Информационные процессы, Том 5, № 1, 2005, стр. 74–86

Динамическая маршрутизация в системе с отключающимися серверами[1]

Н.Д. Введенская*, **Ю.М. Сухов*,****

**Институт проблем передачи информации РАН, Москва, Россия*
***Стат. лаб., фак. чистой мат. и мат. стат., Кэмбриджский ун-т, Великобритания*
Поступило в редколлегию 17.02.2005

Аннотация—Рассматривается система обслуживания с N серверами, у каждого из которых имеется свой буфер; N велико ($N \to \infty$). На систему поступает пуассоновский поток заявок интенсивности $N\lambda$. Мы сравниваем две модели системы. В "линейной" модели заявка становится в очередь к случайно выбиранному серверу, а в "нелинейной" модели заявка случайно выборает два сервера и становится в наименьшую из очередей. Времена обслуживания заявок независимы и распределены экспоненциально со средним 1. При этом каждый из серверов может быть в рабочем состоянии ("on") или быть сломан (состояние "off") с марковскими переходами из состояния в состояние. Мы изучаем распределение длин очередей на серверах. Показывается, что если система не перегружена, то в линейной модели вероятности больших очередей убывают экспоненциально, а в нелинейной модели - сверхэкспоненциально.

Dynamic Routing Queueing Systems with Vacations[1]

N. D. Vvedenskaya*, **Yu. M. Suhov*,****

**Institute for Information transmission Problems RAS, Moscow, Russia*
***Statistical Lab., DPMMS, Univ. of Cambridge, Great Britain*

Abstract Consider a system with N single servers (each with its own queueing buffer) and a common Poisson input flow of tasks of intensity $N\lambda$ where N is large ($N \to \infty$). We compare two service models. In a 'linear' model, a task selects a server at random and joins the corresponding queue. In the 'nonlinear' model, a task selects two servers at random and is despatched to the one with the shorter queue. The task service time is distributed exponentially with mean 1, independently of other tasks. In addition, every server may be in an active stage (*'on'*) or broken down (stage *'off'*), the transition from one stage to the other being Markovian. We investigate the distribution of the queue lengths at the servers. It is shown that when the system is not overloaded the probability of long queues in linear model decreases exponentially while in nonlinear model superexponentially.

1. INTRODUCTION

Consider a system with N servers and with a common Poisson arrival flow of tasks of intensity $N\lambda$, $\lambda > 0$. We are interested in two models of selection of a server:

A *linear model* : upon arrival each task selects a server at random and is instantly dispatched to this server where it joins the corresponding queue.

[1] Работа НДВ выполнена при поддержке гранта РФФИ No 02.01.00068
[1] The work of NDV is partly supported by RFFI grant No 02.01.00068

A *nonlinear model* : upon arrival each task selects two servers at random and is instantly dispatched to the one where the queue is shorter. Possible ties are broken at random.

The service time of a task is distributed exponentially, the service times of different tasks being independent. Without loss of generality we take that the mean is equal to one. Servers have infinite buffers and use a conservative service discipline.

A server can be active (stage *'on'*) or broken (stage *'off'*). In the active stage the server processes the tasks one by one at rate one. If a server is broken, the service is interrupted, and resumed in when it returns to the *on*-stage. The duration of stage *'on'* is distributed exponentially with mean $1/\tau$, and of stage *'off'* exponentially with mean $1/\sigma$, where $\tau, \sigma > 0$. The lengths of different stages are independent.

The above description suggests that both models lead to homogeneous Markov processes where a state is represented by a collection of occupation numbers indicating how many servers are at a given stage and have a given queue length. It is not hard to show that for

$$\lambda < \frac{\sigma}{\sigma + \tau} \tag{1}$$

both Markov processes are positive recurrent. In addition in the linear model the queue lengths at different servers are independent

Let $L_k(t)$ be the number of servers in state *'on'* for which at time t the queue length is at least k (the served task is also taken into account) and $M_k(t)$ be the number of servers in state *'off'* for which at time t the queue length is at least k, $k = 0, 1, \ldots$. Let

$$U_{k,N}(t) = L_k(t)/N, \;\; V_{k,N}(t) = M_k(t)/N,$$

be the fractions of these servers. Obviously $\forall t \geq 0$, $k = 0, 1, \ldots$ and $U_{k,N}(t)$, $V_{k,N}(t)$ take values in $\{0, 1/N, 2/N, \ldots, 1\}$. (We will systematically omit the argument t and index N.) Furthermore, $\forall t$

$$U_{0,N} \geq U_{1,N} \geq \ldots \geq 0, \;\; V_{0,N} \geq V_{1,N} \geq \ldots \geq 0, \;\; \text{and} \;\; U_{0,N} + V_{0,N} = 1.$$

Consider the mean values

$$u_k(t) = \mathbb{E}U_k(t), \;\; v_k(t) = \mathbb{E}V_k(t)$$

with

$$u_0 \geq u_1 \geq \ldots \geq 0, \;\; v_0 \geq v_1 \geq \ldots \geq 0, \;\; \text{and} \;\; u_0 + v_0 = 1. \tag{2}$$

In the linear model they satisfy the infinite system of linear ordinary differential equations

$$\dot{u}_k = \lambda(u_{k-1} - u_k) + u_{k+1} - u_k - \tau u_k + \sigma v_k, \;\; k \geq 1, \tag{3.1}$$

$$\dot{v}_k = \lambda(v_{k-1} - v_k) + \tau u_k - \sigma v_k, \;\; k \geq 1, \tag{3.2}$$

$$\dot{u}_0 = -\tau u_0 + \sigma v_0, \;\; \dot{v}_0 = +\tau u_0 - \sigma v_0, \tag{3.3}$$

in the class of sequences of functions described by (2), and with Cauchy data

$$u_k|_{t=0} = g_k, \;\; v_k|_{t=0} = h_k, \;\; k \geq 0 \tag{3.4}$$

generated by the initial distribution of the corresponding Markov process.

Equations (3) are Kolmogorov's backward equations for the so-called coordinate functions in the Markov process of the linear model. Observe that they do not depend on N (and coincide with Kolmogorov's forward equations for the transitional probabilities in the single-server M/M/1 queue with server's vacations). These equations have a global solution $(\mathbf{u}(t), \mathbf{v}(t))$ which is unique in a class described by (2). Here $\mathbf{u}(t) = \mathbf{u}(t, \mathbf{g}, \mathbf{h}) = \{u_k(t, \mathbf{g}, \mathbf{h})\}$ and $\mathbf{v}(t) = \mathbf{v}(t, \mathbf{g}, \mathbf{h}) = \{v_k(t, \mathbf{g}, \mathbf{h})\}$ are infinite-dimensional vector-functions, and $\mathbf{g} = \{g_k\}, \mathbf{h} = \{h_k\}$, $k = 0, 1, \ldots$, are sequences of initial data.

By $(\mathbf{a}, \mathbf{b})$, with $\mathbf{a} = \{a_k\}$, $\mathbf{b} = \{b_k\}$, $k = 0, 1, ...$, we denote the stationary solution (a fixed point) to (3.1)–(3.4), its entries satisfy

$$\lambda(a_{k-1} - a_k) + a_{k+1} - a_k - \tau a_k + \sigma b_k = 0, \ \ k > 0, \tag{4.1}$$

$$\lambda(b_{k-1} - b_k) + \tau a_k - \sigma b_k = 0, \ \ k > 0, \tag{4.2}$$

$$-\tau a_0 + \sigma b_0 = 0 \tag{4.3}$$

with the boundary values

$$a_0 + b_0 = 1. \tag{4.4}$$

Under condition (1) such a solution exists and is unique in the class described by (2).

In the nonlinear model the queue lengths at different servers are dependent. Here, as $N \to \infty$, the mean values $u_{k,N}(t)$, $v_{k,N}(t)$ tend to the solution to an infinite system of nonlinear ODEs:

$$\dot{u}_k = \lambda \left(u_{k-1}^2 - u_k^2\right) + \lambda \left(u_{k-1} - u_k\right)\left(v_{k-1} + v_k\right) + u_{k+1} - u_k - \tau u_k + \sigma v_k, \ \ k > 0, \tag{5.1}$$

$$\dot{v}_k = \lambda \left(v_{k-1}^2 - v_k^2\right) + \lambda \left(v_{k-1} - v_k\right)\left(u_{k-1} + u_k\right) + \tau u_k - \sigma v_k, \ \ k > 0, \tag{5.2}$$

$$\dot{u}_0 = -\tau u_0(t) + \sigma v_0(t), \quad \dot{v}_0 = +\tau u_0 - \sigma v_0 \tag{5.3}$$

As before these equations are considered in class defined by (2) and with initial values (3.4). The global solution to (5.1)–(5.3) again exists and is unique in this class, as before we denoted it $(\mathbf{u}, \mathbf{v})$.

Equations (5.1)–(5.3) give rise to there own stationary solution. Here the equations are

$$\lambda \left(a_{k-1}^2 - a_k^2\right) + \lambda \left(a_{k-1} - a_k\right)\left(b_{k-1} + b_k\right) + a_{k+1} - a_k - \tau a_k + \sigma b_k = 0, \ \ k > 0, \tag{6.1}$$

$$\lambda \left(b_{k-1}^2 - b_k^2\right) + \lambda \left(b_{k-1} - b_k\right)\left(a_{k-1} + a_k\right) + \tau a_k - \sigma b_k = 0, \ \ k > 0, \tag{6.2}$$

$$\tau a_0(t) - \sigma b_0(t) = 0, \tag{6.3}$$

with boundary values $a_0 + b_0 = 1$. As in the linear model under condition (1) a solution to (6.1)–(6.3) exists and is unique in the same class.

To stress particular features of the two models, we will write $\mathbf{u}^L(t)$, $\mathbf{v}^L(t)$ for solutions to Cauchy problem (3) and $\mathbf{a}^L$, $\mathbf{b}^L$ for solutions to (4). Similarly, $\mathbf{u}^{NL}(t)$, $\mathbf{v}^{NL}(t)$ will stand for solutions to (5) and $\mathbf{a}^{NL}$, $\mathbf{b}^{NL}$ for solutions to (6).

Equation (3) and (5) are the main tool of studying both the linear and nonlinear model with large number of servers N. They capture some interesting aspects of behavior. For example if $\tau = \sigma = 0$ then in the linear model the stationary solution has $a_k = \lambda^k, b_k = 0$. The nonlinear model with $\tau = \sigma = 0$ was studied in [1]; it was shown that here $a_k = \lambda^{2^k - 1}, b_k = 0, \ \ k \geq 0$.

A somewhat different model with *'on/off'* stages of service process was investigated in [2].

A similar theory can be proposed for the case of Jackson-type queueing networks.

The sections that follow elaborate on the outlined approach. In section 2 we discuss mathematical facts about Markov processes and equations (3) and (5). In Section 3 we give the results of numerical simulations for nonlinear model with $N = 5, \ 10, \ 20$ and compare linear and nonlinear models with each other and with the limiting model as $N \to \infty$. In Section 4 we present a sketch of some proofs. As the proofs follow that of [1] and [3] many technical constructions are only briefly quoted.

2. MAIN DEFINITIONS AND THEOREMS

Let $\bar{\mathcal{U}}$ denote the set of sequences $(\mathbf{g},\mathbf{h}) = \{g_k, h_k\}_{i=0}^{\infty}$

$$g_0 \geq g_1 \geq \ldots \geq 0, \quad h_0 \geq h_1 \geq \ldots \geq 0, \quad g_0 + h_0 = 1,$$

with the metric

$$\rho(\mathbf{g},\mathbf{h}),(\mathbf{g}',\mathbf{h}')) = \sup_{k>0} \max\Big(\frac{|g_k - g'_k|}{k}, \frac{|h_k - h'_k|}{k}\Big), \ \ (\mathbf{g},\mathbf{h}), \ \ (\mathbf{g}',\mathbf{vh}') \in \bar{\mathcal{U}}.$$

Under this metric $\bar{\mathcal{U}}$ is compact, and the convergence in $\bar{\mathcal{U}}$ is equivalent to the coordinate–wise convergence.

Denote by $\mathcal{U}$ the subset of $\bar{\mathcal{U}}$, formed by pairs $(\mathbf{g},\mathbf{h})$ with $\sum_k (g_k + h_k) < \infty$, and by $\mathcal{U}_N$ the subset of $\mathcal{U}$, with $g_k = l_k/N, \ h_k = m_k/N$ where l_k, m_k are nonnegative integers.

The evolution of our system which was informally described in the previous section (in both linear and nonlinear cases) is governed by homogeneous Markov processes $\Omega_N(t) = (\mathbf{U_N}(t), \mathbf{V_N}(t))$ with state space $\mathcal{U}_N$.

Now, let $(\mathbf{g},\mathbf{h}) \in \mathcal{U}_N$. In the linear case the Markov process is determined by the generating operator $\mathbf{A}_N$ acting on functions $f : \mathcal{U} \to \mathbb{R}$

$$\mathbf{A}_N = \mathbf{A}_N^{In} + \mathbf{A}_N^{Out},$$

$$\mathbf{A}_N^{In} f(\mathbf{g},\mathbf{h}) = \lambda N \sum_{k=1}^{\infty} \Big[(g_{k-1} - g_k)\Big(f(\mathbf{g} + \frac{\mathbf{e}_k}{N}, \mathbf{h}) - f(\mathbf{g},\mathbf{h})\Big)$$

$$+(h_{k-1} - h_k)\Big(f(\mathbf{g}, \mathbf{h} + \frac{\mathbf{e}_k}{N}) - f(\mathbf{g},\mathbf{h})\Big)\Big], \tag{7.1}$$

$$\mathbf{A}_N^{Out} f(\mathbf{g},\mathbf{h}) = N \sum_{k=1}^{\infty} \Big[(g_k - g_{k+1})\Big(f(\mathbf{g} - \frac{\mathbf{e}_k}{N}, \mathbf{h}) - f(\mathbf{g},\mathbf{h})\Big)\Big]$$

$$+\tau N \sum_{k=0}^{\infty} (g_{k-1} - g_k)\Big(f(\mathbf{g} - \sum_{j=0}^{k} \frac{\mathbf{e}_j}{N}, \mathbf{h} + \sum_{j=0}^{k} \frac{\mathbf{e}_j}{N}) - f(\mathbf{g},\mathbf{h})\Big)$$

$$+\sigma N \sum_{k=0}^{\infty} (h_{k-1} - h_k)\Big(f(\mathbf{g} + \sum_{j=0}^{k} \frac{\mathbf{e}_j}{N}, \mathbf{h} - \sum_{j=0}^{k} \frac{\mathbf{e}_j}{N}) - f(\mathbf{g},\mathbf{h})\Big). \tag{7.2}$$

Here $\mathbf{e}_k$ is a sequence with all entries but the kth one equal to 0 and the kth entry 1.

In the nonlinear model the Markov process is determined by the generating operator

$$\mathbf{B}_N = \mathbf{B}_N^{In} + \mathbf{B}_N^{Out} \tag{8}$$

of the form similar to (7.1)–(7.2). Furthermore, $\mathbf{A}_N^{Out} = \mathbf{B}_N^{Out}$, and the only difference in $\mathbf{A}_N$ and $\mathbf{B}_N$ is between $\mathbf{A}_N^{In}$ and $\mathbf{B}_N^{In}$

$$\mathbf{B}_N^{In} = \lambda N \sum_{k=1}^{\infty} \Big[\Big((g_{k-1})^2 - (g_k)^2 + (g_{k-1} - g_k)(h_{k-1} + h_k) \Big)$$

$$\times \Big(f(\mathbf{g} + \frac{\mathbf{e}_k}{N}, \mathbf{h}) - f(\mathbf{g},\mathbf{h}) \Big)$$

$$+\Big((h_{k-1})^2 - (h_k)^2 + (h_{k-1} - h_k)(g_{k-1} + g_k) \Big)$$

$$\times\Big(f(\mathbf{g},\mathbf{h}+\frac{\mathbf{e}_k}{N})-f(\mathbf{g},\mathbf{h})\Big)\Big].$$

The sums with factor λN corresponds to the increase of the queue length caused by the exogenous arrival. More precisely, the probability of a task to select a server in stage $'on'$ with a queue length $k-1$ is the sum of 1) the probability that both selected server are in stage $'on'$ and both queues are not shorter then $k-1$ (this probability is equal to $(g_{k-1})^2-(g_k)^2$), and 2) the probability that a task is directed to a server in stage $'on'$ with queue length $k-1$ while the other selected server is in stage $'off'$ and has the queue length not shorter then $k-1$ (this probability is $(g_{k-1}-g_k)(h_{k-1}+h_k)$). This explains the term with the factor $\Big((g_{k-1})^2-(g_k)^2+(g_{k-1}-g_k)(h_{k-1}+h_k)\Big)$. The second term with the factor $\Big((h_{k-1})^2-(h_k)^2+(h_{k-1}-h_k)(g_{k-1}+g_k)\Big)$ is explained by a similar argument.

Rewrite the last terms in (7.2):

$$f\Big(\mathbf{g}\mp\sum_{j=0}^{k}\frac{\mathbf{e}_j}{N},\mathbf{h}\pm\sum_{j=0}^{k}\frac{\mathbf{e}_j}{N}\Big)-f\Big(\mathbf{g},\mathbf{h}\Big)$$

$$=\sum_{i=j}^{k}\Big[f\Big(\mathbf{g}\mp\sum_{l=i}^{k}\frac{\mathbf{e}_l}{N},\mathbf{h}\pm\sum_{l=i}^{k}\frac{\mathbf{e}_l}{N}\Big)-f\Big(\mathbf{g}\mp\sum_{l=i+1}^{k}\frac{\mathbf{e}_l}{N}\mathbf{h}\pm\sum_{l=i+1}^{k}\frac{\mathbf{e}_l}{N}\Big)\Big].$$

Now we can present $\mathbf{A}_N f(\mathbf{g},\mathbf{h})$ in the form

$$\mathbf{A}_N f(\mathbf{g},\mathbf{h})=\lambda\sum_{k=1}^{\infty}\Big[(g_{k-1}-g_k)\frac{\partial f(\mathbf{g},\mathbf{h})}{\partial g_k}+(h_{k-1}-h_k)\frac{\partial f(\mathbf{g},\mathbf{h})}{\partial h_k}\Big]$$

$$-\sum_{k=1}^{\infty}(g_k-g_{k+1})\frac{\partial f(\mathbf{g},\mathbf{h})}{\partial g_k}+\sum_{k=0}^{\infty}\Big[(\sigma h_k-\tau g_k)\frac{\partial f(\mathbf{g},\mathbf{h})}{\partial g_k}+(\tau g_k-\sigma h_k)\frac{\partial f(\mathbf{g},\mathbf{h})}{\partial h_k}\Big]$$

$$+\frac{1}{N}\mathrm{O}\Big(\frac{\partial^2 f(\tilde{\mathbf{g}},\tilde{\mathbf{h}})}{\partial g_i\partial h_j}\Big),\tag{9}$$

where $(\tilde{\mathbf{g}},\tilde{\mathbf{h}})=\Big(\mathbf{g}+\mathrm{O}\Big(\frac{\mathbf{e}_i}{N}\Big),\mathbf{h}+\mathrm{O}\Big(\frac{\mathbf{e}_j}{N}\Big)\Big)$.

The first mathematical result in this paper is that as $N\to\infty$ the transition probabilities of both Markov processes (with the generating operators $\mathbf{A}_N$ and $\mathbf{B}_N$) converge uniformly on any finite time interval to the transition probabilities of the limiting processes (see Theorem 1 and 2). The limiting processes are deterministic dynamical systems acting in the infinite-dimensional state space. This result is derived from the existing theorems that the convergence of generating operators implies the convergence of the transition functions of the corresponding Markov processes (see, e.g., [4], ch. 1).

As was mentioned in Introduction, the condition of ergodicity for these Markov processes is given in (1). We prove that under this condition their equilibrium distributions converge to the delta-measures concentrated at fixed points of the corresponding limiting dynamical systems (see Theorem 3). Here we need to verify a uniform compactness with respect to N of the family of stationary distributions of the processes under consideration. The proofs use properties of solutions to boundary-value problems (3) and (5).

Associated with process $(\mathbf{U}_N(t),\mathbf{V}_N(t))$ is the operator semigroup $\mathbf{T}_N=\mathbf{T}_N(t)$ acting on functions defined on $\mathcal{U}_N$. Namely, if $f:\mathcal{U}_N\to\mathbb{R}$ then

$$\Big(\mathbf{T}_N(t)f\Big)(\mathbf{g},\mathbf{h})=\mathbb{E}\left(f(\mathbf{U}_N(t),\mathbf{V}_N(t))\mid\mathbf{U}_N(0)=\mathbf{g},\mathbf{V}_N(0)=\mathbf{h}\right),\quad \mathbf{g},\mathbf{h}\in\mathcal{U}_N.\tag{10}$$

Here and below $\mathbb{E}$ stands for the expectation relative to the distribution of the respective Markov process.

Theorem 1. *Let f be continuous function $\bar{\mathcal{U}} \to \mathbb{R}$. Then in both linear and nonlinear models for any $t > 0$*

$$\lim_{N\to\infty} \sup_{\mathbf{g},\mathbf{h}\in\mathcal{U}_N} \left| \Big(\mathbf{T}_N(t)f\Big)(\mathbf{g},\mathbf{h}) - f(\mathbf{u}(t,\mathbf{g},\mathbf{h}),\mathbf{v}(t,\mathbf{g},\mathbf{h}))\right| = 0. \tag{11}$$

The convergence is uniform in $t \in [0,T]$ for any $T > 0$.

Let $G^{(k)}$ and $H^{(k)}$, $k = 0,1,...$ stand for coordinate functions on $\bar{\mathcal{U}}$:

$$G^{(k)} : (\mathbf{g},\mathbf{h}) \to g_k, \quad H^{(k)} : (\mathbf{g},\mathbf{h}) \to h_k. \tag{12}$$

Theorem 2. *Given $\mathbf{g},\mathbf{h} \in \mathcal{U}$, let sequences of points $(\mathbf{g}_N,\mathbf{h}_N) = \{(g_N)_j,(h_N)_j\}_{j=0}^{\infty} \in \mathcal{U}_N$, $N = 1,2,...$, be such that $\mathbf{g}_N \to \mathbf{g}$, $\mathbf{h}_N \to \mathbf{h}$ in $\bar{\mathcal{U}}$ as $N \to \infty$, and the series $\sum_{k=1}^{\infty}(g_N)_k$, $\sum_{k=1}^{\infty}(h_N)_k$ converge uniformly in N. Then in both the linear and nonlinear models, for all $t \geq 0$*

$$\lim_{N\to\infty} \Big(\mathbf{T}_N(t)\sum_{k=1}^{\infty} G^{(k)}\Big)(\mathbf{g}_N,\mathbf{h}_N) = \sum_{k=1}^{\infty} u_k(t,\mathbf{g},\mathbf{h}),$$

$$\lim_{N\to\infty} \Big(\mathbf{T}_N(t)\sum_{k=1}^{\infty} H^{(k)}\Big)(\mathbf{g}_N,\mathbf{h}_N) = \sum_{k=1}^{\infty} v_k(t,\mathbf{g},\mathbf{h}).$$

The convergence is uniform in $t \in [0,T]$ for any $T > 0$.

Theorem 2 does not follow from Theorem 1 because the functions $(\mathbf{g},\mathbf{h}) \to \sum_{k=1}^{\infty} u_k$, $(\mathbf{g},\mathbf{h}) \to \sum_{k=1}^{\infty} v_k$ are not continuous in $\bar{\mathcal{U}}$.

Observe that $N\sum_{k=1}^{\infty}(U_k(t) + V_k(t))$ is equal to the number of tasks in the queues of a system.

Theorem 3. *Let condition* (1) *hold. Then*

(a) there exists the unique equilibrium distribution π_N of the Markov process $\Big(\mathbf{U}_N(t),\mathbf{V}_N(t)\Big)$, π_N obey

$$\sup_N \int_{\mathcal{U}_N} \sum_{k=0}^{\infty} \Big(G^{(k)}(\mathbf{g},\mathbf{h}) + H^{(k)}(\mathbf{g},\mathbf{h})\Big) d\pi_N(\mathbf{g},\mathbf{h}) < \infty. \tag{13}$$

(b) As $N \to \infty$ π_N weakly converges to the delta-measure supported by a fixed point $(\mathbf{a},\mathbf{b})$ satisfying equations (4) *in the linear and* (6) *in nonlinear case.*

Theorem 4.

(a) Let $\mathbf{g},\mathbf{h} \in \bar{\mathcal{U}}$. Then there exists in $\bar{\mathcal{U}}$ a unique solution $(\mathbf{u}(t),\mathbf{v}(t))= (\mathbf{u}(t;\mathbf{g},\mathbf{h})\mathbf{v}(t;\mathbf{g},\mathbf{h}))$ of problem (3). *If $\mathbf{g},\mathbf{h} \in \mathcal{U}$, then $\mathbf{u}(t),\mathbf{v}(t) \in \mathcal{U}$ for all $t > 0$.*

The same is true for problem (5).

(b)*Assume that condition* (1) *holds. Then there exists in $\mathcal{U}$ a unique solution $(\mathbf{a}^L,\mathbf{b}^L)$ to problem* (4) *and a unique solution $(\mathbf{a}^{NL},\mathbf{b}^{NL})$ to problem* (6).

If $(\mathbf{g},\mathbf{h}) \in \mathcal{U}$, then $\lim_{t\to\infty} u_k(t) = a_k$, $\lim_{t\to\infty} v(t)_k = b_k$.

The entries a_k^L, b_k^L of $(\mathbf{a}^L,\mathbf{b}^L)$ show exponential decay in k. More precisely

$$a_k^L = \alpha_1\Lambda_1^k + \alpha_2\Lambda_2^k, \quad b_k^L = \beta_1\Lambda_1^k + \beta_2\Lambda_2^k, \quad k = 0,1,....$$

Here Λ_1,Λ_2 are the eigenvalues of the recursion matrix in (3)*:*

$$\det\begin{pmatrix} \lambda - \Lambda_i & \lambda \\ \frac{\lambda\tau}{\lambda+\sigma} & \frac{\lambda(1+\tau)}{\lambda+\sigma} - \Lambda_i \end{pmatrix} = 0,$$

$$\Lambda_i = \frac{1}{2}(\lambda + \frac{\lambda(1+\tau)}{\lambda+\sigma}) \pm \sqrt{\frac{1}{4}(\lambda + \frac{\lambda(1+\tau)}{\lambda+\sigma})^2 - \frac{\lambda^2\tau}{\lambda+\sigma}} \in (0,1)$$

and α_i, β_i *are constants.*

The entries a_k^{NL}, b_k^{NL} *of* $(\mathbf{a}^{NL}, \mathbf{b}^{NL})$ *decay superexponentially:* $\forall\theta, 2 > \theta > 0$,

$$\frac{a_k^{NL}}{e^{\theta k}}, \ \frac{b_k^{NL}}{e^{\theta k}}, \quad \to 0 \ \text{ as } \ k \to \infty.$$

(c) *Assume that condition* (1) *holds. Than* $\forall(\mathbf{g}, \mathbf{h}) \in \mathcal{U}$, *as* $t \to \infty$, *the solution* $(\mathbf{u}^L(t), \mathbf{v}^L(t))$ *of Cauchy problem* (3) *converges to* $(\mathbf{a}^L, \mathbf{b}^L)$ *and the solution* $(\mathbf{u}^{NL}(t), \mathbf{v}^{NL}(t))$ *of Cauchy problem* (5) *converges to* $(\mathbf{a}^{NL}, \mathbf{b}^{NL})$, *in the topology of space* $\bar{\mathcal{U}}$.

A similar theorem can be developed for models of open Jackson-type systems which generalize the system considered in [3].

Namely, consider a network with J stations (nodes) $S^{(j)}$, $j = 1, ..., J$, that is fed by a family of independent Poisson flows of intensity $(N\lambda^{(1)}, ..., N\lambda^{(J)})$. Each station is a system of the kind described above: it contains N servers switching between '*on*' and '*off*' states and is characterized by parameters τ^i, σ^i, $i = 1, ..., J$. As before, we can consider the linear or the nonlinear model. In the former, a task arriving in station $S^{(i)}$ chooses a server at random, with probability $\frac{1}{N}$, whereas in the later it selects first two servers at random and then chooses the one with the shorter queue. The Jackson rule is that after completing service in station $S^{(i)}$ the task is dispatched to station $S^{(j)}$ with probability $p_i^{(0)} p_{ij}$ (where it joins one of the queues following the rules of the model) and lives the network with probability $1 - p_j^{(0)}$, (here $0 \le p_i^{(0)} \le 1$, $0 \le p_{ij} \le 1$, and $\sum_j p_{ij} = 1, \ \forall \ i, j = 1, ..., J$.)

Like before, let $U_{k,N}^{(j)}(t), V_{k,N}^{(j)}(t)$ denote the fractions of servers at time t in nod $S^{(j)}$ at stage *'on'* and *'off'* correspondingly, which have queue lengths at least k, and $u_k^j(t) = \mathbb{E}U_{k,N}^{(j)}(t)$, $v_k^j(t) = \mathbb{E}V_{k,N}^{(j)}(t)$, $t > 0, \ \ k = 0, 1, ...$ Then in the linear model

$$\dot{u}_k^j = \bar{\lambda}^j(u_{k-1}^j - u_k^j) + u_{k+1}^j - u_k^j - \tau^j u_k^j + \sigma^j v_k^j, \quad k \ge 1,$$

$$\dot{v}_k^j = \bar{\lambda}^j(v_{k-1}^j - v_k^j) + \tau^j u_k^j - \sigma^j v_k^j, \quad k \ge 1,$$

$$\dot{u}_0^j = -\tau^j u_0^j + \sigma^j v_0^j, \quad \dot{v}_0^j = +\tau^j u_0^j - \sigma^j v_0^j,$$

$$v_0^j + u_0^j = 1,$$

$$u_k^j|_{t=0} = g_k^j, \quad v_k^j|_{t=0} = h_k^j, \quad k \ge 0,$$

$$\bar{\lambda}^j = \bar{\lambda}^j(t) = \lambda^{(j)} + \sum_{i=1}^J p_i^{(0)} p_{ij} u_1^i.$$

Following (5) it is easy to present similar equations for the nonlinear limiting model as $N \to \infty$.

It is natural to presume that in an open system $u_k \to 0, \ v_k \to 0, k > 0$, if all $\lambda^{(j)} = 0$. For the algebraic system corresponding to the network

$$\rho^j = \lambda^{(j)} + \sum_{i=1}^J p_i^{(0)} p_{ij} \rho^i$$

that presumption indicates that as $n \to \infty$ the iterations

$$\rho_n^j = \lambda^{(j)} + \sum_{i=1}^J p_i^{(0)} p_{ij} \rho_{n-1}^i$$

converge to the solution for any $\rho_0^j \geq 0$.

Under this condition the next Theorem takes place:

Theorem 5 *The Markov process that guides the system (with linear or/and nonlinear subsystems S^j) is ergodic if the algebraic system corresponding to the network has a solution* $(\rho^1, ..., \rho^J), 0 < \rho^j < \frac{\sigma^j}{\sigma^j + \tau^j}$, $j = 1, ..., J$ (compare with [3]).

3. SIMULATION RESULTS

In this section, we report results of numerical simulations.

Simulation have been done for a Markov chain that approximates the performance of the system. Namely, we consider a system with N servers operating in a discrete time. Each server may be in state 'on' or '*off*'. At times $K\Delta,\ K = 0, 1, ...,$ where $0 < \Delta \ll 1$ is a fixed constant, the following events may happen independently :

- with probability $N\lambda\Delta$ a new task arrives upon the system and is directed to one of the servers according to the rules of nonlinear system,
- with probability Δ the nonzero queue length of each server that is in state 'on' decreases by 1.
- with probability $\tau\Delta$ each server that is in state 'on' changes its state to '*off*'
- with probability $\sigma\Delta$ each server that is in state '*off*' changes its state to 'on'.

At time $t = 0$ all servers are in state 'on' and all queues are empty. In the experiments, $N\Delta \cong 0.05$, and the number M of simulation steps is so that $M\Delta > 1000$. This practically allowed us to reach the stationary regime.

In the following tables the lines marked by $N = 1$ give the stationary data for the linear case. They are calculated analytically. The lines marked by $N = 5, 10, 20$ represent the results of simulations in the nonlinear case for $N = 5, 10, 20$ correspondently. The line marked by $N = \infty$ gives the stationary data for the nonlinear limiting case, they are calculated analytically.

For several values of τ, σ we produce data for $\lambda = 0.4$ and for $\lambda = 0.9\sigma/(\tau+\sigma)$ (compare with condition (1)). All values are given with accuracy 0.01 (thus we write 0.00 for the values that are less than 0.005). We also give the value of x - the largest eigenvalues of the recursion matrix B where

$$\begin{pmatrix} a_k^L \\ b_k^L \end{pmatrix} = B \begin{pmatrix} a_{k-1}^L \\ b_{k-1}^L \end{pmatrix},$$

x guides the decrease of a_k^L, b_k^L as $k \to \infty$.

The values $a_0 = \frac{\sigma}{\sigma + \tau}$, $b_0 = \frac{\tau}{\sigma + \tau}$ and $a_1 = \lambda$ are the same for all lines in a table; the values of b_1^L and b_1^{NL} are also presented,

$$b_1^L = \frac{\lambda b_0 + \tau\lambda}{\sigma + \lambda},$$

$$b_1^{NL} = -\frac{a_0 + \lambda + \sigma/\lambda}{2} + \sqrt{\left(\frac{a_0 + \lambda + \sigma/\lambda}{2}\right)^2 + b_0^2 + b_0(a_0 + \lambda) + \tau}.$$

Observe that $b_1^{NL} > b_1^L$, indicating that in the nonlinear stationary model there are more occupied servers in stage *'off'* then in the liner one. The inequality $b_1^{NL} > b_1^L$ was unexpected, we hoped that in nonlinear model less servers are occupied (for $\lambda = 0.9\sigma/(\sigma + \tau)$ the values $b_1^L,\ b_1^{NL}$ are presented with the accuracy 0.001).

Knowing the rate of fast decrease of a_k^L, b_k^L and a_k^{NL}, b_k^{NL} as $k \to \infty$ we were not interested in there exact values, therefore we do not give any graphs and do not present a_k, b_k for large k. Our tables are to demonstrate how close are the values $a_k,\ b_k$ in the nonlinear models with 5 - 10 stations to the limit values a_k^{NL}, b_k^{NL} as k=2 - 6.

It is shown that for $\lambda = 0.4$ the values for $a_k,\ b_k$ in the models with finite number of servers are close to limit $a_k^{NL},\ b_k^{NL}$ even for $N = 5$. As $\lambda = 0.9\sigma/(\tau + \sigma)$ the values for $a_k, b_k,\ \ k \le 6$, decrease rather slow, but a_6, b_6 in the nonlinear models with $N = 10 - 20$ (and limit a_6^{NL}, b_6^{NL}) are already considerably less than a_6^L, b_6^L.

Both calculations and simulations show that for τ/σ =const the values of a_k, b_k decrease faster, when the lengths of *'on'/'off'* periods are shorter (thus the model performs 'better' when the *'on'* and *'off'* stages are short). And, of cause, the larger $\dfrac{\sigma}{\tau} = \dfrac{\text{mean length of 'on' stage}}{\text{mean length of 'off' stage}}$ the better.

We start with 5 values for a_k for the simplest case where $\tau = \sigma = 0,\ \ a_0 = 1$ and $b_k \equiv 0$. Here $a_k^L = \lambda^k$ and $a_k^{NL} = \lambda^{2^k - 1}$ (see [1]).

$\lambda = 0.4,\ \tau = 0.,\ \ x = 0.4$

$N \setminus k$	2	3	4	5
1	0.16	0.64	0.26	0.01
5	0.09	0.01	0.00	0.00
10	0.08	0.01	0.00	0.00
20	0.07	0.00	0.00	0.00
∞	0.06	0.00	0.00	0.00

$\lambda = 0.9,\ \tau = 0.,\ \ x = 0.9$

2	3	4	5
0.81	0.73	0.66	0.59
0.74	0.57	0.39	0.25
0.74	0.53	0.34	0.18
0.73	0.49	0.25	0.08
0.73	0.48	0.21	0.04

Next some values are given in a 'good' case where the mean length of an $'on'$ period is 10 times larger then the mean length of an $'off'$ period. Here $a_0 = 0.91,\ b_0 = 0.09$

$\lambda = 0.4, \tau = 0.025, \sigma = 0.25,\ x = 0.66,\ b_1^L = 0.071,\ b_1^{NL} = 0.076$

	a_k				b_k			
$N \setminus k$	2	3	4	5	2	3	4	5
1	0.19	0.1	0.05	0.3	0.05	0.04	0.02	0.02
5	0.11	0.03	0.01	0.00	0.05	0.02	0.01	0.00
10	0.11	0.02	0.00	0.00	0.05	0.02	0.00	0.00
20	0.1	0.01	0.00	0.00	0.05	0.01	0.00	0.00
∞	0.09	0.01	0.00	0.00	0.04	0.01	0.00	0.00

$\lambda = 0.818, \tau = 0.025, \sigma = 0.25, x = 0.93,\ b_1^L = 0.089,\ b_1^{NL} = 0.09$

	a_k					b_k				
$N \setminus k$	2	3	4	5	6	2	3	4	5	6
1	0.74	0.68	0.62	0.57	0.53	0.09	0.08	0.08	0.07	0.07
5	0.72	0.59	0.47	0.35	0.25	0.09	0.08	0.07	0.06	0.05
10	0.69	0.49	0.3	0.18	0.1	0.09	0.08	0.07	0.05	0.03
20	0.68	0.48	0.3	0.15	0.06	0.09	0.08	0.07	0.05	0.03
∞	0.68	0.47	0.25	0.08	0.01	0.09	0.08	0.06	0.04	0.01

$\lambda = 0.4, \tau = 0.05, \sigma = 0.5,\ x = 0.53,\ b_1^L = 0.063,\ b_1^{NL} = 0.068$

	a_k				b_k			
$N \setminus k$	2	3	4	5	2	3	4	5
1	0.19	0.09	0.04	0.02	0.04	0.02	0.01	0.01
5	0.11	0.02	0.00	0.00	0.03	0.01	0.00	0.00
10	0.1	0.01	0.00	0.00	0.03	0.01	0.00	0.00
20	0.09	0.01	0.00	0.00	0.03	0.00	0.00	0.00
∞	0.09	0.01	0.00	0.00	0.03	0.00	0.00	0.00

$\lambda = 0.82, \tau = 0.05, \sigma = 0.5, x = 0.92,\ b_1^L = 0.089,\ b_1^{NL} = 0.086$

	a_k					b_k				
$N \setminus k$	2	3	4	5	6	2	3	4	5	6
1	0.74	0.67	0.61	0.56	0.51	0.08	0.08	0.07	0.07	0.06
5	0.69	0.55	0.41	0.29	0.2	0.08	0.07	0.06	0.04	0.03
10	0.69	0.52	0.35	0.21	0.11	0.08	0.07	0.06	0.04	0.02
20	0.67	0.48	0.28	0.13	0.04	0.08	0.07	0.06	0.03	0.01
∞	0.67	0.47	0.24	0.07	0.01	0.08	0.07	0.06	0.02	0.00

Below are the values where the mean length of an ‘*on*’ period is 5 times larger then the mean length of an ‘*off*’ period. Here $a_0 = 0.83, b_0 = 0.17$.

$\lambda = 0.4, \tau = 0.025, \sigma = 0.125, x = 0.8,\ b_1^L = 0.146,\ b_1^{NL} = 0.152$

	a_k				b_k			
$N \setminus k$	2	3	4	5	2	3	4	5
1	0.21	0.14	0.09	0.07	0.12	0.09	0.08	0.06
5	0.15	0.05	0.02	0.01	0.12	0.06	0.02	0.02
10	0.13	0.03	0.01	0.00	0.12	0.06	0.02	0.01
20	0.13	0.02	0.01	0.00	0.12	0.06	0.02	0.01
∞	0.12	0.02	0.00	0.00	0.12	0.06	0.02	0.00

$\lambda = 0.75, \tau = 0.025, \sigma = 0.125, x = 0.96,\ b_1^L = 0.164,\ b_1^{NL} = 0.165$

	a_k					b_k				
$N \setminus k$	2	3	4	5	6	2	3	4	5	6
1	0.69	0.64	0.59	0.57	0.53	0.16	0.16	0.15	0.15	0.14
5	0.65	0.55	0.46	0.38	0.31	0.16	0.16	0.15	0.14	0.14
10	0.62	0.48	0.35	0.25	0.17	0.16	0.16	0.14	0.12	0.09
20	0.62	0.47	0.32	0.2	0.11	0.16	0.16	0.14	0.12	0.09
∞	0.62	0.47	0.29	0.14	0.05	0.16	0.16	0.14	0.12	0.09

$\lambda = 0.4, \tau = 0.05, \sigma = 0.25, x = 0.69,\ b_1^L = 0.133,\ b_1^{NL} = 0.142$

	a_k				b_k			
$N \setminus k$	2	3	4	5	2	3	4	5
1	0.21	0.13	0.08	0.05	0.1	0.7	0.05	0.03
5	0.15	0.05	0.02	0.01	0.1	0.05	0.02	0.01
10	0.13	0.03	0.01	0.00	0.09	0.04	0.01	0.00
20	0.12	0.02	0.00	0.00	0.09	0.03	0.01	0.00
∞	0.12	0.02	0.00	0.00	0.09	0.03	0.00	0.00

$\lambda = 0.75, \tau = 0.05, \sigma = 0.25, x = 0.94,\ b_1^L = 0.162,\ b_1^{NL} = 0.164$

	a_k					b_k				
$N \setminus k$	2	3	4	5	6	2	3	4	5	6
1	0.68	0.63	0.58	0.54	0.51	0.16	0.15	0.14	0.13	0.13
5	0.65	0.54	0.44	0.46	0.28	0.16	0.15	0.13	0.11	0.09
10	0.64	0.51	0.37	0.24	0.15	0.16	0.14	0.12	0.09	0.07
20	0.63	0.45	0.28	0.16	0.07	0.16	0.13	0.11	0.09	0.05
∞	0.63	0.46	0.28	0.12	0.03	0.16	0.15	0.12	0.09	0.04

$\lambda = 0.4, \tau = 0.1, \sigma = 0.5, x = 0.59,\ b_1^L = 0.119,\ b_1^{NL} = 0.128$

	a_k				b_k			
$N \setminus k$	2	3	4	5	2	3	4	5
1	0.21	0.11	0.06	0.04	0.08	0.05	0.03	0.02
5	0.14	0.03	0.01	0.00	0.06	0.02	0.00	0.00
10	0.12	0.02	0.00	0.00	0.06	0.01	0.00	0.00
20	0.12	0.02	0.00	0.00	0.06	0.01	0.00	0.00
∞	0.11	0.01	0.00	0.00	0.06	0.01	0.00	0.00

$\lambda = 0.75, \tau = 0.1, \sigma = 0.5, x = 0.92,\ b_1^L = 0.162,\ b_1^{NL} = 0.164$

	a_k					b_k				
$N \setminus k2$	3	4	5	6		2	3	4	5	6
1	0.68	0.63	0.57	0.53	0.49	0.15	0.14	0.13	0.12	0.11
5	0.65	0.53	0.41	0.3	0.21	0.15	0.14	0.12	0.09	0.07
10	0.62	0.45	0.29	0.17	0.09	0.15	0.13	0.1	0.07	0.03
20	0.62	0.45	0.29	0.16	0.07	0.15	0.13	0.1	0.07	0.03
∞	0.62	0.45	0.26	0.09	0.02	0.15	0.13	0.1	0.05	0.01

The last is the 'worst' case where the mean length of an 'on' period is only 2.5 times longer than the mean length of an '*off*' period. Here $a_0 = 0.71, b_0 = 0.29$.

$\lambda = 0.4, \tau = 0.1, \sigma = 0.25, x = 0.75,\ b_1^L = 0.237,\ b_1^{NL} = 0.251$

	a_k				b_k			
$N \setminus k$	2	3	4	5	2	3	4	5
1	0.26	0.18	0.13	0.09	0.19	0.14	0.11	0.08
5	0.19	0.08	0.03	0.01	0.19	0.11	0.06	0.02
10	0.19	0.07	0.02	0.01	0.19	0.1	0.04	0.01
20	0.18	0.06	0.02	0.00	0.18	0.09	0.03	0.01
∞	0.17	0.05	0.01	0.00	0.18	0.09	0.02	0.01

$\lambda = 0.64, \tau = 0.1, \sigma = 0.25, x = 0.95,\ b_1^L = 0.278,\ b_1^{NL} = 0.281$

	a_k					b_k				
$N \setminus k$	2	3	4	5	6	2	3	4	5	6
1	0.59	0.55	0.52	0.49	0.46	0.27	0.25	0.24	0.23	0.22
5	0.55	0.47	0.37	0.32	0.25	0.27	0.25	0.22	0.19	0.15
10	0.55	0.45	0.34	0.25	0.16	0.27	0.25	0.22	0.18	0.13
20	0.55	0.45	0.33	0.22	0.13	0.27	0.25	0.22	0.18	0.13
∞	0.55	0.43	0.3	0.17	0.08	0.27	0.25	0.22	0.17	0.11

4. SKETCH OF PROOFS

The proofs of Theorems 1–3 use the statements of theorem 4, this theorem deals with differential equations and its proof does not depend on the proofs of Theorems 1–3.

The proof of Theorem 1 uses the known result connecting the convergence of semigroups with the convergence of their generators, [4] (ch. 1, Theorem 6.1).

The proofs of Theorems 1 and 2 follow those of [1] and [3] and are omitted.

Proof of theorem 3

The process $(\mathbf{U}_N(t), \mathbf{V}_N(t))$ is a Markov process with denumerable number of states $\mathcal{U}_N$, all states are attainable. To prove topic (a) of the theorem it is sufficient to show that:

i) any subset of $\mathcal{U}_N$ is bounded if on this subset $\sum_{k=1}^{\infty}(g_k + h_k)$ is bounded,

ii) for any $(\mathbf{g}, \mathbf{h}) \in \mathcal{U}_N$ the inequality $\sup_{t>0} \sum_{k=1}^{\infty}(u_k(t) + v_k(t)) < \infty$ takes place, where $u_k(t) = \left(\mathbf{T}_N(t)G^{(k)}\right)(\mathbf{g}, \mathbf{h})$, $v_k(t) = \left(\mathbf{T}_N(t)H^{(k)}\right)(\mathbf{g}, \mathbf{h})$ $G^{(k))} and\ H^{(k))}$ are defined in (12). .

Statement i) is checked easily.

To prove ii) consider sequences $(\mathbf{g}, \mathbf{h}) \in \mathcal{U}_N$, and find the action of operators $\mathbf{A}_N,\ \mathbf{B}_N$ on functions $G^{(k))},\ H^{(k))},\ k = 0, 1, 2, \dots$.

For linear model

$$\mathbf{A}_N G^{(k)}(\mathbf{g}, \mathbf{h}) = \lambda(g_{k-1} - g_k) - g_k + g_{k+1} - \tau g_k + \sigma h_k. \tag{14.1}$$

$$\mathbf{A}_N H^{(k)}(\mathbf{u},\mathbf{v}) = \lambda(h_{k-1}-h_k)+\tau g_k-\sigma h_k, \tag{14.2}$$

$$\mathbf{A}_N\left(G^{(k)}(\mathbf{g},\mathbf{h})+H^{(k)}(\mathbf{g},\mathbf{h})\right) = \lambda(g_{k-1}+h_{k-1}-g_k-h_k)-g_k+g_{k+1}, \tag{14.3}$$

For nonlinear model

$$\mathbf{B}_N G^{(k)}(\mathbf{g},\mathbf{h}) = \lambda(g_{k-1}-g_k)(g_{k-1}+g_k+h_{k-1}+h_k)-g_k+g_{k+1}$$
$$-\tau g_k+\sigma h_k, \tag{15.1}$$

$$\mathbf{B}_N H^{(k)}(\mathbf{g},\mathbf{h}) = \lambda(h_{k-1}-h_k)(g_{k-1}+g_k+h_{k-1}+h_k)+\tau g_k-\sigma h_k$$
$$\le 2\lambda(h_{k-1}-h_k)+\tau g_k-\sigma h_k, \tag{15.2}$$

$$\mathbf{B}_N\left(G^{(k)}(\mathbf{g},\mathbf{h})+H^{(k)}(\mathbf{g},\mathbf{h})\right) = \lambda((g_{k-1}+h_{k-1})^2-(g_k+h_k)^2)-g_k+g_{k+1}. \tag{15.3}$$

Consider $Z^u=\sum_{k=1}^{\infty} g_k z^k$, $Z^v=\sum_{k=1}^{\infty} h_k z^k$. By (14.2,3),(15.2,3) we have both in the linear and nonlinear cases

$$\frac{d(Z^u+(1+\varepsilon)Z^v)}{dt} \le \lambda(g_0+(1+\varepsilon)h_0-g_1,$$
$$+\lambda(z-1)(Z^u+Z^v+2\varepsilon Z^v)+(\frac{1}{z}-1)Z^u+\varepsilon\tau Z^u-\varepsilon\sigma Z^v. \tag{16}$$

The direct check shows that if inequality (1) is valid it is possible to find such $z, z>1$, $\varepsilon>0$, $A>0$ that

$$\lambda(z-1)+\frac{1}{z}-1+\varepsilon\tau<0,\ \ \lambda(z-1)(1+2\varepsilon)-\varepsilon\sigma<0,$$

$$\frac{d(Z^u+(1+\varepsilon)Z^v}{dt}<0 \text{ as } Z^u+Z^v>A.$$

That proves the boundedness of $\sum_{k=1}^{\infty}(u_{k,N}+v_{k,N})$ (and the ergodicity of the processes $(\mathbf{U}_N,\mathbf{V}_N)$). That also proves that the inequality (13) is valid and $u_k(t)$ and $v_k(t)$ decrease in k at least exponentially. The bounds for z, ε and A do not depend on N, therefore the estimates are uniform in N.

(b). The set of invariant measures μ_N for $(\mathbf{U}_N,\mathbf{V}_N)$ is compact in compact set $\bar{\mathcal{U}}$. It follows from Theorem 1 that any measure μ that is a limiting point for μ_N is invariant under the semigroup $\mathbf{T}$. The exponential and uniform in N decrease in k of $(u_{k,N}+v_{k,N}$ indicates that $\mu(\mathcal{U})=1$, and Theorem 4 states that such μ is unique. △

Proof of Theorem 4. Most of statements presented below are similar to the corresponding statements of [1] and we omit there proofs.

Lemma 1. *If initial values* $(g,h)\in\bar{\mathcal{U}}$ *then for all* $t,\ t<\infty$, *there exists a unique solution* $(u(t),v(t))$ *to problems* (3) *and* (5) *and* $(u(t),v(t))\in\bar{\mathcal{U}}$.

The proof is omitted.

Lemma 2. *Let the initial conditions* $(g,h)\in\mathcal{U}$ *and condition* (1) *be valid. Then for all* $t,\ t<\infty$, *the solution to* (3), (5) *is* $\in\mathcal{U}$.

Proof. The arguments similar to the arguments of the proof of Theorem 3 (see (14) - (17)) show that $u_k,\ v_k$ decrease at least exponentially as $k\to\infty$. △

Lemma 3. *If* $(g^{(1)},h^{(1)})\in\mathcal{U}$, $(g^{(2)},h^{(2)})\in\mathcal{U}$, *then both in linear and nonlinear case*

$$\lim_{t\to\infty}|u_k(t,g^{(1)},h^{(1)})-u_k(t,g^{(2)},h^{(2)})|=0,\ \ \lim_{t\to\infty}|v_k(t,g^{(1)},h^{(1)})-v_k(t,g^{(2)},h^{(2)})|=0.$$

The proof is omitted.

Lemma 4. *Under condition* (1) *problems* (3.1) – (3.3) *and* (5.1) – (5.3) *have stationary solutions* $(a^L, b^L) \in \mathcal{U}$, $(a^{NL}, b^{NL}) \in \mathcal{U}$, *both pairs are unique,* (a_k^{NL}, b_k^{NL}) *decrease superexponentially in* k.

PROOF. In both cases formal summation of equations (4) and (6) shows that a_k and b_k are expressed via a_{k-1} and b_{k-1}. By (3.3), (5.3) both in linear and nonlinear case there exist the limit $a_0 = \lim_{t\to\infty} u_0(t) = \dfrac{\sigma}{\sigma+\tau}$, $b_0 = \lim_{t\to\infty} v_0(t) = \dfrac{\tau}{\sigma+\tau}$. The induction in k shows the existence of $\lim_{t\to\infty} u_k(t)$, $\lim_{t\to\infty} v_k(t)$, $k = 1, 2, \ldots$. The limit sequences are in $\mathcal{U}$ and it is easy to prove that they satisfy (4) and (6). Stationary solutions are unique in $\mathcal{U}$ and $\lim_{t\to\infty}(u(t), v(t)) = (a, b)$ by Lemma 3.

In nonlinear case for large k by (6.2)

$$b_k \sim \frac{\lambda}{\sigma}[b_{k-1}^2 + a_{k-1}b_{k-1}] + \frac{\lambda\tau}{\sigma}(a_{k-1} + b_{k-1})^2 \le C(a_{k-1} + b_{k-1})^2,$$

where C is a constant. Let $C_1 = \max\{\lambda, C\}$ and k_0 be such that $C_1(a_{k_0} + b_{k_0}) = \kappa < 1$. Then

$$a_{k_0+1} + b_{k_0+1} < C_1(a_{k_0} + b_{k_0})^2, \ \ a_{k_0+2} + b_{k_0+2} < C_1^3(a_{k_0} + b_{k_0})^4,$$

$$a_{k_0+j} + b_{k_0+j} < C_1^{2^j-1}(a_{k_0} + b_{k_0})^{2^j} = \kappa^{2^j-1}(a_{k_0} + b_{k_0}),$$

and $a_k + b_k$ decrease superexponentially as $k \to \infty$. △

Theorem 4 follows from Lemmas 1 - 4.△

We omit the proof of Theorem 5.

СПИСОК ЛИТЕРАТУРЫ

1. Н.Д. Введенская, Р.Л. Добрушин, Ф.И. Карпелевич, Система обслуживания с выбором наименьшей из двух очередей – асимптотический подход, *Пробл. передачи информ.*, 1996, **32**, № 1, стр. 15–27.
2. P. Jacquet, N.D. Vvedenskaya, *ON/OFF Sources in an Interconnection Network: Performance Analysis when Packets are Routed to the Shortest Queue of two Randomly Selected Nodes*, Rapport de Recherche No 3570, INRIA, pp. 1–35, December 1998.
3. J.B. Martin, Yu. M. Suhov, Fast Jackson networks, *J. Appl. Prob.*,1999, **9**, pp. 854–870.
4. S. N. Ethier, T. G. Kurtz, *Markov Processes Characterization and Convergence*, New York: Willey, 1986.