transform (LST)

[1] This research was partially supported by Grant no. Sci.Sch.–934.2003.1 (Head R.A.Minlos) and Grant to the Program of Fundamental Researches of Russian Academy of Sciences (the Division of Informatics) "New physical and structural solutions in infotelecommunications"(Head N.A.Kuznetsov).

$\beta(s) = \int_{0-}^{+\infty} \mathrm{e}^{-sx} dB(x)$ [2]. The service discipline is the EPS: every job is being served with rate $1/n$, when $n > 0$ jobs are present in the system. 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 queue intrinsically harder to analyze than, say, the classical First Come—First Served (FCFS) queue or many other classical disciplines. The system works in steady state. In other words, $\rho = \lambda\beta_1 < 1$ and very long time went from the start of the work of the EPS system till current time.

It is well known that the stationary distribution $(\mathsf{P}_n)_{n\geq 0}$ of L, the number of jobs in the M/G/1—EPS queue, has the geometrical form

$$\mathsf{P}_n = \mathsf{P}(L = n) = (1-\rho)\rho^n, \quad n \in 0 \cup \mathbb{N}, \tag{2.1}$$

where $\rho = \lambda \int_0^\infty (1 - B(x))dx < 1$. We note that $(\mathsf{P}_n)_{n\geq 0}$ depends on the service time only through its mean.

We shall let that $V(u)$ denotes the conditional sojourn time of a job of the size u upon its arrival. Define the LST of $V(u)$ by $v(s,u) = \mathbb{E}[\mathrm{e}^{-sV(u)}]$ for Re $s \geq 0$ and $u \geq 0$.

Let $\pi(r)$ be the LST of the busy period distribution . In other words, it is the positive root of the well–known Takács functional equation

$$\pi(s) = \beta(s + \lambda - \lambda\pi(s)) \tag{2.2}$$

with the smallest absolutely value.

It is known from [1], [2] the following theorem which is given below in somewhat different (but equivalent) form that resembles Theorem 3.2 in [2] (except for a difference in notations of the components which, in turn, are drawn from Theorem 4 [1]).

Theorem 2.1. *When $\rho < 1$,*

$$v(s,u) \doteq \mathbb{E}[\mathrm{e}^{-sV(u)}] == \frac{(1-\rho)\mathrm{e}^{-u(s+\lambda)}}{\psi(s,u) - \tilde{a}(s,u)}\,. \tag{2.3}$$

Here

$$\tilde{a}(s,u) = \lambda\psi(s,u) * \left[\mathrm{e}^{-u(s+\lambda)}(1 - B(u))\right] + \lambda\mathrm{e}^{-u(s+\lambda)} \int_u^\infty (1 - B(x))dx, \tag{2.4}$$

where "$*$" is the Stieltjes convolution sign (on variable u), and $\psi(s,u)$ is the LST (with respect to x) of some function $\Psi(x,u)$ of two variables (possessing the probability density on variable x), which, in turn, has a Laplace transform (LT) with respect to u(argument q)

$$\tilde{\psi}(s,q) = \frac{q + s + \lambda\beta(q+s+\lambda)}{(q+s+\lambda)(q + \lambda\beta(q+s+\lambda))} \quad (s \geq 0,\ q > -\lambda\pi(s)). \tag{2.5}$$

In (2.5), $\beta(s) = \int_{0-}^{+\infty} \mathrm{e}^{-sx}\, dB(x)$ and $\pi(s)$ (in the conditions imposed on (2.5) is understood as the minimal solution of the functional equation (2.2).

[2] There is no loss of generality in assuming that $\beta_1 = 1$, since this case may be handled by rescaling the rate of the Poisson process. Furthermore, we assumed that $B(x)$ has no atom in the origin. For otherwise, the pattern of busy and idle periods is essentially the same as in a queueing process for which arrival rate is reduced to $\lambda[1 - \mathsf{P}(B = 0)]$, and service time has the distribution of B given that $B > 0$.

Thus, the function $\tilde{\psi}(s,q)$ is given in the form of the two–dimensional transform of the function $\Psi(x,u)$)

$$\tilde{\psi}(s,q) = \int_0^\infty \int_0^\infty \mathrm{e}^{-sx-qu} d_x \Psi(x,u) du. \tag{2.6}$$

In other words, $\psi(s,u)$ in equality (2.3) is the Laplace transform inversion operator, $\psi(s,u) = \mathcal{L}^{-1}(\tilde{\psi}(s,q))(s,u)$, that is, the contour Bromvich integral

$$\psi(s,u) = \frac{1}{2\pi i} \int_{-i\infty+0}^{+i\infty+0} \tilde{\psi}(s,q) \mathrm{e}^{qu}\, dq.$$

Remark 21. Briefly, we have derived the expression for $\mathbb{E}[\mathrm{e}^{-sV_K(u)}]$ by writing the sojourn time as some generalized functional on a branching process (like the processes by Crump–Mode–Jagers) by means of simple extensions of (non–trivial) arguments from [1]. Using the structure of the branching process, we found and solved a system of partial differential equations (of the first order) determining the components of a (non–trivial, too) decomposition of $V(u)$. It leads to $\mathbb{E}[\mathrm{e}^{-sV(u)}]$.

It is instructive to rewrite (2.3) as

$$v(s,u) = (1-\rho)\delta(s,u) \left[1 - \rho \int_0^\infty \varphi(s,x,u) \frac{(1-B(x))}{\beta_1} dx \right]^{-1} = \frac{(1-\rho)\delta(s,u)}{1-\rho\varphi(s,u)}, \tag{2.7}$$

where

$$\varphi(s,x,u) = \begin{cases} \delta(s,u) & \text{for } x \ge u, \\ \delta(s,u)/\delta(s,u-x) & \text{for } x < u, \end{cases} \tag{2.8}$$

and

$$\delta(s,u) = \mathrm{e}^{-u(s+\lambda)}/\psi(s,u), \quad u \ge 0. \tag{2.9}$$

(The equality for $\psi(s,u)$ is given above.)

Remark 22. In some cases, it can be useful the equivalent forms of (2.8). For example,

$$\varphi(s,x,u) = \mathrm{e}^{-(x\wedge u)(s+\lambda)+\lambda\int_0^{x\wedge u} \varphi_B(s,u-y)\,dy}, \; x \in [0,\infty), \tag{2.10}$$

where

$$\varphi_B(s,t) \doteq \int_0^\infty \varphi(s,x,t)\,dB(x) = \int_0^t \mathrm{e}^{-\int_{t-x}^t (s+\lambda-\lambda\varphi_B(s,y))\,dy}\,dB(x) + (1-B(t))\mathrm{e}^{-\int_0^t (s+\lambda-\lambda\varphi_B(s,y))\,dy}. \tag{2.11}$$

The equality (2.11) represents the functional equation to which $\varphi_B(s,\cdot)$ satisfies. $\varphi_B(s,t)$ is the LST of the distribution of the terminating busy period (it terminates at time t) for the M/G/1—EPS queue. This is a non–trivial notion in queueing theory (among the others which were introduced in [1,3]): the distribution of such terminating busy period is not insensitive to the work–conserving discipline. For example, the M/G/1—FBPS queue has another distribution of the terminating busy period [3,6]. The solution of the equation (2.11) was obtained in terms of the function ψ $(\psi(s,t) \doteq \exp(-\lambda \int_0^t \varphi_B(s,y)\,dy))$ (more precisely, in terms of the LT for this function, see (2.5)). This remark also shows that the study of the sojourn time in the M/G/1 queue (even in steady–state) requires much deeper analysis in comparison with an analysis that is expected on the customary level.

Next we consider two special cases of the M/G/1—EPS queue in equilibrium: the M/D/1 and the M/M/1 systems with egalitarian processor sharing.

3. THE M/D/1—EPS QUEUE

Let us begin from the case M/D/1—EPS queueing system when the job's sizes are i.i.d. random variables with the deterministic distribution

$$B(x) = \begin{cases} 0, & 0 \le x < u, \\ 1, & x \ge u. \end{cases}$$

Hence the LST of this distribution has the form $\beta(s) = \exp(-su)$ with the moments $\beta_i = u^i$, $i = 1, 2, \ldots$. The offered load is equal to $\rho = \lambda u < 1$. In this special case, the distributions of conditioned and unconditioned sojourn times coincide, hence we may use $V = V(u)$ to denote the steady–state sojourn time of a job in the queue M/D/1—FBPS.

Corollary 3.1. *The LST of the stationary distribution of $V(u)$ in the special case M/D/1—EPS has simpler form than in the general case reflected in Teorem 2.1*

$$v(s) = v(s,u) = \frac{(1-\rho)(s+\lambda)^2 \mathrm{e}^{-u(s+\lambda)}}{s^2 + \lambda[s + (s+\lambda)(1-\rho)]\mathrm{e}^{-u(s+\lambda)}}. \tag{3.1}$$

In the case considered, the formula (2.9) takes the form

$$\delta(s,t) = \frac{s+\lambda}{\lambda + s\mathrm{e}^{t(s+\lambda)}}, \quad t \le u. \tag{3.2}$$

Proof. The solution for $v(s,u)$ for the case M/D/1—EPS can be found from Theorem 2.1 in explicit form (see [3] for details). It is convenient for this to use the equation (2.7) (cf. the equality (2.33) from [3]). In our case, the equation mentioned is reduced to the form

$$v(s) = v(s,u) = \frac{(1-\rho)\delta(s,u)}{1 - \lambda\delta(s,u)\int_0^u \frac{dx}{\delta(s,u-x)}} \tag{3.3}$$

where $\delta(s,u)$ is given by (3.2). To obtain (3.2), it is better to use the equation (3.15) from [1] for the unknown function $\delta(s,u)$ (see also (2.29) in [3] or the first equation from (2.18) in [6]) instead of inverting the function $\tilde{\psi}(s,q)$ that is given by (2.5). Then the equation mentioned reduces to the form

$$\frac{\partial\delta(s,t)}{\partial t} + (s+\lambda)\delta(s,t) - \lambda\delta(s,t)^2 = 0 \tag{3.4}$$

with the additional condition $\delta(s,0) = 1$. This is a Bernoulli equation. It is reduced to linear one after division of each term by $\delta(s,t)^2$ and the change of variable $1/\delta(s,t) = u$. The solution of (3.4) is given by (3.2). The final result (3.1) follows after the substitution (3.2) into (3.3). □

Remark 31. The expression for the variance of $V(u) = V$ (see (3.20) in [1]) reduces to the form

$$\mathrm{Var}V(u) = \frac{u^2}{(1-\rho)^2} - \frac{2u^2(\mathrm{e}^\rho - 1 - \rho)}{\rho^2(1-\rho)}.$$

4. THE M/M/1—EPS QUEUE

The following corollary holds for the second special case.

Corollary 4.1. *The LST of the stationary distribution of $V(u)$ in the special case M/M/1—EPS has the following form*

$$v(s,u) = \frac{(1-\rho)(1-\rho b^2)\mathrm{e}^{-u(s+\lambda-\lambda b)}}{(1-\rho b)^2 - \rho(1-b)^2\mathrm{e}^{-\mu u(1-\rho b^2)/b}}, \tag{4.1}$$

where

$$b = \pi(s) = \left\{(s + \lambda + \mu) - \left[(s + \lambda + \mu)^2 - 4\lambda\mu\right]^{1/2}\right\}/(2\lambda) \tag{4.2}$$

is the solution of the functional equation (2.2) in our special case.

Proof. The exponential distribution of the sizes of jobs $B(x) = 1 - \mathrm{e}^{-\mu x}$ has the LST $\beta(s) = \mu/(\mu + s)$. Taking it into account, the following quadratic equation follows from (2.2)

$$\lambda\pi(s)^2 - (s + \lambda + \mu)\pi(s) + \mu = 0, \tag{4.3}$$

the solution of which is given by (4.2) (we take the sign "minus" before the root since $|\pi(s)| \le 1$).

In our case, the differential equation for the function $\psi(s, u)$ (see, for example, (3.17) in [1]) can be rewritten to simpler form. It makes easier the conversion of (2.3) to the form

$$v(s, u) = \frac{(1 - \rho)\mathrm{e}^{-u(s+\lambda)}}{\psi(s, u) + \frac{1}{\mu}\frac{\partial\psi(s,u)}{\partial u}}. \tag{4.4}$$

It remains to invert the formula (2.5) on complex argument q, which, in our case, can be represented as

$$\tilde{\psi}(s, q) = \frac{q + s + \mu}{(q - q_1)(q - q_2)},$$

where $q_1 = -\lambda b$, $q_2 = -\mu/b$ are two simple poles of the function $\tilde{\psi}(s, q)$.

We shall use the inversion integral for carrying out the inversion procedure

$$\psi(s, u) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \tilde{\psi}(s, q)\mathrm{e}^{qu}dq \tag{4.5}$$

where $c \ge 0$. The integration in the complex q–plane is taken to be a straight–line integration parallel to the imaginary axis and lying to the right of an abscissa of absolute convergence for $\tilde{\psi}(s, q)$. The standard means for this is to apply the Cauchy residue theorem to the integral in the complex domain around a closed contour. We choose such contour in the form of a semicircle of very large radius and the straight–line segment mentioned above. Since the function $\tilde{\psi}(s, q)$ satisfies to the conditions of Jordan's lemma, the integral along the semicircle tends to zero as the radius tends to infinity. Therefore the limit of the integral along the entire closed contour coincides with the right–hand side of the inversion integral (4.5). In virtue of the Cauchy's residue theorem we have

$$\int \tilde{\psi}(s, q)\mathrm{e}^{qu}dq = 2\pi i\sum_{n=1}^{2}\mathrm{res}\left[\tilde{\psi}(s, q_n)\mathrm{e}^{q_n u}\right].$$

It remains to calculate two residues of the function $f(s, q) = \tilde{\psi}(s, q)\mathrm{e}^{qu}$ (relatively q), which are located at $q_1 = -\lambda b$ and $q_2 = -\mu/b$. Since q_1 and q_2 are the simple poles of this function $f(s, q)$, then

$$\mathrm{res} f(s, q_n) = \lim_{q\to q_n}(q - q_n)f(s, q_n).$$

Hence $\mathrm{res} f(s, -\lambda b) = (s + \mu - \lambda b)\exp(-\lambda bu)/(-\lambda b + \mu/b)$ and also $\mathrm{res} f(s, -\mu/b) = (s + \mu - \mu/b)$ $\exp(-\mu u/b)/(\lambda b - \mu/b)$. It leads to

$$\psi(s, u) = \frac{[s + \mu(1 - \rho b)]\mathrm{e}^{-\lambda bu} - [s - \mu(1 - b)/b]\mathrm{e}^{-\mu u/b}}{\mu(1 - \rho b^2)/b}.$$

Note that

$$\frac{\partial\psi(s, u)}{\partial u} = \frac{-[s + \mu(1 - \rho b)]\lambda b\mathrm{e}^{-\lambda bu} + [s - \mu(1 - b)/b]\mu\mathrm{e}^{-\mu u/b}/b}{\mu(1 - \rho b^2)/b}.$$

The substitution of two last equalities into (4.4) leads to the assertion (4.1). □

The equality (4.1) coincides with the classical result in [4] (1970) (see also [5, Chapter 4]). The method of analysis in [4] is based on birth and death processes for description of the service process, technically, it is rather cumbersome (for example, it requires solution of first–order partial differential equations) and it is not suitable for the analysis of more general cases. This method does not carry over into the M/G/1—EPS queue.

Remark 41. Sengupta and Jagerman [7] found an alternative expression for the LST of the distribution of the sojourn time conditioned only on the number of jobs seen upon arrival.

Now the formula for the variance of $V(u)$ (see (3.20) in [1]) reduces to the form

$$\mathrm{Var}[V(u)] = \frac{2\rho u}{\mu(1-\rho)^3} - \frac{2\rho}{\mu^2(1-\rho)^4}\left[1 - \mathrm{e}^{-\mu u(1-\rho)}\right]. \tag{4.6}$$

5. CONCLUSION

We gave the details of the derivations of two important corollaries of Theorem 4 [1] and Theorem 3.2 [2].

REFERENCES

1. Yashkov S.F. A derivation of response time distribution for an M/G/1 processor-sharing queue. *Probl. of Control and Information Theory*, 1983, vol. 12, no. 2, pp. 133–148.
2. Schassberger R. A new approach to the M/G/1 processor–sharing queue. *Adv. Appl. Prob.*, 1984, vol. 16, no. 1, pp. 202–213.
3. Yashkov S.F. *Analysis of Queues in Computers* (in Russian with English Summary). Moscow: Radio i Svyaz, 1989 (review in *Mathematics and Computers in Simulation*, 1991, vol. 33, no. 2, pp. 177–178).
4. Coffman E.G., Muntz R. and Trotter H. Waiting time distributions for processor–sharing systems. *J. Assoc. Comput. Mach.*, 1970, vol. 17, no. 1, pp. 123–130.
5. Kleinrock L. *Queueing Systems*. New-York: Wiley, 1976, vol. 2. Russian edition: Kleinrock L. *Computer systems with Queues*. Moscow: Mir, 1979.
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, pp. 101–147.
7. Sengupta B. and Jagerman D. A conditional response time of the M/M/1 processor–sharing queue. *AT&T Techn. J.*, 1985, vol. 64, no. 2, pp. 409–421.

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