Problems of Information Transmission, 1990

12. G. I. Falin, "Comparability of migration processes," Teor. Veroyatn. Primen., 33, No. 2, 392-396 (1988).
13. G. I. Ivchenko, V. A. Kashtanov, and I. N. Kovalenko, Queueing Theory [in Russian], Vysshaya Shkola, Moscow (1982).

ASYMPTOTIC BEHAVIOR OF THE STATIONARY DISTRIBUTION FOR A CLOSED QUEUEING SYSTEM

A. L. Stolyar

UDC 621.395.74:519.27

We consider a closed queueing system consisting of M identical servers with fixed unit service time. The number of customers is fixed and equal to N. Each served customer is instantaneously routed with equal probability to one of M servers in the system (or is enqueued if the server is busy). An asymptotic result is proved for the stationary distribution of the queueing process as $N, M \to \infty$, $N/M \to \nu = const$, and also a result on deterministic approximation of the process on a finite time interval.

1. INTRODUCTION

Throughput analysis of computing systems and networks often involves queueing system models for which explicit expressions of the operating characteristics are not readily derivable. If the system is large, numerical analysis is also quite complex, and it is therefore natural to investigate the asymptotic behavior of the characteristics of such systems with system size increasing to infinity.

The queueing system considered in this paper models a real-life situation. A multiprocessor computing system consists of N identical processors and M identical memory modules, with each processor connected to each memory module. The memory operates in fixed-length cycles. Each memory module can process one request from any processor in one cycle. We assume that the processor, after its current request has been processed, immediately sends another request to one of the memory modules with equal probability. The throughput of the system is determined by the average number of requests processed in memory in one cycle, which in turn depends on the probability (in the stationary mode) that the request occupies one memory module (by symmetry, this may be any of the modules). The distribution of the waiting time to receive an answer to a request also depends on the distribution of the number of requests being processed and awaiting processing by some fixed memory module.

Other interpretations of the model are also possible. For example, the model may comprise N user terminals in a computer network and M host computers processing their requests, etc.

Statement of the Problem

We consider a symmetric closed queueing system with M servers. The number of customers in the system is fixed and equal N. The service time on each server is fixed and equal 1, and each server has infinitely many waiting places in its queue. At time $n = 0$, the customers are arbitrarily allocated to servers and service still has not begun. Each customer leaving a server is instantaneously directed with equal probability to one of the servers, including the one that it has just left. This system obviously changes its state (allocation of customers to servers) only at discrete time moments $n = 0,1,2,\ldots$. For definiteness, we define the state of the system at the moment n as its state at the moment $(n + 0)$. The system actually functions as follows. Let the state at time n be given. Extract one customer from each nonempty server, and then distribute all the extracted customers with equal probability between the servers. The result is the state of the system at time $n + 1$, and so on. This interpretation of the process is the most useful for our purposes.

Translated from Problemy Peredachi Informatsii, Vol. 25, No. 4, pp. 80-92, October-December, 1989. Original article submitted February 18, 1988.

0032-9460/89/2504-0321$12.50

Let $p^{N,M} = (p_i^{N,M})_{i\geq 0}$ be the distribution of a random number of customers on some designated server (say, the first server) in the stationary mode for fixed N and M. It is required to find the asymptotic behavior of the distribution $p^{N,M}$ as N, $M \to \infty$, $N/M \to \nu > 0$.

The problem of finding the distribution $p^{N,M}$ has been considered by many authors, and a fairly comprehensive bibliography can be found in [1, 2]. So far, however, no explicit expression has been obtained for $p^{N,M}$ for arbitrary N, M. Most authors propose heuristic formulas for estimating the quantity $(1 - p_0^{N,M})$, which is of practical interest. In [3] it is also suggested from heuristic considerations that

$$p^{N,M} \to p, \quad N, M \to \infty, \quad N/M \to \nu, \tag{1.1}$$

where the distribution p is specified explicitly and convergence is coordinatewise.

In this paper, we provide a rigorous proof of a stronger proposition (Theorem 1), from which (1.1) is obtained as a corollary. As an intermediate result, Theorem 2 supplies a deterministic approximation of the stochastic queueing process on a finite time interval. The method used in the proof of Theorem 1 is also applicable to the analysis of other, more complex queueing systems and networks (see [4]) and is therefore of independent interest.

2. THE MAIN RESULT ON THE ASYMPTOTIC BEHAVIOR OF THE STATIONARY DISTRIBUTION. OUTLINE OF THE PROOF. DETERMINISTIC APPROXIMATION ON A FINITE TIME INTERVAL

In what follows, we assume for simplicity that the system is identified by a single parameter N, and that $M = M(N)$ is a function of N such that $N/M(N) \to \nu = \text{const} > 0$ as $N \to \infty$.

Let N (and thus also $M = M(N)$) be fixed. Denote by $M_i^N(n)$, $i \geq 0$, the total number of servers that are occupied at the moment n, $n \geq 0$, by precisely i customers and let

$$D_i^N(n) = M_i^N(n)/M \quad i\geq 0, \quad n\geq 0, \tag{2.1}$$

i.e., $D_i^N(n)$ is the proportion of servers occupied by precisely i customers at the moment n. Note that for any $n \geq 0$ with probability 1

$$D_i^N(n) = 0, \quad i>N; \quad \sum_{i=0}^{\infty} D_i^N(n) = 1; \quad \sum_{i=0}^{\infty} iD_i^N(n) = \frac{N}{M}. \tag{2.2}$$

The sequence of (infinite-dimensional) vectors $D^N = (D^N(n))_{n\geq 0}$, where $D^N(n) = (D_i^N(n))_{i\geq 0}$, form a Markov chain with a finite (by (2.1) and (2.2)) state set

$$S^N = \{x = (x_i)_{i\geq 0} \in R^\infty \mid x_i = m_i/M, \quad m_i \in Z_+;$$
$$x_i = 0, \quad i>N; \quad \sum_{i=0}^{\infty} x_i = 1; \quad \sum_{i=0}^{\infty} ix_i = \frac{N}{M}\Big\}. \tag{2.3}$$

It is easy to see that the finite Markov chain D^N is ergodic. Thus, it has a stationary distribution $Q^N = \{Q^N(x), x \in S^N\}$. To the distribution Q^N we associate the probability measure Q^N defined on a measurable space $(E, \mathscr{E})$:

$$Q^N(B) = \sum_{x\in B\cap S^N} Q^N(x), \quad B \in \mathscr{E}, \tag{2.4}$$

where $E = \{x \in R^\infty \mid x_i \geq 0,\ i \geq 0,\ \sum_{i=0}^{\infty} x_i = 1,\ \sum_{i=0}^{\infty} ix_i \leq \nu_*\}$, $\nu_* = \sup N/M(N)$, $N \geq 1$, and $\mathscr{E}$ is the σ-algebra generated by the metric

$$d(x, y) = \sum_{i=0} \frac{|x_i - y_i|}{2^i}, \quad x, y \in E.$$

The measure Q^N can be interpreted as ergodic distribution of the process D^N, treated as a process with values in $(E, \mathscr{E})$, because $S^N \subset E$, $\forall N$.

Denote by $\delta^\rho = (\delta_i^\rho)_{i\geq 0}$, $0 \leq \rho < 1$, the probability distribution on the set $Z_+ = \{0, 1, 2, \ldots\}$, $\delta_i^\rho \geq 0$, $i \geq 0$, $\sum_{i=0}^{\infty} \delta_i^\rho = 1$, such that its generating function $F(z) = \sum_{i=0}^{\infty} \delta_i^\rho z^i$, $|z| \leq 1$, satisfies the equation

$$F(z)=[(F(z)-F(0))/z+F(0)]e^{-\rho(1-z)},$$

whence

$$F(z)=(1-\rho)(z-1)e^{-\rho(1-z)}/(z-e^{-\rho(1-z)}).$$

Note that as a numerical sequence $\delta^\rho \in R^\infty$. Let

$$\mu(\rho)=\sum_{i=0}^{\infty} i\delta_i^\rho=\rho+\frac{\rho^2}{2(1-\rho)}, \quad 0\leqslant\rho<1,$$

and by $\rho(\mu)$ denote the inverse of the function $\mu(\rho)$:

$$\rho(\mu)=1+\mu-\sqrt{1+\mu^2}, \quad 0\leqslant\mu<\infty.$$

THEOREM 1. As $N \to \infty$,

$$Q^N \xrightarrow{w} Q_{\rho(\nu)},$$

where Q_ρ, $0 \le \rho < 1$, is the Dirac measure concentrated at the point $x = \delta^\rho$, $Q_\rho\{\delta^\rho\} = 1$. (Here and in what follows, $\xrightarrow{w}$ denotes weak convergence.)

If $X^N(n)$ is the random number of customers on the first server at the moment n for a fixed N, and $p^N = (p_i^N)_{i\ge0}$ is the ergodic distribution of $X^N(n)$, i.e.,

$$p_i^N=\lim_{n\to\infty} \mathrm{P}(X^N(n)=i), \quad i\geqslant0,$$

then by symmetry of the servers

$$\int_E x_i\, dQ^N=p_i^N, \quad i\geqslant0.$$

By Theorem 1, $Q^N \xrightarrow{w} Q_{\rho(\nu)}$. Now, since x_i is a continuous function of $x \in E$,

$$\lim_{N\to\infty} p_i^N=\int_E x_i\, dQ_{\rho(\nu)}=\delta_i^{\rho(\nu)}, \quad i\geqslant0.$$

Thus, Theorem 1 leads to

COROLLARY 1. For $N \to \infty$, $p^N \to \delta^{\rho(\nu)}$.

We introduce some notation. Let

$$A=\left\{x\in R^\infty \mid x_i\geqslant0,\ i\geqslant0,\ \sum_{i=0}^{\infty} x_i=1\right\},$$

$$A^\nu=\left\{x\in A \,\Big|\, \sum_{i=0}^{\infty} ix_i=\nu\right\}; \quad A^\nu\subset E\subset A.$$

Each element $x \in A$ may be interpreted as a distribution on the set Z_+, where x_i is the measure of the number $i \in Z_+$. Then each element $x \in A^\nu$ may be regarded as a distribution on Z_+ with mean ν. On the set A define a standard stochastic order relation, specifically, $x\prec y$, x, $y\in A$, if

$$\sum_{j=i}^{\infty} x_j \leqslant \sum_{j=i}^{\infty} y_j, \quad i\geqslant0.$$

Let U and W be random variables with values in Z_+, and $u = (u_i)_{i\ge0}$ and $w = (w_i)_{i\ge0}$ their distributions. Then, by definition, $U\prec W$, if $u\prec w$. Let us outline the proof.

The metric space E is complete and separable. Moreover, E is a compact set, whence it follows that the family of measures $\{Q^N\}$ is dense, and therefore (by Prokhorov's theorem [5]) relatively compact. Then there exists a probability measure Q on $(E, \mathscr{E})$ such that

$$Q^{N_k} \xrightarrow{w} Q, \quad N_k\to\infty,$$

where $\{Q^{N_k}\}_{k\geqslant1}\subseteq\{Q^N\}$. To avoid overcomplicated notation, we assume in what follows that $Q^N \xrightarrow{w} Q$. The proof of Theorem 1 consists of the following propositions.

Proposition 1. The measure Q is concentrated on the set A^ν, i.e.,

$$Q(A^\nu)=1.$$

Let the transformation $T\colon\ A\mapsto A$ be defined as follows ($y = Tx$):

$$\begin{cases} y_0=(x_0+x_1)\,e^{-(1-x_0)}, \\ y_i=(x_0+x_1)\,\dfrac{(1-x_0)^i}{i!}\,e^{-(1-x_0)}+ \\ +\displaystyle\sum_{j=1}^{i} x_{j+1}\,\dfrac{(1-x_0)^{i-j}}{(i-j)!}\,e^{-(1-x_0)}, \qquad i\geq 1. \end{cases} \tag{2.5}$$

It is easy to see that:

a) T is continuous on E;

b) the sets E and A^ν are closed relative to T, i.e.,

$$T(E)\subseteq E,\quad T(A^\nu)\subseteq A^\nu.$$

Proposition 2. The transformation T preserves the measure Q, i.e.,

$$Q(T^{-1}B)=Q(B),\quad \forall B\in\mathscr{E}.$$

Proposition 3. For any $x \in A^\nu$,

$$T^n x\to\delta^{\rho(\nu)},\quad n\to\infty.$$

Proposition 4. The measure Q coincides with $Q_{\rho(\nu)}$, i.e.,

$$Q\{\delta^{\rho(\nu)}\}=1.$$

We will show that Propositions 1-3 imply Proposition 4. Consider the set $J_\varepsilon=\{x\in A^\nu \mid d(x,\delta^{\rho(\nu)})>\varepsilon\}$. Then by Proposition 3 and Poincaré recurrence theorem [5, p. 392]

$$Q(J_\varepsilon)=0,\quad \forall\varepsilon>0,$$

whence

$$Q\{\delta^{\rho(\nu)}\}=\lim_{k\to\infty} Q(A^\nu\backslash J_{1/k})=1,$$

Q.E.D.

The auxiliary Lemma 5 (see Sec. 6) and uniform continuity of T on E (E is a compact set!) lead to

THEOREM 2. For any integer $l \geq 0$ and any $\varepsilon > 0$

$$\lim_{N\to\infty}\sup_{x\in S^N} P\,[\sup_{0\leq k\leq l}\{d(D^N(k),T^k x)\}>\varepsilon \mid D^N(0)=x]=0.$$

This theorem shows that for large N the Markov process $(D^N(n))_{n\geq 0}$ is close to the deterministic process $(T^n x)_{n\geq 0}$ uniformly in the initial state $D^N(0) = x \in S^N$ on a finite time interval.

3. PROOF OF PROPOSITION 1

Let the following condition hold. There exists a distribution $r = (r_i)_{i\geq 0} \in A$ and $N_* \in Z_+$ such that for all $N > N_*$

$$p^N\prec r, \tag{3.1}$$

$$\sum_{i=0}^{\infty} i r_i<\infty. \tag{3.2}$$

Recall that p^N is the ergodic distribution of the random number X^N of customers on the first server, when N (and therefore M) is fixed.

By (2.3) and (2.4)

$$\sum_{i=0}^{\infty} ix_i = \frac{N}{M} \quad (Q^N\text{-a. s.}). \tag{3.3}$$

Since $N/M \to \nu$,

$$\sum_{i=0}^{\infty} ix_i \leqslant \nu \quad (Q\text{-a. s.}). \tag{3.4}$$

Indeed, for any $\varepsilon > 0$, the set $J_\varepsilon' = \left\{x \in E \,\middle|\, \sum_{i=0}^{\infty} ix_i \leqslant \nu + \varepsilon\right\}$ is closed. From (3.3) it follows that for all large N, $Q^N(J_\varepsilon') = 1$, and therefore $Q(J_\varepsilon') = 1$. Integrating the expression in the left-hand side of (3.4), we have

$$\int_E \left(\sum_{i=0}^{\infty} ix_i\right) dQ = \sum_{i=0}^{\infty} ip_i,$$

where

$$p_i = \int_E x_i \, dQ = \lim_{N\to\infty} \int_E x_i \, dQ^N = \lim_{N\to\infty} p_i^N, \quad i \geqslant 0,$$

i.e., $p = \lim p^N$, $N \to \infty$.

If to each distribution p^N we associate a random variable with values in Z_+ that has this distribution, then the resulting family of random variables is uniformly integrable by (3.1) and (3.2). Then

$$\sum_{i=0}^{\infty} ip_i = \lim_{N\to\infty} \sum_{i=0}^{\infty} ip_i^N = \nu.$$

Thus,

$$\int_E \left(\sum_{i=0}^{\infty} ix_i\right) dQ = \nu.$$

This and (3.4) give

$$\sum_{i=0}^{\infty} ix_i = \nu \quad (Q\text{-a.s.}).$$

The last equality proves Proposition 1.

We will show that (3.1) and (3.2) indeed hold. To this end, construct a queueing process $(Z(n))_{n\geq 0}$ that majorizes each of the processes $(X^N(n))_{n\geq 0}$ for all N greater than some N_*.

For the sequence $(X^N(n))_{n\geq 0}$ we have the following recurrence:

$$X^N(n) = (X^N(n-1) - 1)^+ + U_n, \quad n \geqslant 1,$$

where $(a)^+ = \max\{a, 0\}$ and U_n is the random number of customers arriving at the first server at the moment n. Lemma 4 (Sec. 6) claims that there exist integers k, $N_* > 0$, and an integer-valued nonnegative random variable V such that for any $N > N_*$ and $n \geq 0$, independently of the values of $X^N(j)$, U_j, $j = 0, \ldots, n$,

$$U_{n+1} + \ldots + U_{n+k} < V, \tag{3.5}$$

and $MV < k$.

Define the process $(Z(n))_{n\geq 0}$ recursively by

$$Z(n) = \max\{Z(n-1) - 1, k\} + V_n, \quad n \geqslant 1,$$

where $\{V_n, n \geq 1\}$ are jointly independent random variables,

$$V_n \overset{d}{=} \begin{cases} V, & n = kl + 1, \quad l = 0, 1, 2, \ldots, \\ 0 & \text{otherwise}, \end{cases}$$

where $\stackrel{d}{=}$ is equality in distribution.

The process $(Z(n))_{n\geq 0}$ is a homogeneous Markov chain, and if $r(n)$ is the distribution of $Z(n)$, then

$$r(kl+j)\to r^j,\quad l\to\infty,\quad j=\overline{1,\ k}. \tag{3.6}$$

Indeed, ergodicity of the embedded Markov chain $(Z(kl+j))_{l\geq 0}$ is easily established by Mustafa's criterion [6]. Now, if we hold the initial values $X^N(0) = Z(0) = 0$ fixed, then the processes $(X^N(n), D^N(n))_{n\geq 0}$ and $(Z(n))_{n\geq 0}$ can be constructed by (3.5) on one probability space so that

$$X^N(n)\leqslant Z(n)\quad \text{(a.s.)}.$$

Hence it follows that

$$p^N(n)\prec r(n),\quad n\geqslant 0.$$

Then from $p^N(n)\to p^N$, $n\to\infty$ and (3.6) we obtain

$$p^N\prec r^j,\quad j=\overline{1,\ k}.$$

Thus, for $r = r^1$, condition (3.1) is satisfied. Let us prove (3.2). Let $Y(l) \equiv Z(kl + 1) - k$, $l \geq 0$, and let $\tilde{r}$ be the stationary distribution of the process $(Y(l))_{l\geq 0}$. The relationship

$$\sum_{i=0}^{\infty} ir_i^1=\sum_{i=0}^{\infty} i\tilde{r}_i+k$$

is obvious and therefore it remains to show that $\sum i\tilde{r}_i<\infty$.

Denote by

$$\Pi(z)=\sum_{i=0}^{\infty}\tilde{r}_iz^i,\qquad |z|\leqslant 1,$$

the generating function of the distribution $\tilde{r}$, and let

$$R(z)=\sum_{i=0}^{k-1}\tilde{r}_iz^i,\qquad z^kW(z)=\sum_{i=k}^{\infty}\tilde{r}_iz^i.$$

Then the stationarity condition of $\tilde{r}$ has the form

$$\Pi(z)=R(z)+z^kW(z)=(R(1)+W(z))V(z),$$

where $V(z)$ is the generating function of V, whose explicit form is known from Lemma 4 (Sec. 6). Hence

$$W(z)=(R(z)-R(1)V(z))/(V(z)-z^k)=\Phi(z)/\Psi(z).$$

Given the explicit form of $V(z)$, we can easily verify that $\Phi(z)=(z-1)\tilde{\Phi}(z)$, $\Psi(z)=(z-1)\tilde{\Psi}(z)$, , where $\tilde{\Phi}(z)$ and $\tilde{\Psi}(z)$ are analytical in the neighborhood of 1, and $\tilde{\Psi}(1)\neq 0$. Then $W'(1 - 0) < \infty$, whence $\Sigma i\tilde{r}_i = \Pi'(1) < \infty$, as required. Q.E.D.

4. PROOF OF PROPOSITION 2

We will show that if $B\in\mathscr{B}$ is closed then

$$Q(T^{-1}B)=Q(B). \tag{4.1}$$

Let

$$B_\varepsilon=\bigcup_{x\in B}O_\varepsilon(x),$$

where $O_\varepsilon(x)$ is an open ε-neighborhood of the point x.

By continuity of T, there exists an open set $C \in \mathscr{B}$ such that

$$T^{-1}(B)\subset C,\quad T(C)\subseteq B_\varepsilon,\quad \varepsilon>0.$$

Fix ε, $\alpha > 0$. Then we can choose N_* such that for $N > N_*$

$$\mathbf{P}(d(D^N(n+1), Tx) \leqslant \varepsilon \mid D^N(n) = x) \geqslant 1 - \alpha \tag{4.2}$$

uniformly in $x \in S^N \subset E$ (see Sec. 6, Lemma 5).

Seeing that $[B_{2\varepsilon}]$ is closed and C is open, $Q^N \xrightarrow{w} Q$, where Q^N is the stationary distribution of the process $(D^N(n))_{n\geq 0}$, we obtain using (4.2)

$$Q([B_{2\varepsilon}]) \geqslant \overline{\lim_{N\to\infty}} Q^N([B_{2\varepsilon}]) \geqslant \overline{\lim} Q^N(C)(1-\alpha) \geqslant \underline{\lim} Q^N(C)(1-\alpha) \geqslant Q(C)(1-\alpha) \geqslant Q(T^{-1}B)(1-\alpha).$$

Since ε and α are arbitrary, and $[B_{2\varepsilon}] \downarrow B$, $\varepsilon \downarrow 0$, we obtain

$$Q(T^{-1}B) \leqslant Q(B). \tag{4.3}$$

But $\overline{B}_\varepsilon$ is closed and $\overline{B}_\varepsilon \uparrow \overline{B}$, $\varepsilon \downarrow 0$. Replacing B with $\overline{B}_\varepsilon$ in (4.3) and passing to the limit as $\varepsilon \downarrow 0$, we obtain

$$Q(T^{-1}\overline{B}) \leqslant Q(\overline{B}). \tag{4.4}$$

From (4.3) and (4.4) we obtain (4.1).

Since $(E, \mathscr{E})$ is a complete separable metric space with the σ-algebra generated by open sets, then for any $H \in \mathscr{E}$ and any $\varepsilon > 0$ there exists a compact set $B \in H$ such that $Q(H \setminus B) < \varepsilon$. Hence we easily obtain that $Q(T^{-1}H) = Q(H)$.

5. PROOF OF PROPOSITION 3

In this section, we will first prove Proposition 4 using only Propositions 1 and 2, and then prove Proposition 3. The proof of Proposition 4 is of independent interest (see Remark 2), and moreover it contains some auxiliary definitions and constructions that are used in the proof of Proposition 3.

The transformation T may be expressed in generating functions, i.e., if $F(z) = \sum_{i=0}^{\infty} x_i z^i$, $|z| \leqslant 1$ is the generating function of x, $x \in A$, and $(TF)(z)$ is the generating function of Tx, then from (2.5) we have

$$(TF)(z) = [(F(z) - F(0))/z + F(0)]\, e^{-(1-F(0))(1-z)}.$$

Consider (also on A) the transformation T_ρ, $0 \leq \rho < 1$, which is expressible in generating functions as

$$(T_\rho F)(z) = [(F(z) - F(0))/z + F(0)]\, e^{-\rho(1-z)},$$

where F and $T_\rho F$ are the generating functions of x and $T_\rho x$, respectively. It is easy to see that the transformation T_ρ indeed maps A to A. It has been well studied in queueing theory, because it describes the embedded Markov chain for the $M|D|1$ system with traffic ρ. We know (see, e.g., [7]) that the transformation T_ρ has in A a unique fixed point $x = \delta^\rho$, $T_\rho\delta^\rho = \delta^\rho$, which is the definition of δ^ρ in Sec. 2. Here $\delta_0^\rho = 1 - \rho$. Moreover, for any $x \in A$,

$$T_\rho^n x \to \delta^\rho, \quad n \to \infty. \tag{5.1}$$

Since $Tx \equiv T_{1-x_0}x$, the fixed point δ^ρ of the transformation T_ρ is a fixed point for T:

$$T\delta^\rho = \delta^\rho, \quad 0 \leqslant \rho < 1. \tag{5.2}$$

It is easy to see that the distribution δ^ρ and also the transformations T and T_ρ have the following properties of monotonicity and continuity:

$$\rho_1 \leqslant \rho_2 \Rightarrow \delta^{\rho_1} \prec \delta^{\rho_2}. \tag{5.3}$$

$$\delta^{\rho_1} \to \delta^{\rho_2}, \quad \rho_1 \to \rho_2; \tag{5.4}$$

for all $x, y \in A$:

$$x \prec y \Rightarrow Tx \prec Ty, \tag{5.5}$$

$$(x \prec y,\ 1 - x_0 \leqslant \rho) \Rightarrow Tx \prec T_\rho y. \tag{5.6}$$

Remark 1. Using (5.2), we can easily show that the transformation T has precisely one fixed point on the set A^ν, $0 \leq \nu < \infty$, which is $\delta^{\rho(\nu)}$. Since A^ν is closed relative to T, we can naturally expect Proposition 3 to be true.

Proof of Proposition 4. By Proposition 1, $Q(A^\nu) = 1$. Moreover, A^ν is closed relative to T. We use the notation $x(n) \equiv (x_i(n))_{i\geq 0} \equiv T^n x$. From the explicit expressions (2.5) we can easily see that if $x \in A$, then

$$x_0(k) \geqslant \left(\sum_{i=0}^{k} x_i\right) e^{-k}, \quad k \in Z_+. \tag{5.7}$$

For any $x \in A^\nu$, we have by Chebyshev inequality

$$\sum_{i=0}^{k} x_i \geqslant 1 - \frac{\nu}{k+1}.$$

Fix $k \geq \nu$ and $\varepsilon > 0$ such that

$$\sum_{i=0}^{k} x_i > \varepsilon, \quad \forall x \in A^\nu.$$

From (5.7) we see that for any $x \in A^\nu$,

$$1 - x_0(n+k) < 1 - \varepsilon e^{-k} = \eta < 1, \quad n \geqslant 0.$$

Then from (5.6) it follows that

$$T^{n+k} x = T^n T^k x \prec T_\eta^{\,n} T^k x, \quad n \geqslant 0. \tag{5.8}$$

But by (5.1)

$$T_\eta^{\,n} T^k x \to \delta^\eta, \quad n \to \infty. \tag{5.9}$$

Recalling that T is a transformation preserving the measure Q, we obtain from (5.8) and (5.9), using the Poincaré recurrence theorem [5, p. 392],

$$Q(G^\eta) = 1, \quad G^\eta = \{x \in A^\nu \mid x \prec \delta^\eta\}.$$

Let $\rho_* = \min\{\eta \mid 0 \leq \eta < 1, Q(G^\eta) = 1\}$. The minimum is attained, which follows from (5.3) and (5.4), and therefore $Q(G^{\rho_*}) = 1$. If $\rho_* < \rho(\nu)$, then $G^{\rho_*} = \varnothing$, which is impossible. Therefore $\rho_* \geq \rho(\nu)$. We will show that $\rho_* = \rho(\nu)$.

Assume that this is not so: $\rho_* > \rho(\nu)$. From (5.5) and (5.2) it follows that G^{ρ_*} is closed relative to T. From the explicit expressions (2.5) we see that for any n, $x_0(n)$ depends (continuously) only on $x_0, \ldots, x_n$, and we can show that in the region

$$G_n^{\rho_*} = \left\{(x_i)_{i \in \overline{0,n}} \in [0,1]^{n+1} \,\middle|\, \sum_{j=0}^{i} x_j \geqslant \sum_{j=0}^{i} \delta_j^{\rho_*}, \quad 0 \leqslant i \leqslant n\right\}$$

$x_0(n)$ as a function of $(x_0, \ldots, x_n)$ attains a strict minimum $\delta_0^{\rho_*}$ for $x_i = \delta_i^{\rho_*}$, $i = \overline{0,n}$. This means that for any integer $l > 0$, $\alpha > 0$, there exists $\beta > 0$ such that if $x \in G^{\rho_*}$ and $y \in (T^{-1}x) \cap G^{\rho_*}$, then $|x_0 - \delta_0^{\rho_*}| \leqslant \beta$ implies $|y_i - \delta_i^{\rho_*}| \leqslant \alpha$, $i = \overline{0,l}$. Take l so that

$$\sum_{i=0}^{l} i\delta_i^{\rho_*} > \nu + \varepsilon_1, \quad \varepsilon_1 > 0.$$

(This can be done, because $\sum_{i=0}^{\infty} i\delta_i^{\rho_*} = \mu(\rho_*) > \mu(\rho(\nu)) = \nu$.) Then set $\alpha = \varepsilon_1/l^2$ and select an appropriate $\beta > 0$. Now, if $x \in G^{\rho_*}$ and $|x_0 - \delta_0^{\rho_*}| \leqslant \beta$, then for any $y \in (T^{-1}x) \cap G^{\rho_*}$,

$$\sum_{i=0}^{\infty} iy_i \geqslant \sum_{i=0}^{l} i(\delta_i^{\rho_*} - \alpha) > \nu,$$

i.e., $(T^{-1}x) \cap G^{\rho_*} = \varnothing$. Hence it follows that for any $x \in G^{\rho_*}$ we have $x_0(l) > (1 - \rho_*) + \beta$, and so

$$1 - x_0(l+n) < \rho_* - \beta, \quad n \geqslant 0.$$

Again applying relationships similar to (5.8), (5.9) and the Poincaré theorem, we obtain $Q(G^{\rho_* - \beta}) = 1$, which contradicts the definition of ρ_*. Thus, $\rho_* = \rho(\nu)$, and therefore $Q(G^{\rho(\nu)}) = 1$. But the set $G^{\rho(\nu)}$ consists precisely of one element $\delta^{\rho(\nu)}$. Therefore $Q = Q_{\rho(\nu)}$. Q.E.D.

Proof of Proposition 3. Fix an arbitrary element $x \in A^\nu$. From the proof of Proposition 4 it follows that

$$\overline{\lim_{n\to\infty}}\,(1-x_0(n))=\rho_*<1.$$

We will show that $\rho_* \le \rho(\nu)$. Assume that this is not so, $\rho_* > \rho(\nu)$. As in the proof of Proposition 4, we choose an integer l so that

$$\sum_{i=0}^{l} i\delta_i^{\rho_*}>\nu+\varepsilon_1,\quad \varepsilon_1>0,$$

and again for $\alpha = \varepsilon_1/l^2$ choose $\beta > 0$ so that

$$f^{-1}\{[0,1-\rho_*+\beta]\}\cap G_l^{\rho_*}\subset\{(y_i)_{i\in\overline{0,l}}\,|\,|y_i-\delta_i^{\rho_*}|<\alpha\}\equiv H_\alpha,$$

where the continuous function $f(y_0, ..., y_l)$ expresses the dependence of $y_0(l)$ on $(y_0, ..., y_l)$, $y = (y_i)_{i\ge0} \in A$.

For $n \to \infty$ we have $\overline{\lim}(1-x_0(n))=\rho_*$, and therefore for any $\gamma > 0$ there exists $n_* = n_*(\gamma) > 0$ such that for all $n \ge n_*$, $(x_0(n),\dots,x_l(n))\in G_l^{\rho_*+\gamma}$.

Take γ so that

$$f^{-1}\{[0,1-\rho_*+\beta]\}\cap G^{\rho_*+\gamma}\subset H_\alpha$$

(this can be done because H_α is open, G^{ρ_*}, $G^{\rho_*+\gamma}$ are compact sets, $G_l^{\rho_*+\gamma}\downarrow G^{\rho_*}$ for $\gamma\downarrow 0$), and choose $n_* = n_*(\gamma)$. Then for all $n \ge n_* + l$, from $1 - x_0(n) \ge \rho_* - \beta$ we have $(x_0(n-l), ..., x_l(n-l)) \in H_\alpha$, whence $\sum_{i=0}^{l} ix_i(n-l)>\nu$, which is impossible because $x(n-l) \in A^\nu$. Thus, for all $n \ge n_* + l$, $1 - x_0(n) < \rho_* - \beta$, which contradicts the definition of ρ_*.

Thus, $\overline{\lim}(1-x_0(n))\le\rho(\nu)$. Hence it follows that any subsequence of the sequence $(x(n))_{n\ge0}$ contains a limiting point $z \in A$, and $z<\delta^{\rho(\nu)}$.

We will show that necessarily $z = \delta^{\rho(\nu)}$. (This will indicate that $x(n) \to \delta^{\rho(\nu)}$.) It suffices to show that for any $\varepsilon > 0$

$$\limsup_{m\to\infty}\ \sup_{n\ge0}\sum_{i=m}^{\infty} ix_i(n)=0.$$

This in turn follows from a simple fact, which is stated without proof:

Let $0 \le \eta < 1$, $y \in A$, $\zeta(y) < \infty$, where $\zeta(y)=\sum_{i=0} iy_i$. Then for $n \to \infty$ we have $T_\eta^n y \to \delta^\eta$ as well as

$$\zeta(T_\eta^n y)\to\zeta(\delta^\eta).$$

This completes the proof of Proposition 3.

Remark 2. In our proof of Proposition 4, the support of the measure Q is restricted stepwise by the Poincaré lemma until it shrinks to a point. This proof of Proposition 4 is technically simpler than the approach that first proves Proposition 3 and then applies the Poincaré lemma. A similar situation is also observed for more complex systems (see [4]).

6. AUXILIARY LEMMAS

The following lemmas are of technical interest and are given without proof. (The proofs of Lemmas 1 and 2 are obvious. Lemmas 3-5 are proved in [4].) The notation used in Lemmas 1-3 is independent of the notation introduced in Secs. 1-5.

In what follows, distributions are elements of the set A (see Sec. 2), i.e., infinite sequences $x=(x_i)_{i\ge0}$, $x_i\ge0$, $i\ge0$, $\sum x_i=1$. By $\pi^\gamma = (\pi_i^\gamma)_{i\ge0}$ we denote the Poisson distribution with the parameter $\gamma \ge 0$:

$$\pi_i^\gamma=\gamma^i e^{-\gamma}/i!,\quad i\ge0.$$

and by $\theta^{m,n} = (\theta_i^{m,n})_{i\geq 0}$ the binomial distribution with the parameters 1/m, n (m, n $\in Z_+$):

$$\theta_i^{m,n} = \begin{cases} (n!/i!(n-i)!)(1/m)^i(1-1/m)^{n-i}, & 0\leq i\leq n, \\ 0, & i>n. \end{cases}$$

LEMMA 1. Let p and r be distributions such that

$$r_{i+1}/r_i \leq p_{i+1}/p_i \text{ for } i\geq 0,$$

where 0/0 = 0. Then $r\prec p$.

LEMMA 2. Let $0 < \alpha < \beta$ be fixed. Then there exists M such that for m > M and $n \leq \alpha m$,

$$\theta^{m,n}\prec\pi^{\beta}.$$

(The proof uses Lemma 1.)

LEMMA 3. Given are m servers, am of which constitute a distinguished set J; m, am $\in Z_+$. bm customers are allocated once with equal probabilities to all m servers, bm $\in Z_+$. Denote by R_i, $i \geq 0$, the random number of servers in the set J that are occupied by precisely i customers, and $r_i = R_i/(am)$. Then for $\forall\varepsilon, \xi > 0$ and $\forall l \in \{1,2,...\}$ uniformly in (a, b, i) $\in L = [\varepsilon, 1] \times [\varepsilon, \xi] \times \{0,1, ..., l\}$

$$\mathbf{M}(r_i)\to\pi_i^b,$$
$$\mathbf{D}(r_i)\to 0, \quad m\to\infty.$$

The following bound on the rate of convergence of the binomial distribution to Poisson [8] plays a central role in the proof of the lemma:

$$|C_n^i(\lambda/n)^i(1-\lambda/n)^{n-i}-\lambda^i/i!e^{-\lambda}|\leq\lambda^2/n, \lambda>0, \quad i=\overline{0, n}.$$

LEMMA 4 (notation from Secs. 2 and 3). The family of Markov processes $\{(X^N(n), D^N(n))_{n\geq 0}\}_{N\in Z}$ has the following property. There exists an integer k > 0 and N_0 such that for $N > N_0$, for any $n \geq 0$ and for any state of the process at the moment n,

$$U_{n+1}+\ldots+U_{n+k}\prec V,$$

where V is the random variable with the generating function

$$V(z)=\mathbf{M}z^V=[e^{-(k-1)\beta(1-z)}-e^{-(k-1)\beta}]e^{-\beta(1-z)}+e^{-(k-1)\beta}[(1-\eta)e^{-\beta(1-z)}+\eta e^{-\alpha(1-z)}].$$

where $0<\alpha<1<\beta$, $0<\eta\leq 1$, $\mathbf{M}V=k\beta+e^{-(k-1)\beta}[(1-\eta)\beta+\eta\alpha-1]<k$.

The proof uses Lemmas 2 and 3.

LEMMA 5 (notation of Sec. 2). For any ε, $\alpha > 0$, there is N_0 such that for $N > N_0$

$$\mathbf{P}(d(D^N(n+1), Tx)>\varepsilon|D^N(n)=x)<\alpha, \ n\geq 0,$$

uniformly in $x \in S^N$.

The proof relies on Lemma 3.

7. DISCUSSION

The method of analysis used in this paper treats the queueing process as a process of proportions, i.e., the state of the system as a whole at each time moment is defined by the proportions of the elements (out of the total number of system elements) that occupy various fixed states. (In our case, a system element is a server and the state of an element is the number of customers at the server.) The process of proportions, regardless of the total number of system elements (in our case M), takes values in the phase space (S^N), whose points are naturally interpreted as probability distributions on the state set (Z_+) of one element. Thus, the processes of proportions corresponding to systems with *different* number of elements may be treated as processes with values *in the same* phase space (A) — the space of *all* probability distributions on the state set of one element.

This make it possible to study the asymptotic behavior of the process of proportions when, as in our case, the system parameters, including the number of elements, increase without bound. Note that the process of proportions is usually meaningful only if it is Markov, which is so if the system being studied has a sufficiently symmetric structure.

In asymptotic analysis of the process of proportions, we can use the probabilistic interpretation of the points of the phase space. Thus, in our paper, in the proofs of Propositions 1 and 3, which constitute the main difficulty, we essentially use the stochastic order relation on the phase space and the associated monotonicity considerations.

The results of this paper are generalizable to a much more complex queueing system (see [4]). Other generalizations are also possible.

I an deeply grateful to Yu. M. Baryshnikov and L. B. Boguslavskii for numerous useful comments and to a referee for a number of useful suggestions.

LITERATURE CITED

1. M. A. Holliday and M. K. Vernon, "Exact performance estimates for multiprocessor memory and bus interference," IEEE Trans. Comput., **36**, No. 1, 76-85 (1987).
2. D. W. L. Jen, J. H. Patel, and E. S. Davidson, "Memory interference in synchronous multiprocessor systems," IEEE Trans. Comput., **31**, No. 11, 1116-1121 (1982).
3. F. Baskett and A. J. Smith, "Interference in multiprocessor computer systems with interleaved memory," Commun. ACM., **19**, No. 6, 327-334 (1976).
4. A. L. Stolyar, Asymptotic Behavior of the Stationary Distribution for One Class of Closed Queueing Networks [in Russian], Preprint, IPPI, Moscow (1988).
5. A. N. Shiryaev, Probability [in Russian], Nauka, Moscow (1980).
6. G. P. Klimov, Stochastic Queueing Systems [in Russian], Nauka, Moscow (1966).
7. V. F. Matveev and V. G. Ushakov, Queueing Systems [in Russian], Moscow State Univ. (1984).
8. A. A. Borovkov, Probability Theory [in Russian], Nauka, Moscow (1976).