A queueing model for multiprogrammed computer systems with different I/O times

By J. SZTRIK

1. Introduction. Multiprogrammed computer systems are important subject of research in modern queueing theory. We can model such a system as the collection of a Central Processor Unit (CPU) or CPU-s, terminals (e. g. rotating disk memory, magnetic tape, card reader, etc.) and the jobs. Each job is associated with a terminal at which it suffers no delay, queues of programs may occur only at the CPU (or CPU-s).

To analyse this kind of systems we often use the finite-source queueing model which is sometimes called the "machine interference model". For a FIFO multiprogrammed computer system a new mathematical model can be given in the following way. Let the number of jobs in the system be n $(n \geqq r)$. Jobs (or programs) emanate from the peripheral devices where various input-output (I/O) operations are carried out. An arriving program is immediately served by one of r CPU-s if there is an idle one, otherwise a waiting line is formed. The jobs are served in the order of their arrival, that is the service discipline is FIFO (first-in, first-out). The service times of the jobs are assumed to be exponentially and identically distributed random variables with mean $1/\mu$. After completing CPU operations the program i returns to its peripheral device and stays there for a random time having an arbitrary distribution function $F_i(x)$ with density $f_i(x)$. All random variables involved here are supposed to be mutually independent of each other.

Since there is a huge literature for the problem in question we refer only to the latest results. Bunday and Scraton [3] have recently proved that the probability distribution of the number of machines running in steady state is the same in the $M/M/r$ and $G/M/r$ cases. In connection with the mathematical description of multiprogrammed computer system for the interested reader the following papers can be recommended: Avi-Itzhak and Heyman [2], Asztalos [1], Csige and Tomkó [5] Gaver [6], Kameda [8], Schatte [10], Sztrik [11]. In Kleinrock's book [9] further models and good bibliography on this subject can be found.

The present paper deals with a possible generalization of the $G/M/r$ case and gives the main steady-state operational characteristics of the system, such as CPU utilization, mean waiting and response times of the jobs.

2. The mathematical model. Let the random variable $\nu(t)$ denote the number of jobs processing I/O operations at time t and $(\alpha_1(t), \ldots, \alpha_{\nu(t)}(t))$ indicate their indices ordered lexicographically. Let us denote by $(\beta_1(t), \ldots, \beta_{n-\nu(t)}(t))$ the indices of the jobs waiting or served at the CPU-s in the order of their arrival. Clearly the sets $\{\alpha_1(t), \ldots, \alpha_{\nu(t)}(t))$ and $\{\beta_1(t), \ldots, \beta_{n-\nu(t)}(t)\}$ are disjoint.

Let us introduce the process

$$\underline{Y}(t) = (\nu(t);\ \alpha_1(t), \ldots, \alpha_{\nu(t)}(t)\colon \beta_1(t), \ldots, \beta_{n-\nu(t)}(t)).$$

The stochastic process $(\underline{Y}(t), t \geqq 0)$ is not a Markovian one unless the distribution functions $F_i(x)$ are exponential, $i=1, \ldots, n$.

Let us introduce the supplementary variable $\xi\alpha_l(t)$ denoting the random time that the job $\alpha_l(t)$ has been staying at a peripheral device in the time period $(0, t)$, $l=1, \ldots, \nu(t)$. Define

$$X(t) = (\nu(t);\ \alpha_1(t), \ldots, \alpha_{\nu(t)}(t);\ \xi\alpha_j(t), \ldots, \xi\alpha_{\nu(t)}(t);\ \beta_1(t), \ldots, \beta_{n-\nu(t)}(t)).$$

the process $(\underline{X}(t), t \geqq 0)$ has the Markov property.

Let V_k^n and C_k^n denote the set of all variations and combinations, respectively, of order k of the integers $1, 2, \ldots, n$ ordered lexicographically. Then the state space of the process $(\underline{X}(t), t \geqq 0)$ consists of the points $(i_1, \ldots, i_k;\ x_1, \ldots, x_k; j_1, \ldots, j_{n-k})$, where $(i_1, \ldots, i_k) \in C_k^n$, $(j_1, \ldots, j_{n-k}) \in V_{n-k}^n$, $x_i \in \mathbf{R}_+$, $i=1, \ldots, k$, $k=1, \ldots, n$.

The process is in state $(i_1, \ldots, i_k;\ x_1, \ldots, x_k;\ j_1, \ldots, j_{n-k})$ if k jobs with indices $(i_1, \ldots, i_k)$ have been processing I/O operations for times $(x_1, \ldots, x_k)$, respectively, while the rest of jobs need the CPU-s. The indices of these programs in the order of their arrival are $(j_1, \ldots, j_{n-k})$.

To derive the Kolmogorov equations we should consider the transitions that can occur in an arbitrary time interval $(t, t+h)$. The transition probabilities are given in the following way

$$P\{\underline{x}(t+h) = (i_1, \ldots, i_k;\ x_1+h, \ldots, x_k+h;\ j_1, \ldots, j_{n-k})/$$

$$\underline{x}(t) = (i_1, \ldots, i_k;\ x_1, \ldots, x_k;\ j_1, \ldots, j_{n-k})\} =$$

$$(1-(n-k)\mu h) \prod_{l=1}^{k} \frac{1-F_{i_l}(x_l+h)}{1-F_{i_l}(x_l)} + o(h),$$

$$P\{\underline{x}(t+h) = (i_1, \ldots, i_k;\ x_1+h, \ldots, x_k+h;\ j_1, \ldots, j_{n-k})/$$

$$\underline{x}(t) = (i_1', \ldots, j_{n-k}', \ldots, i_k';\ x_1', \ldots, y', \ldots, x_k';\ j_1, \ldots, j_{n-k-1})\} =$$

$$\frac{f_{j_{n-k}}(y)h}{1-F_{j_{n-k}}(y)} \prod_{l=1}^{k} \frac{1-F_{i_l}(x_l+h)}{1-F_{i_l}(x_l)} + o(h),$$

$$\text{for} \quad 0 \leqq n-k < r,$$

where $(i'_i, \ldots, j'_{n-k}, \ldots, i'_k)$ denotes the lexicographical order of the indices $(i_1, \ldots, \ldots, i_k, j_{n-k})$, while $(x'_1, \ldots, y', \ldots, x'_k)$ indicates the corresponding times.

$$P\{\underline{x}(t+h) = (i_1, \ldots, i_k;\ x_1+h, \ldots, x_k+h;\ j_1, \ldots, j_{n-k})/$$
$$\underline{x}(t) = (i_1, \ldots, i_k;\ x_1, \ldots, x_k;\ j_1, \ldots, j_{n-k})\} =$$
$$(1-r\mu h) \prod_{l=1}^{k} \frac{1-F_{i_l}(x_l+h)}{1-F_{i_l}(x_l)} + \sigma(h),$$
$$P\{\underline{x}(t+h) = (i_1, \ldots, i_k;\ x_1+h, \ldots, x_k+h;\ j_1, \ldots, j_{n-k})/$$
$$\underline{x}(t) = (i'_1, \ldots, j'_{n-k}, \ldots, i'_k;\ x'_1, \ldots, y', \ldots, x'_k;\ j_1, \ldots, j_{n-k-1})\} =$$
$$\frac{f_{j_{n-k}}(y)h}{1-F_{j_{n-k}}(y)} \prod_{l=1}^{k} \frac{1-F_{i_l}(x_l+h)}{1-F_{i_l}(x_l)} + o(h).$$
$$\text{for} \quad r \leqq n-k \leqq n.$$

To calculate the distribution of $\underline{X}(t)$ consider the following functions.

$$Q_{0;\, j_1, \ldots, j_n}(t) = P(v(t) = 0;\ \beta_1(t) = j_1, \ldots, \beta_k(t) = j_n),$$
$$Q_{i_1, \ldots, i_k;\, j_1, \ldots, j_{n-k}}(x_1, \ldots, x_k;\ t) =$$
$$P(v(t) = k;\ \alpha_1(t) = i_1, \ldots, \alpha_k(t) = i_k;\ \xi_i \leqq x_1, \ldots, \xi_{i_k} \leqq$$
$$\leqq x_k;\ \beta_1(t) = j_1, \ldots, \beta_{n-k}(t) = j_{n-k}) \tag{2.1}$$

Let λ_i be defined by $1/\lambda_i = \int_0^\infty x\, dF_i(x)$, then we have

Theorem 1. If $1/\lambda_i < \infty$, $i=1, \ldots, n$, then the process $(\underline{X}(t), t \geqq 0)$ possesses the unique limiting (stationary) ergodic distribution independently of the initial conditions, namely

$$Q_{0;\, j_1, \ldots, j_n} = \lim_{t\to\infty} Q_{0;\, j_1, \ldots, j_n}(t), \tag{2.2}$$
$$Q_{i_1, \ldots, i_k;\, j_1, \ldots, j_{n-k}}(x_1, \ldots, x_k) = \lim_{t\to\infty} Q_{i_1, \ldots, i_k;\, j_1, \ldots, j_{n-k}}(x_1, \ldots, x_k;\ t).$$

Note that $\underline{X}(t)$ belongs to the class of piecewise-linear Markov processes subject to discontinuous changes. The proof follows immediately from a theorem on page 211 of Gnedenko—Kovalenko's [7] monograph. Theorem 1 provides the existence and uniqueness of the following limits

$$q_{i_1, \ldots, i_k;\, j_1, \ldots, j_{n-k}}(x_1, \ldots, x_k) =$$
$$\lim_{t\to\infty} P(v(t) = k;\ \alpha_1(t) = i_1, \ldots, \alpha_k(t) = i_k;\ x_l \leqq \xi_{i_l} < x_l + dx_l,\ l =$$
$$= \overline{1, k};\ \beta_1(t) = j_1, \ldots, \beta_{n-k}(t) = j_{n-k}) \tag{2.3}$$
$$\text{for} \quad k = 1, \ldots, n,$$

where $q_{i_1, \ldots, i_k;\, j_1, \ldots, j_{n-k}}(x_1, \ldots, x_k)$ denotes a state probability density associated with state $(i_1, \ldots, i_k;\ x_1, \ldots, x_k;\ j_1, \ldots, j_{n-k})$ as $t\to\infty$, $k=1, \ldots, n$. Note that we have assumed here that the ergodic distributions (2.2) of $\underline{X}(t)$ for fixed k have densities, $k=1, \ldots, n$. This assumption is justified if we suppose for simplicity that $F_i(x)$ has density $f_i(x)$, $i=1, \ldots, n$. (cf. Gnedenko—Kovalenko [7] pp. 224). We can make this assumption as in many applications we use distributions having density.

In order to formulate the following theorem introduce a further notation, namely

$$q^*_{i_1, \ldots, i_k; j_1, \ldots, j_{n-k}}(x_1, \ldots, x_k) = \frac{q_{i_1, \ldots, i_k; j_1, \ldots, j_{n-k}}(x_1, \ldots, x_k)}{(1-F_{i_1}(x_1)) \ldots (1-F_{i_k}(x_k))} \tag{2.4}$$

which is the so-called normed density function, $k=1, \ldots, n$. Then we have

Theorem 2. The normed density functions introduced above satisfy the following system of integro-differential equations (2.5,), (2.7) with boundary conditions (2.6), (2.8):

$$\left[\frac{\partial}{\partial x_1} + \ldots + \frac{\partial}{\partial x_k}\right]^* q^*_{i_1, \ldots, i_k; j_1, \ldots, j_{n-k}}(x_1, \ldots, x_k) =$$

$$= -(n-k)\mu q^*_{i_1, \ldots, i_k; j_1, \ldots, j_{n-k}}(x_1, \ldots, x_k) + \int_0^\infty q^*_{i'_1, \ldots, j'_{n-k}, \ldots, i'_k; j_1, \ldots, j_{n-k-1}}$$

$$(x'_1, \ldots, y', \ldots, x'_k) f_{j_n}(y)\, dy \tag{2.5}$$

$$q^*_{i_1, \ldots, i_k; j_1, \ldots, j_{n-k}}(x_1, \ldots, x_{l-1}, 0, x_{l+1}, \ldots, x_k) =$$

$$\mu \sum_{V^{i_l}_{j_1, \ldots, j_{n-k}}} q^*_{i_1, \ldots, i_{l-1}, i_{l+1}, \ldots, i_k; j_1, \ldots, i_l, \ldots, j_{n-k}}(x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_k), \tag{2.6}$$

$$\text{for} \quad l = 1, \ldots, k, \quad 0 \leqq n-k < r.$$

$$\left[\frac{\partial}{\partial x_1} + \ldots + \frac{\partial}{\partial x_k}\right]^* q^*_{i_1, \ldots, i_k; j_1, \ldots, j_{n-k}}(x_1, \ldots, x_k) =$$

$$= -r\mu q^*_{i_1, \ldots, i_k; j_1, \ldots, j_{n-k}}(x_1, \ldots, x_k) + \int_0^\infty q^*_{i'_1, \ldots, j'_{n-k}, \ldots i'_k; j_1, \ldots, {n-k-1}}$$

$$(x'_1, \ldots, y', \ldots, x'_k) f_{j_{n-k}}(y)\, dy, \tag{2.7}$$

$$q^*_{i_1, \ldots, i_k; j_1, \ldots, j_{n-k}}(x_1, \ldots, x_{l-1}, 0, x_{l+1}, \ldots, x_k) =$$

$$\mu \sum_{V^{i_l}_{j_1, \ldots, j_{r-1}}} q^*_{i_1, \ldots, i_{l-1}, i_{l+1}, i_k; j_1, \ldots, i_l, \ldots, j_{n-k}}(x_1, \ldots, x_{l-1}, x_{l+1}, x_k), \tag{2.8}$$

$$\text{for} \quad l = 1, \ldots, k, \quad r \leqq n-k \leqq n-1.$$

$$r\mu Q_{0; j_1, \ldots, j_n} = \int_0^\infty q^*_{j_n; j_1, \ldots, j_{n-1}}(y) f_{j_n}(y)\, dy.$$

Symbol [ ]* will be explained in the proof, while $V^{i_l}_{j_1, \ldots, j_s}$ is defined as follows

$$V^{i_l}_{j_1, \ldots, j_s} = \{(i_l, j_1, \ldots, j_s), (j_1, i_l, j_2, \ldots, j_s), \ldots, (j_1, \ldots, j_s, i_l) \in V^n_{s+1}\}.$$

Proof. Since the process $(\underline{X}(t), t \geqq 0)$ is Markovian its densities must satisfy the Chapman—Kolmogorov equations. A derivation is based on the examination of the sample paths of the process during an infinitesimal interval of width h. The following relations hold:

$$q_{i_1,\dots,i_k;\,j_1,\dots,j_{n-k}}(x_1+h,\dots,x_k+h) = q_{i_1,\dots,i_k;\,j_1,\dots,j_{n-k}}(x_1,\dots,x_k)\times$$

$$\times(1-(n-k)\mu h)\prod_{l=1}^{k}\frac{1-F_{i_l}(x_{l+h})}{1-F_{i_l}(x_l)}+\prod_{l=1}^{k}\frac{1-F_{i_l}(x_l+h)}{1-F_{i_l}(x_l)}\times \tag{2.9}$$

$$\times\int_0^\infty q_{i'_1,\dots,j'_{n-k},\dots,i'_k;\,j_1,\dots,j_{n-k-1}}(x'_1,\dots,y',\dots,x'_k)\frac{f_{i_{n-k}}(y)h\,dy}{1-F_{i_l}(x_l)}+o(h),$$

$$q_{i_1,\dots,i_k;\,j_1,\dots,j_{n-k}}(x_1+h,\dots,x_{l-1}+h,x_{l+1}+h,\dots,x_k+h)h = \prod_{\substack{s=1\\ s\neq l}}^{k}\frac{1-F_{i_s}(x_s+h)}{1-F_{i_s}(x_s)}\times$$

$$\times\mu h\sum_{V^{i_l}_{j_1,\dots,j_{n-k}}} q_{i_1,\dots,i_{l-1}i_{l+1},\dots,i_k;\,j_1,\dots,i_l,\dots,j_{n-k}}(x_1,\dots,x_{l-1},x_{l+1},\dots,x_k)+o(h)$$

$$\text{for}\quad 0\leqq n-k<r,\quad l=1,\dots,k.$$

Similarly

$$q_{i_1,\dots,i_k;\,j_1,\dots,j_{n-k}}(x_1+h,\dots,x_k+h) = q_{i_1,\dots,i_k;\,j_1,\dots,j_{n-k}}(x_1,\dots,x_k)\times$$

$$\times(1-r\mu h)\prod_{l=1}^{k}\frac{1-F_{i_l}(x_l+h)}{1-F_{i_l}(x_l)}+\prod_{l=1}^{k}\frac{1-F_{i_l}(x_l+h)}{1-F_{i_l}(x_l)}\times$$

$$\times\int_0^\infty q_{i'_1,\dots,j'_{n-k},\dots,i'_k;\,j_1,\dots,j_{n-k-1}}(x'_1,\dots,y',\dots,x'_k)\frac{f_{j_{n-k}}(y)h\,dy}{1-F_{j_{n-k}}(y)}+o(h), \tag{2.10}$$

$$q_{i_1,\dots,i_k;\,j_1,\dots,j_{n-k}}(x_1+h,\dots,x_{l-1}+h,0,x_{l+1}+h,\dots,x_k+h)h = \prod_{\substack{s=1\\ s\neq l}}^{k}\frac{1-F_{i_s}(x_s+h)}{1-F_{i_s}(x_s)}\times$$

$$\times\mu h\sum_{V^{i_l}_{j_1,\dots,j_{r-1}}} q_{i_1,\dots,i_{l-1},i_{l+1},\dots,i_k;\,j_1,\dots,i_l,\dots,j_{n-k}}(x_1,\dots,x_{l-1},x_{l+1},\dots,x_k)+o(h)$$

$$\text{for}\quad r\leqq n-k\leqq n-1,\quad l=1,\dots,k.$$

Finally

$$Q_{0;\,j_1,\dots,j_n} = Q_{0;\,j_1,\dots,j_n}(1-r\mu h)+\int_0^\infty q_{j_n;\,j_1,\dots,j_{n-1}}(y)\frac{f_{j_n}(y)h\,dy}{1-F_{j_n}(y)}+o(h). \tag{2.11}$$

Hence the derivation of eq. (2.5), (2.7) and boundary conditions (2.6), (2.8) is quite simple. Indeed, dividing the lefthand side of eq. (2.9), (2.10), (2.11) by the factor $\prod_{l=1}^{k}(1-F_{i_l}(x_l+h))$, taking into account the definition of the normed densities (2.4) and taking the limits as $h\to 0$ we get the desired result.

In the lefthand side of (2.5), (2.7), used for the notation of the limit in the righthand side, the usual notation for partial differential quotients has been applied. Strictly speaking this is not allowed since the existence of the individual partial differential quotients is not assured. This is why the operator is notated by []*. Actually this is a $(1, 1, \ldots, 1)\in R^k$ directional derivative. (See Cohen [4] pp. 252).

In the following we solve eq. (2.5), (2.7) subject to boundary conditions (2.6), (2.8) to determine the ergodic distribution

$$(Q_{0;\, j_1, \ldots, j_n}, Q_{i_1, \ldots, i_k;\, j_1, \ldots, j_{n-k}}),$$
$$(i_1, \ldots, i_k)\in C_k^n, \quad (j_1, \ldots, j_{n-k})\in V_{n-k}^n, \quad k = 1, \ldots, n.$$

If we set

$$Q_{0;\, j_1, \ldots, j_n} = C_0,$$
$$q^*_{i_1, \ldots, i_k;\, j_1, \ldots, j_{n-k}}(x_1, \ldots, x_k) = c_k, \quad k = 1, \ldots, n,$$

then it can be shown by substitution that they satisfy the eq. (2.5), (2.7), and boundary conditions (2.6), (2.8). Moreover the sequence $\{c_k\}$ can be obtained in succession and expressed in a neat form by the help of c_n. Using the relations (2.5), (2.6), (2.7), (2.8) it is easy to see that

$$c_k = (r!\, r^{n-r-k} \mu^{n-k})^{-1} \cdot c_n \quad \text{for} \quad 0 \leqq k \leqq n-r,$$

and, similarly

$$c_k = ((n-k)!\, \mu^{n-k})^{-1} \cdot c_n \quad \text{for} \quad n-r \leqq k \leqq n.$$

Since these equations completly describe the system, this is the required solution.

Let $Q_{i_1, \ldots, i_k;\, j_1, \ldots, j_{n-k}}$ denote the steady state probability that jobs with indices $(i_1, \ldots, i_k)$ are at the peripheral devices and the order of the rest arrival to the CPU-s is $(j_1, \ldots, j_{n-k})$. Furthermore, denote by $Q_{i_1, \ldots, i_k}$ the steady state probability that programs with indices $(i_1, \ldots, i_k)$ are processing I/O operations. It can be verified that

$$Q_{i_1, \ldots, i_k;\, j_1, \ldots, j_{n-k}} = (\lambda_{i_1} \ldots \lambda_{i_k})^{-1} c_k, \quad \text{for} \quad k = 1, \ldots, n. \tag{2.12}$$

Using the relations for c_k we get

$$Q_{i_1, \ldots, i_k} = (n-k)!\, [r!\, r^{n-k-r} \mu^{n-k} \cdot \lambda_{i_1} \ldots \lambda_{i_k}]^{-1} \cdot c_n, \tag{2.13}$$
$$(i_1, \ldots, i_k)\in C_k^n, \quad k = 0, 1, \ldots, n-r.$$

Similarly

$$Q_{i_1, \ldots, i_k} = [\mu^{n-k} \cdot \lambda_{i_1} \ldots \lambda_{i_k}]^{-1} \cdot c_n, \tag{2.14}$$
$$(i_1, \ldots, i_k)\in C_k^n, \quad k = n-r, \ldots, n.$$

Let $\hat{Q}_k$ and $\hat{P}_l$ denote the steady state probabilities that k jobs are staying at the peripheral devices and l jobs need service at the CPU-s, respectively. Clearly

$$Q_{i_1, \ldots, i_n} = Q_{1, \ldots, n} = \hat{Q}_n = \hat{P}_0, \quad \hat{Q}_k = \hat{P}_{n-k},$$
$$\text{for} \quad k = 0, \ldots, n.$$

It is easy to see that

$$C_n = \hat{Q}_n(\lambda_1 \ldots \lambda_n),$$

and

$$\hat{Q}_k = \sum_{(i_1, \ldots, i_k)\in C_k^n} Q_{i_1, \ldots, i_k},$$

where $\hat{Q}_n$ can be obtained with the aid of the norming condition

$$\sum_{k=0}^{n} \hat{Q}_k = 1.$$

In the homogenenous case relations (2.13), (2.14) yield

$$\hat{Q}_k = \frac{n!}{k!\,r!\,r^{n-k-r}} \cdot \left(\frac{\lambda}{\mu}\right)^{n-k} \cdot \hat{Q}_n, \quad \text{for} \quad 0 \leqq k \leqq n-r,$$

$$\hat{Q}_k = \binom{n}{k}\left(\frac{\lambda}{\mu}\right)^{n-k} \cdot \hat{Q}_n, \quad \text{for} \quad n-r \leqq k \leqq n.$$

Thus

$$\hat{P}_k = \binom{n}{k}\left(\frac{\lambda}{\mu}\right)^{k} \cdot \hat{P}_0, \quad \text{for} \quad 0 \leqq k \leqq r, \quad \text{and}$$

$$\hat{P}_k = \frac{n!}{(n-k)!\,r!\,r^{k-r}} \left(\frac{\lambda}{\mu}\right)^{k} \cdot \hat{P}_0, \quad \text{for} \quad r \leqq k \leqq n.$$

This is exactly the same result obtained by Bunday and Scraton [3]. The equivalence of the $E_k/M/1$ and $M/M/1$ models and that of $G/M/r$ and $M/M/r$ models as noted by Benson, Bunday and Scraton, respectively, is just a special case of our more general result obtained here.

Before determining the operational characteristics of the system we need one more theorem. In order to formulate it we introduce some notations. Let $Q^{(i)}(P^{(i)})$ denote the stationary probability that job i is processing I/O operation (need service at the CPU-s) for $i=1, \ldots, n$. It is clear that the process $(\underline{Y}(t), t \geqq 0)$ is semi-Markovian with state space

$$\bigcup_{\substack{(i_1,\ldots,i_k)\in C_k^n,\ (j_1,\ldots,j_{n-k})\in V_{n-k}^n \\ \{i_1,\ldots,i_k\}\cap\{j_1,\ldots,j_{n-k}\}=\emptyset \\ k=0,1,\ldots,n,}} \{(i_1, \ldots, i_k;\ j_1, \ldots, j_{n-k})\}.$$

Let H_i be the event that job i is processing I/0 operation and $Z_{H_i}(t)$ its indicator function i.e.

$$Z_{H_i}(t) = \begin{cases} 1 & \text{if} \quad \underline{Y}(t) \in H_i, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 4.

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T Z_{H_i}(t)\,dt = \frac{1/\lambda_i}{1/\lambda_i + W_i + 1/\mu} = Q^{(i)} = 1 - P^{(i)},$$

where W_i denotes the mean waiting time of program i.

Proof. The statement is a special case of a theorem concerning the expected sojourn time for semi-Markov processes, see Tomkó [12] pp. 297.

3. The main characteristics of the system

(i) *Utilizations.* Utilizations can now be considered for individual servers or for the system as a whole. The process $\underline{X}(t)$ is assumed to be in equilibrium. Considering the system as a whole it will be empty only when there are no jobs at the CPU-s and will be busy at other times. As usual, using renewal-theoretic arguments for the system utilization we have $U=1-\hat{Q}_n$, and

$$\hat{Q}_n = \frac{M\eta^*}{M\eta^*+M\delta},$$

where $\eta^*=\min(\eta_1, \ldots, \eta_n)$, the random variable η_i denotes the I/0 times of program i, $i=1, \ldots, n$, and $M\delta$ means the average busy period of the system, respectively. Thus the expected busy period lenght is given by

$$M\delta = M\eta^* \cdot \frac{1-\hat{Q}_n}{\hat{Q}_n}.$$

Specially, if $F_i(x)=1-\exp(-\lambda_i x)$, $i=1, \ldots, n$ we get $M\delta=(1-\hat{Q}_n)\left(\hat{Q}_n \cdot \sum_{i=1}^{n} \lambda_i\right)^{-1}$. It is easy to see that for CPU utilization the following relation holds:

$$U_{\text{CPU}} = \frac{1}{r}\left(\sum_{k=1}^{r} k\hat{P}_k + r \sum_{k=r+1}^{n} \hat{P}_k\right) = \frac{\bar{r}}{r},$$

where $\bar{r}$ denotes the mean number of busy CPU-s.

(ii) *Mean waiting times.* By the virtue of Theorem 4 we have

$$Q^{(i)} = (1/\lambda_i)(1/\lambda_i+W_i+1/\mu)^{-1}.$$

Consequently, the expected waiting time of job i is

$$W_i = \frac{1}{\lambda_i}\frac{1-Q^{(i)}}{Q^{(i)}} - \frac{1}{\mu}, \quad \text{for} \quad i = 1, \ldots, n.$$

It follows that the mean response time of program i, that is, the waiting and CPU time together, can be obtained by

$$T_i = W_i+1/\mu = (1-Q^{(i)})(\lambda_i Q^{(i)})^{-1}, \quad \text{for} \quad i = 1, \ldots, n. \tag{3.1}$$

Since

$$\sum_{i=1}^{n}(1-Q^{(i)}) = \bar{n}$$

where $\bar{n}$ denotes the mean number of jobs staying at the CPU-s, by reordering and adding (3.1) we have

$$\sum_{i=1}^{n} \lambda_i T_i Q^{(i)} = \bar{n} \tag{3.2}$$

which is Little's formula for the finite-source $\vec{G}/M/r$ queue. In particular, if $F_i(x)=F(x)$, $i=1, \ldots, n$, (3.2) can be written as $\lambda T\bar{Q}=\bar{n}$ where $\bar{Q}$ denotes the average number of programs processing I/O operations.

Acknowledgement. I am very grateful to Dr. J. Tomkó for his helpful discussion. My special thanks are also due to Prof. M. Arató for providing valuable comments and very thorough readings of earlier version of the paper which have greatly improved the presentation.

Abstract

The aim of the present paper is to give a new queueing model for a multiprogrammed computer system, where r CPU-s serve the jobs according to the FIFO discipline. The programs are stochastically different, job i is characterised by exponentially distributed CPU time with rate μ and I/O time with an arbitrary distribution function $F_i(x)$ possessing density $f_i(x)$. In steady state we deal with the main performance measures, such as CPU utilization, mean waiting and response times of the jobs.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF DEBRECEN
PF. 12, DEBRECEN, HUNGARY
H—4010

References

[1] ASZTALOS, D., Finite source queueing systems and their applications to computer systems, Alk. Mat. Lapok, 5 (1979), 89—101 (in Hungarian).
[2] AVI-ITZHAK, B. and D. P. HEYMAN, Approximate Queuing Models for Multiprogramming Computer Systems, Opns. Res. 21 (1973), 1212—1230.
[3] BUNDAY, B. D. and R. E. SCRATON, The $G/M/r$ machine interference model, Eur. J. Operation Res., 4 (1980), 399—402.
[4] COHEN, J. W., The multiple phase service network with generalised processor sharing, Acta Informatica, 12 (1979), 245—284.
[5] CSIGE, L. and J. TOMKÓ, The machine interference for exponentially distributed operating and repair times, Alk. Mat. Lapok, 8 (1982), 107—124 (in Hungarian).
[6] GAVER, D. P., Probability models for multiprogramming computer systems, J. ACM, 3 (1967), 423—438.
[7] GNEDENKO, B. V. and J. N. KOVALENKO, Introduction to Queueing Theory, Nauka, Moscow, 1966 (in Russian).
[8] KAMEDA, H., A Finite-Source Queue with Different Customers, J. ACM, 29 (1982), 478—491.
[9] KLEINROCK, L., Queueing Systems, Vol. 2: Computer Applications, Wiley-Interscience, New York, 1976.
[10] SCHATTE, P., On the Finite Popoulation $GI/M/1$ Queue and its Application to Multiprogrammed Computers, Journal of Information Processing and Cybernetics, 16 (1980), 433—441.
[11] SZTRIK, J., Probability model for non-homogeneous multiprogramming computer system, Acta Cybernetica, 6 (1983), 93—101.
[12] TOMKÓ, J., On sojourn times for semi-Markov processes, 14-th European Meeting of Statisticians, Wroclaw, Poland, 1981.

(Received Oct. 19, 1983)