UNIFORM IN TIME BOUNDS FOR "NO-WAIT" PROBABILITY IN QUEUES OF $M_t/M_t/S$ TYPE

Alexander Zeifman
Vologda State University,
Vologda, Russia,
Institute of Informatics Problems
of the FRC CSC RAS, Moscow, Russia
ISEDT RAS

Anna Korotysheva
Yacov Satin
Galina Shilova
Vologda State University,
Vologda, Russia,
Institute of Informatics Problems
of the FRC CSC RAS, Moscow, Russia

Rostislav Razumchik
Institute of Informatics Problems
of the FRC CSC RAS, Moscow, Russia,
Peoples' Friendship University,
Moscow, Russia

Victor Korolev
Moscow State University,
Moscow, Russia
Institute of Informatics Problems
of the FRC CSC RAS, Moscow, Russia

Sergey Shorgin
Institute of Informatics Problems
of the FRC CSC RAS, Moscow, Russia

KEYWORDS

inhomogeneous continuous-time Markov chain, approximation bounds

ABSTRACT

In this paper we present new analytical results concerning long-term staffing problem in high-level telecommunication service systems. We assume that a service system can be modelled either by a classic $M_t/M_t/S$ queue, or $M_t/M_t/S$ queue with batch service or $M_t/M_t/S$ with catastrophes and batch arrivals when empty. The question under consideration is: how many servers guarantee that in the long run the probability of zero delay in a queue is higher than the target probability at all times? Here the methodology is presented, which allows one to construct uniform in time upper bound for the value of S in each of the three cases and does not require the calculation of the limiting distribution. These upper bounds can be easily computed and are accurate enough whenever the arrival intensity is low, but become rougher as the arrival intensity is further increased. In the numerical section one compares the accuracy of the obtained bounds with the exact vales of S, obtained by direct numerical computation of the limiting distribution.

1 INTRODUCTION

In this paper consideration is given to three classes of continuous-time Markov chains, which describe the behaviour of multi-server queueing systems of type $M_t/M_t/S$ with a single queue of infinite capacity. Specifically our attention is paid to a classic $M_t/M_t/S$ queue, $M_t/M_t/S$ queue with batch service, and $M_t/M_t/S$ with catastrophes and batch arrivals when empty. The problem under consideration can be formulated as follows: determine the number of servers S, which guarantees that in the long run the probability of zero waiting time in a queue is higher than the target probability at all times[1].

One can think of two common approaches to the problem. The first one is the numerical. It requires the calculation of the limiting system's state probability distribution (which still depends on time t due to inhomogeneity), by truncating the countable state space of the Markov chain. The truncation must be wise in the sense that the calculation errors must be kept low. Having obtained the limiting probabilities one can find the appropriate value for S by exhaustive search. The second approach is the construction of bounds for the value of S without the calculation of the limiting distribution but by using general inequalities known for some classes of continuous-time Markov chains. The bounds obtained in such a way are not sharp, but do not require the solution of the system of ordinary differential equations. Due to the simplicity of their calculation in some cases (for example, when one needs to know only the order of magnitude of S) they still can be valuable. Moreover even crude bounds facilitate the search of exact values of S, especially when the traffic intensity is high.

In this paper we dwell on the second approach and provide the new methodology which allows one to compute bounds for S in case of three different classes of continuous-time Markov chains introduced above. This methodology heavily relies on the results of Zeifman and Sipin et al. (2015) and thus here one will omit most of the intermediate calculations. It is also worth noticing

[1] That is a uniform in time bound.

Proceedings 30th European Conference on Modelling and
Simulation ©ECMS Thorsten Claus, Frank Herrmann,
Michael Manitz, Oliver Rose (Editors)
ISBN: 978-0-9932440-2-5 / ISBN: 978-0-9932440-3-2 (CD)

that the study of qualitative and quantitative properties of inhomogeneous continuous-time Markov chains has received considerable attention since 1980's (one can refer to papers Di Crescenzo et al. (2008)-Zeifman (1995), Zeifman and Korotysheva et al. (2012)-Margoulis (2013) and references therein). Other interesting results related to this paper can be found in Zeifman and Satin et al. (2009) and Zeifman, Satin and Korolev et al. (2014).

The result, which seems to be the most useful for the computation of the bounds for S, can be found in Zeifman and Sipin et al. (2015). According to it, the limiting state probabilities of an inhomogeneous Markov chain $\{X(t),\ t \geq 0\}$ defined on non-negative integers which described the behaviour of one of the three queueing systems mentioned at the beginning of the Section, satisfy the inequality

$$\limsup_{t\to\infty} \sum_i d_i p_i(t) \leq K, \qquad (1)$$

where K is some constant and $\{d_i\}$, $i = 0, 1, 2, \dots$ is an increasing sequence of positive numbers, with $d_0 = 1$. As the inequality (1) implies that

$$\limsup_{t\to\infty} \sum_{i\geq S} d_i p_i(t) \leq \frac{K}{d_S},$$

then the solution of the considered problem can be found from the condition:

$$\Pr\left(X(t) \leq S\right) \geq 1 - \frac{K}{d_S}, \qquad (2)$$

for sufficiently large t. In the rest of the paper we show how one can obtain in a unified way the values for constants in (2) for each of the chains, mentioned at the beginning of the Section. In order to illustrate the accuracy of the obtained bounds for S we present the comparison of their values with the exact[2] values of S obtained by applying the first approach (i.e. by truncation of the state space and direct calculation of the limiting probabilities). The details of the direct calculation are omitted and can be found in Zeifman and Korotysheva et al. (2015). From the examples one can see, that the bounds are good as long as the arrival intensity is not too high. As the arrival intensity grows, the bounds become crude.

The problem considered in the paper is closely related to the staffing problem in service systems (such as telephone call centers), in which there is an objective of immediately serving all incoming requests. Specifically, the question is: if one needs to keep, say, more than X% of incoming requests to be served without waiting, what is the number of service units which allows the system to achieve this goal? There is a big number of research papers devoted to stochastic modelling of service systems in various settings (see, for example, Whitt (2002), Whitt (2002) and references therein) and specifically to the latter question (see, for example, Whitt (2002), Whitt and Song-Hee (2014), Whitt and Liu (2012), Engblom and Pender (2014) and references therein). But in case when the intensities of all the processes (arrivals, services, breakdowns, etc.) vary periodically over time, this question to our knowledge remains open and challenging. The results presented in this paper may give insights into the influence of periodicities on the management of the service system in case when it is possible to model it as $M_t/M_t/S$ queue with possible group services, breakdowns and group arrivals after recovery.

The paper is organized as follows. In the next section description of the general inhomogeneous Markov chain under consideration is given. In Section 3 we show how one can construct the bounds for the value of S for three particular cases of the chain. In the numerical section one demonstrates the accuracy of the obtained results.

2 DESCRIPTION OF THE MODEL

Let $\{X(t),\ t \geq 0\}$ be an inhomogeneous continuous-time Markov chain describing the evolution of the number of customers in the system which is of $M_t/M_t/S$ type. Denote the state space of $\{X(t),\ t \geq 0\}$ by $E = \{0, 1, 2, \dots\}$. Throughout the paper we assume that for the transition probabilities it holds that

$$\Pr\left(X(t+h) = j | X(t) = i\right) = q_{i,j}(t)\, h + \alpha_{i,j}(t,h),\ \ j \neq i,$$

where all $\alpha_i(t,h) = -\sum_{j\neq i} \alpha_{i,j}(t,h)$ are $o(h)$ uniformly in i, i. e., $\sup_i |\alpha_i(t,h)| = o(h)$. Additionally we assume that all intensity functions are linear combinations of a finite number of locally integrable on $[0,\infty)$ non-negative functions. Let intensity matrix of the chain be $Q(t) = \left(q_{i,j}(t)\right)$ with $q_{ii}(t) = -\sum_{j\neq i} q_{ij}(t)$.

Let $a_{ij}(t) = q_{ji}(t)$ for $j \neq i$, then $a_{ii}(t) = -\sum_{j\neq i} a_{ji}(t)$. According to our approach from Zeifman (1995); Zeifman and Leorato et al. (2006) we assume that the intensity matrix is essentially bounded, i. e.

$$|a_{ii}(t)| \leq L < \infty,$$

for almost all $t \geq 0$.

Denote by $p_{ij}(s,t) = Pr\{X(t) = j | X(s) = i\}$, $i, j \geq 0$, $0 \leq s \leq t$ the transition probability functions of the chain $\{X(t), t \geq 0\}$ and by $p_i(t) = Pr\{X(t) = i\}$ – the state probabilities. By $\mathbf{p}(t) = (p_0(t), p_1(t), \dots)^T$, $t \geq 0$, denote the column vector of state probabilities.

Probabilistic dynamics of the considered chain $\{X(t), t \geq 0\}$ is given by the forward Kolmogorov system

$$\frac{d\mathbf{p}(t)}{dt} = A(t)\,\mathbf{p}(t), \qquad (3)$$

where $A(t) = Q^T(t)$ is the transposed intensity matrix of the chain. Throughout the paper by $\|\cdot\|$ we denote the l_1-norm, i. e., $\|\mathbf{x}\| = \sum_i |x_i|$, and $\|B\| = \sup_j \sum_i |b_{ij}|$ for $B = \left(b_{ij}\right)_{i,j=0}^{\infty}$. Let Ω be the set all stochastic vectors, i. e. l_1 – vectors with nonnegative coordinates and unit norm. Then we have $\|A(t)\| = 2\sup_k |a_{kk}(t)| \leq 2L$ for almost all $t \geq 0$. Hence, the operator function $A(t)$ from l_1 into itself is bounded for almost all $t \geq 0$ and locally integrable

[2] In fact they are approximate too, but sharper.

on $[0;\infty)$. Therefore, we can consider (3) as a differential equation in the space l_1 with bounded operator.

It is well known (see Dalecki and Krein (1974)) that the Cauchy problem for differential equation (3) has a unique solution for arbitrary initial condition, and $\mathbf{p}(s) \in \Omega$ implies $\mathbf{p}(t) \in \Omega$ for $t \geq s \geq 0$.

Denote by $E(t,k) = E\{X(t)|X(0)=k\}$ the expected value (mean) of the chain $X(t)$ at moment t under initial condition $X(0)=k$.

Recall that chain $X(t)$ is called *weakly ergodic*, if $\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| \to 0$ as $t \to \infty$ for any initial conditions $\mathbf{p}^*(0), \mathbf{p}^{**}(0)$, where $\mathbf{p}^*(t)$ and $\mathbf{p}^{**}(t)$ are the corresponding solutions of (3). chain $X(t)$ has the *limiting mean* $\varphi(t)$, if $\lim_{t\to\infty}(\varphi(t) - E(t,k)) = 0$ for any k.

3 THEORETICAL BOUNDS

3.1 $M_t/M_t/S$ system

The inhomogeneous continuous-time Markov chain $\{X(t),\ t \geq 0\}$ describing the behaviour of the ordinary $M_t/M_t/S$ system is of birth-and-death type. Its birth and death intensities are equal to $q_{n,n+1}(t) = \lambda_n(t) = \lambda(t)$ and $q_{n,n-1}(t) = \mu_n(t) = \min(n,S)\mu(t)$, respectively.

Consider an increasing sequence of positive numbers $\{d_i\}$, $i = 1,2,\ldots$, $d_1 = 1$, and the corresponding triangular matrix D:

$$D = \begin{pmatrix} d_1 & d_1 & d_1 & \cdots \\ 0 & d_2 & d_2 & \cdots \\ 0 & 0 & d_3 & \cdots \\ & \ddots & \ddots & \ddots \end{pmatrix}$$

Let l_{1D} be the space of sequences:

$$l_{1D} = \left\{\mathbf{z} = (p_1, p_2, \cdots)^T : \|\mathbf{z}\|_{1D} \equiv \|D\mathbf{z}\| < \infty\right\}.$$

Put

$$d = \inf_{i\geq 1} d_i = 1, \quad W = \inf_{i\geq 1} \frac{d_i}{i}, \quad g_i = \sum_{n=1}^{i} d_n.$$

Consider the following expressions:

$$\alpha_k(t) = \lambda_k(t) + \mu_{k+1}(t) - \frac{d_{k+1}}{d_k}\lambda_{k+1}(t) - \frac{d_{k-1}}{d_k}\mu_k(t), \quad k \geq 0, \tag{4}$$

and

$$\alpha(t) = \inf_{k\geq 0} \alpha_k(t). \tag{5}$$

The property $\mathbf{p}(t) \in \Omega$ for any $t \geq 0$ allows one to use the normalization condition and write $p_0(t) = 1 - \sum_{i\geq 1} p_i(t)$. Then from (3) we obtain the following system of differential equations for the considered chain $\{X(t),\ t \geq 0\}$:

$$\frac{d\mathbf{z}(t)}{dt} = B(t)\mathbf{z}(t) + \mathbf{f}(t), \tag{6}$$

where $\mathbf{z(t)} = (p_1(t), p_2(t), \ldots)^T$, $\mathbf{f}(t) = (\lambda_0(t), 0, 0, \ldots)^T$, $B(t) = \left(b_{ij}(t)\right)_{i,j=1}^{\infty}$ and

$$b_{ij} = \begin{cases} -(\lambda_0 + \lambda_1 + \mu_1), & \text{if} \quad i = j = 1, \\ \mu_2 - \lambda_0, & \text{if} \quad i = 1, j = 2, \\ -\lambda_0, & \text{if} \quad i = 1, j > 2, \\ -(\lambda_j + \mu_j), & \text{if} \quad i = j > 1, \\ \mu_j, & \text{if} \quad i = j-1 > 1, \\ \lambda_j, & \text{if} \quad i = j+1 > 1, \\ 0, & \text{otherwise.} \end{cases}$$

This is a system of linear non-homogeneous differential equations which solution can be written in the form

$$\mathbf{z}(t) \quad = \quad V(t,0)\mathbf{z}(0) + \int_0^t V(t,\tau)\mathbf{f}(\tau)\,d\tau,$$

where $V(t,\tau) = V(t)V^{-1}(\tau)$ is the Cauchy operator of (6).

Consider equation (6) in the space l_{1D}. We have

$$DBD^{-1} = \begin{pmatrix} -(\lambda_0+\mu_1) & \frac{d_1}{d_2}\mu_2 & 0 & \ddots & \\ \frac{d_2}{d_1}\lambda_1 & -(\lambda_1+\mu_2) & \frac{d_2}{d_3}\mu_3 & 0 & \ddots \\ 0 & \frac{d_3}{d_2}\lambda_2 & -(\lambda_2+\mu_3) & \frac{d_3}{d_4}\mu_4 & 0 \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{7}$$

Note that $\mathbf{f}(t)$ and $B(t)$ are bounded and locally integrable on $[0,\infty)$ as being a vector function and an operator function in l_{1D} respectively. Now, taking into consideration (5), we have the following bound for the logarithmic norm $\gamma(B(t))$ in l_{1D}:

$$\gamma(B)_{1D} = \gamma\left(DB(t)D^{-1}\right)_1 = \sup_{i\geq 0}\left(\frac{d_{i+1}}{d_i}\lambda_{i+1}(t) + \frac{d_{i-1}}{d_i}\mu_i(t) - (\lambda_i(t) + \mu_{i+1}(t))\right) = -\inf_{k\geq 0}(\alpha_k(t)) = -\alpha(t),$$

Hence

$$\|V(t,s)\|_{1D} \leq e^{-\int_s^t \alpha(\tau)\,d\tau},$$

and therefore

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\|_{1D} = \|\mathbf{z}^*(t) - \mathbf{z}^{**}(t)\|_{1D} \leq \leq e^{-\int_s^t \alpha(\tau)\,d\tau}\|\mathbf{p}^*(s) - \mathbf{p}^{**}(s)\|_{1D},$$

for any $t \geq s \geq 0$ and any initial conditions $\mathbf{p}^*(s), \mathbf{p}^{**}(s)$.

Moreover, inequality $\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| \leq 2\|\mathbf{z}^*(t) - \mathbf{z}^{**}(t)\| \leq 4\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\|_{1D}$ implies the following bound:

$$\|\mathbf{p}^*(t) - \mathbf{p}^{**}(t)\| \leq 4e^{-\int_s^t \alpha(\tau)\,d\tau}\sum_{i\geq 1} g_i|p_i^*(s) - p_i^{**}(s)|,$$

Assume that the following bounds hold

$$e^{-\int_\tau^t \alpha(u)du} \le M^* e^{-\alpha^*(t-\tau)}, \quad \lambda(t) \le \Lambda, \qquad (8)$$

for almost all $t \ge 0$. Then we obtain the following inequality

$$\|\mathbf{z}(t)\|_{1D} \le \|V(t)\|_{1D}\|\mathbf{z}(0)\|_{1D} + \int_0^t \|V(t,\tau)\|_{1D}\|\mathbf{f}(\tau)\|_{1D}\, d\tau \le M^* e^{-\alpha^* t}\|\mathbf{z}(0)\|_{1D} + \frac{M^*\Lambda}{\alpha^*}.$$

On the other hand, because all $p_i(t)$ are non-negative, we have

$$\|\mathbf{z}(t)\|_{1D} = \sum_{i\ge1} p_i(t)\sum_{k=1}^{i} d_k \ge \sum_{i\ge N} g_i p_i(t).$$

Hence

$$\sum_{i=N}^{\infty} g_i p_i(t) \le M^* e^{-\alpha^* t}\|\mathbf{z}(0)\|_{1D} + \frac{M^*\Lambda}{\alpha^*},$$

and

$$\sum_{i=N}^{\infty} p_i(t) \le g_N^{-1} M^* e^{-\alpha^* t}\|\mathbf{z}(0)\|_{1D} + \frac{M^*\Lambda}{\alpha^* g_N},$$

for any N and any $t \ge 0$, and we obtain the following theorem.

Theorem 1 *Assume that the arrival and service intensities $\lambda(t)$ and $\mu(t)$ for $M_t/M_t/S$ system are known and satisfy (8). Then the chain $\{X(t),\ t \ge 0\}$, describing the number of customers in the system at time t, is exponentially weakly ergodic, and the following bound holds for any N:*

$$\limsup_{t\to\infty} \Pr\left(X(t) < N\right) \ge 1 - \frac{M^*\Lambda}{\alpha^* g_N}.$$

From the previous theorem is follows that the inequality (3) for the $M_t/M_t/S$ system has the form

$$\limsup_{t\to\infty} \Pr\left(X(t) < S\right) \ge 1 - \frac{M^*\Lambda}{\alpha^* g_S}.$$

3.2 $M_t/M_t/S$ system with group services

The inhomogeneous continuous-time Markov chain $\{X(t),\ t \ge 0\}$ describing the behaviour of the ordinary $M_t/M_t/S$ system has non-zero arrival intensities $q_{n,n+1}(t) = \lambda(t)$, and non-zero service intensities $q_{n,n-i}(t) = \mu_i(t) = \frac{\mu(t)}{i}$ for group of i customers, $1 \le i \le S$. The chain $\{X(t),\ t \ge 0\}$ is a SZK chain described in detail in Satin and Zeifman et al. (2013); Zeifman and Satin et al. (2013); Zeifman and Korotysheva et al. (2015).

The analysis of the limiting behaviour for this chain can be carried out in the same way as it was done in the previous subsection. We have

$$\alpha_i(t) = -a_{ii}(t) - \sum_{k\ge1}\lambda_k(t)\frac{d_{k+i}}{d_i} - \sum_{k=1}^{i-1}(\mu_{i-k}(t) - \mu_i(t))\frac{d_k}{d_i}, \quad (9)$$

and

$$\alpha(t) = \inf_{i\ge1} \alpha_i(t), \qquad (10)$$

instead of (4) and (5), where $\lambda_1(t) = \lambda(t)$, $\lambda_k(t) \equiv 0$ for $k \ge 1$, and $\mu_k(t) = \frac{\mu(t)}{k}$, if $k \le S$, $\mu_k(t) \equiv 0$ for $k > S$. Instead of (7) we obtain

$$DBD^{-1} =$$

$$\begin{pmatrix}
a_{11} & (\mu_1-\mu_2)\frac{d_1}{d_2} & (\mu_2-\mu_3)\frac{d_1}{d_3} & \cdots & (\mu_{r-1}-\mu_r)\frac{d_1}{d_r} & \cdots \\
\lambda_1\frac{d_2}{d_1} & a_{22} & (\mu_1-\mu_3)\frac{d_2}{d_3} & \cdots & (\mu_{r-2}-\mu_r)\frac{d_2}{d_r} & \cdots \\
\lambda_2\frac{d_3}{d_1} & \lambda_1\frac{d_3}{d_2} & a_{33} & \cdots & (\mu_{r-3}-\mu_r)\frac{d_3}{d_r} & \cdots \\
\cdots & & & & & \\
\lambda_{r-1}\frac{d_r}{d_1} & \lambda_{r-2}\frac{d_r}{d_2} & \lambda_{r-3}\frac{d_r}{d_3} & \cdots & a_{rr} & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{pmatrix}.$$

Hence the following theorem holds.

Theorem 2 *Assume that the arrival and service intensities $\lambda(t)$ and $\mu(t)$ for $M_t/M_t/S$ system with group services are known and assume that (8), (9), (10) hold. Then the chain $\{X(t),\ t \ge 0\}$, describing the number of customers in the system at time t, is exponentially weakly ergodic, and the following bound holds for any N:*

$$\limsup_{t\to\infty} \Pr\left(X(t) < N\right) \ge 1 - \frac{M^*\Lambda}{\alpha^* g_N}.$$

From the previous theorem is follows that the inequality (3) for the $M_t/M_t/S$ system with group services has the form

$$\limsup_{t\to\infty} \Pr\left(X(t) < S\right) \ge 1 - \frac{M^*\Lambda}{\alpha^* g_S}.$$

3.3 $M_t/M_t/S$ system with catastrophes and batch arrivals when empty

Assume that the inhomogeneous continuous-time Markov chain $\{X(t),\ t \ge 0\}$ describing the behaviour of the ordinary $M_t/M_t/S$ system with catastrophes and batch arrivals when empty has the following intensity matrix:

$$Q(t) = \begin{pmatrix}
a_{00}(t) & r_1(t) & r_2(t) & r_3(t) & r_4(t) & \cdots & \cdots \\
\beta_1(t)+\mu_1(t) & a_{11}(t) & \lambda_1(t) & 0 & 0 & \cdots & \cdots \\
\beta_2(t) & \mu_2(t) & a_{22}(t) & \lambda_2(t) & 0 & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\beta_r(t) & 0 & \cdots & \mu_r(t) & a_{rr}(t) & \lambda_r(t) & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{pmatrix}.$$

For the detailed description of the system one can refer to Zeifman, Korotysheva and Satin et al. (2015); Zeifman and Satin et al. (2016) and references therein. We will apply the approach from Zeifman, Korotysheva and Satin et al. (2015) or study of ergodic properties and for the construction of bounds for $\{X(t),\ t \geq 0\}$.

Put

$$\beta_*(t) = \inf_i \beta_i(t),$$

and rewrite the forward Kolmogorov equation (3) as

$$\frac{d\mathbf{p}(t)}{dt} = A^*(t)\,\mathbf{p}(t) + \mathbf{g}(t), \quad t \geq 0, \tag{11}$$

where $\mathbf{g}(t) = (\beta_*(t), 0, 0, \dots)^T$, $A^*(t) = \left\{a^*_{ij}(t)\right\}$, and

$$a^*_{ij}(t) = \begin{cases} a_{0j}(t) - \beta_*(t) & \text{if } i = 0 \\ a_{ij}(t) & \text{otherwise.} \end{cases}$$

Let now D be a diagonal matrix

$$D = diag\,(d_0, d_1, d_2, \dots)$$

with elements satisfying the inequalities $1 = d_0 \leq d_1 \leq d_2 \leq \dots$. Consider the corresponding space of sequences $l_{1D} = \{\mathbf{z} = (p_0, p_1, p_2, \dots)\}$ such that $\|\mathbf{z}\|_{1D} = \|D\mathbf{z}\|_1 < \infty$.

Put $\beta_{**}(t) = \inf_i \left(|a^*_{ii}(t)| - \sum_{j\neq i} \left|\frac{d_j}{d_i} a^*_{ji}(t)\right|\right)$.

Then one can obtain the following estimate for the logarithmic norm of the operator function $A^*(t)$ in the l_{1D}-norm:

$$\gamma\,(A^*(t))_{1D} = \gamma\left(DA^*(t)\,D^{-1}\right) =$$

$$\sup_i \left(a^*_{ii}(t) + \sum_{j\neq i} \left|\frac{d_j}{d_i} a^*_{ji}(t)\right|\right) = -\beta_{**}(t).$$

Assume now that the following bounds hold:

$$e^{-\int_\tau^t \beta^{**}(u)du} \leq M e^{-a^*(t-\tau)}, \tag{12}$$

$$\beta^*(t) \leq \Theta, \tag{13}$$

for almost all $0 \leq \tau \leq t$. Then one can bound the solution of the system (11) in the following way:

$$\|\mathbf{p}(t)\|_{1D} = \|U^*(t,0)\,\mathbf{p}(0) + \int_0^t U^*(t,\tau)\,\mathbf{g}(\tau)\,d\tau\| \leq$$

$$\leq \|U^*(t,0)\|_{1D}\|\mathbf{p}(0)\|_{1D} + \int_0^t \|U^*(t,\tau)\|_{1D}\|\mathbf{g}(\tau)\|_{1D} d\tau \leq$$

$$\leq M e^{-a^* t}\|\mathbf{p}(0)\|_{1D} + \frac{M\Theta}{a^*}.$$

On the other hand,

$$\sum_{i\geq N} d_i p_i(t) \leq \|\mathbf{p}(t)\|_{1D},$$

and hence

$$\sum_{i=N}^{\infty} d_i p_i(t) \leq M e^{-a^* t}\|\mathbf{p}(0)\|_{1D} + \frac{M\Theta}{a^*},$$

and

$$\sum_{i=N}^{\infty} p_i(t) \leq d_N^{-1} M e^{-a^* t}\|\mathbf{p}(0)\|_{1D} + \frac{M\Theta}{a^* d_N},$$

for any N and any $t \geq 0$. Thus the following theorem holds.

Theorem 3 *Assume that the intensities $\lambda_i(t)$, $\mu_i(t)$, $\beta_i(t)$ and $r_i(t)$ for $M_t/M_t/S$ with catastrophes and batch arrivals when empty are known. Assume that (12) and (13) hold. Then the chain $\{X(t),\ t \geq 0\}$, describing the number of customers in the system at time t, is exponentially weakly ergodic, and the following bound holds for any N:*

$$\limsup_{t\to\infty} \Pr(X(t) < N) \geq 1 - \frac{M\Theta}{a^* d_N}.$$

From the previous theorem is follows that the inequality (3) for the $M_t/M_t/S$ system with catastrophes and batch arrivals when empty has the form

$$\limsup_{t\to\infty} \Pr(X(t) < S) \geq 1 - \frac{M\Theta}{a^* d_S}.$$

4 NUMERICAL EXAMPLES

In this section we will consider six different examples of inhomogeneous Markov chains describing the behaviour of $M_t/M_t/S$ queues. For each example we give two values of the required number of servers S. One of the values is obtained using theorems from the previous section, and the other value is obtained by direct numerical computation of the limiting probabilities (for more details how the limiting probabilities are found one can refer to Zeifman, Korotysheva and Satin et al. (2015); Zeifman and Korotysheva et al. (2015)). As one will see the accuracy of bounds greatly depends on the arrival intensity and drops significantly as it grows.

Example 1

Consider the classic $M_t/M_t/S$ queue with arrival intensity $\lambda(t) = 1 + \sin 2\pi t$ and service intensity $\mu(t) = 3 + \cos 2\pi t$.

Put $d_{n+1} = 2^n$, $n \geq 1$. Then using results from the subsection 3.1 we have that $g_k = 2^k - 1$ and $\alpha_k(t) \geq \mu(t) - \lambda(t)$ for any $k \geq 1$, S and $t \geq 0$. From (5) it follows that $\alpha(t) \geq 2 + \cos 2\pi t - \sin 2\pi t$. Notice that the inequality (8) holds for $\alpha^* = 2$, $M^* = 2$ and $\Lambda = 2$. Hence the *Theorem* 1 gives us the following bound for the value of S:

$$\limsup_{t\to\infty} \Pr(X(t) < S) \geq 1 - \frac{2}{2^S - 1}. \tag{14}$$

One can see that $S = 4$ guarantees that more than 90% of incoming requests will be served without waiting. On the other hand, if one directly computes the limiting probabilities (by truncation of the state space of the chain), then one can find that in fact $S = 2$ guarantees that 90% of requests have zero waiting time. The behaviour of the probabilities of zero waiting time for $S = 1$ and $S = 2$ is depicted in Fig. 1 and 2.

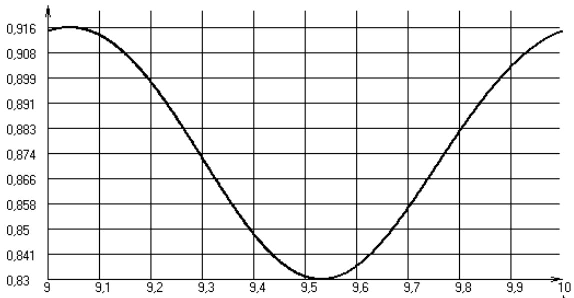

Figure 1: Probability of immediately serving of 80% incoming requests $\Pr(X(t) < 1)$ on $[9, 10]$, with error less then 10^{-6}, one can choose $n = 100$.

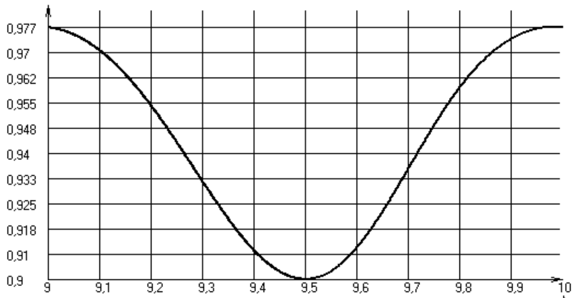

Figure 2: Probability of immediately serving of 90% incoming requests $\Pr(X(t) < 2)$ on $[9, 10]$, with error less then 10^{-6}, one can choose $n = 100$.

Example 2

Consider the $M_t/M_t/S$ queue with group services with the arrival intensity $\lambda(t) = 1+\sin 2\pi t$ and service intensity $\mu(t) = 3+\cos 2\pi t$. This is the same example as in Zeifman and Satin et al. (2013).

Put $d_{n+1} = 2^n$, $n \geq 1$. Then using results from the subsection 3.2 we have that $g_k = 2^k - 1$ and $\alpha_k(t) \geq \mu(t) - \lambda(t)$, for any $k \geq 1$, S and $t \geq 0$. Therefore from (10) it follows that $\alpha(t) \geq 2 + \cos 2\pi t - \sin 2\pi t$, and the inequality (8) holds for $\alpha^* = 2$, $M^* = 2$ and $\Lambda = 2$. Hence the *Theorem* 2 gives us the following bound for the value of S:

$$\limsup_{t\to\infty} \Pr(X(t) < S) \geq 1 - \frac{2}{2^S - 1}. \qquad (15)$$

It can be seen that even with group services, four servers ($S = 4$) guarantee that more than 90% of incoming requests will be served without waiting. But by numerical computation of the limiting probabilities, one can find that in fact $S = 3$ guarantees that 90% of requests have zero waiting time. The behaviour of the probabilities of zero waiting time for $S = 2$ and $S = 3$ is depicted in Fig. 3 and 4.

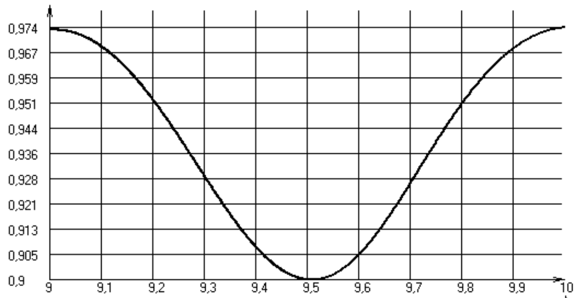

Figure 3: Probability of immediately serving of 87% incoming requests $\Pr(X(t) < 2)$ on $[9, 10]$, with error less then 10^{-6}, one can choose $n = 100$.

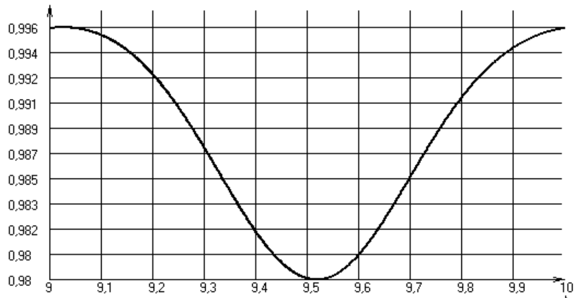

Figure 4: Probability of immediately serving of 90% incoming requests $\Pr(X(t) < 3)$ on $[9, 10]$, with error less then 10^{-6}, one can choose $n = 100$.

Example 3

Consider the classic $M_t/M_t/S$ queue service intensity $\mu(t) = 4 + \cos 2\pi t$ but with very high arrival intensity $\lambda(t) = 100 + 5\sin 2\pi t$. In order to choose the sequence of positive numbers $\{d_i\}$ (see subsection 3.1) we initially suppose that $S \geq 60$. The if one puts $d_{k+1} = 1.02^k$ for $k \geq 1$ then using results from the subsection 3.1 we have that $g_k > 1.02^k$, $k \geq 2$. After little algebra one can find that the inequality (8) is satisfied for $\alpha^* = 2$, $M^* \leq e^2$ and $\Lambda \approx 100$. Hence the *Theorem* 1 gives us the following bound for the value of S:

$$\limsup_{t\to\infty} \Pr(X(t) < S) \geq 1 - \frac{10^3}{2 \cdot 1.02^S}, \qquad (16)$$

Here $S \geq 450$ guarantees that more than 90% of incoming requests will be served without waiting. On the other hand, if one directly computes the limiting probabilities, then one can find that the true value is $S = 37$. The behaviour of the probabilities of zero waiting time for $S = 36$ and $S = 37$ is depicted in Fig. 5 and 6.

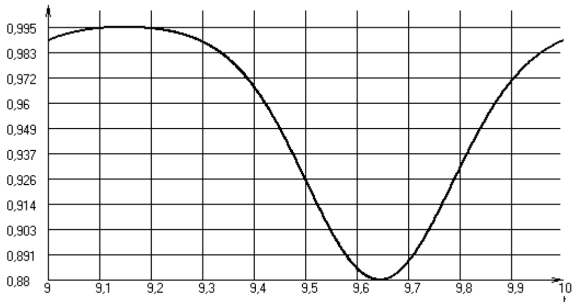

Figure 5: Probability of immediately serving of 85% incoming requests $\Pr(X(t) < 36)$ on $[9, 10]$, with error less then 10^{-6}, one can choose $n = 600$.

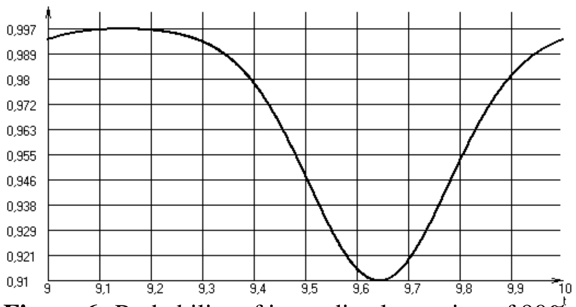

Figure 6: Probability of immediately serving of 90% incoming requests $\Pr(X(t) < 37)$ on $[9, 10]$, with error less then 10^{-6}, one can choose $n = 600$.

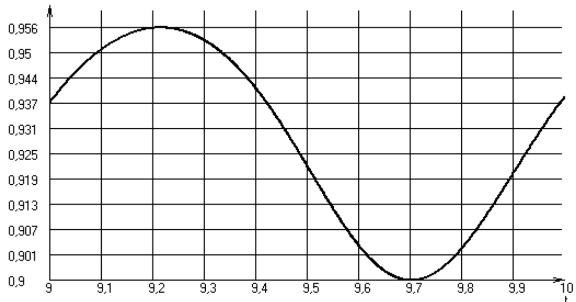

Figure 7: Probability of immediately serving of 88% incoming requests $\Pr(X(t) < 52)$ on $[9, 10]$, with error less then 10^{-6}, one can choose $n = 600$.

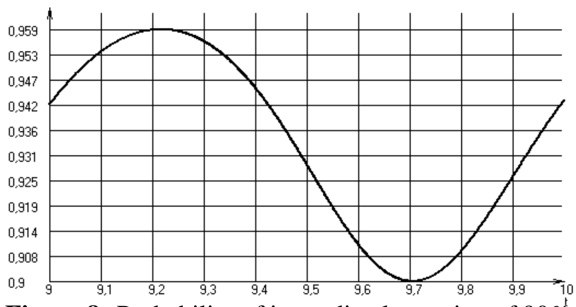

Figure 8: Probability of immediately serving of 90% incoming requests $\Pr(X(t) < 53)$ on $[9, 10]$, with error less then 10^{-6}, one can choose $n = 600$.

Example 4

Consider the $M_t/M_t/S$ queue with group services with the service intensity $\mu(t) = 3 + \cos 2\pi t$ and high arrival intensity $\lambda(t) = 1 + \sin 2\pi t$. Again in order to choose the sequence of positive numbers $\{d_i\}$ (see subsection 3.2) we initially suppose that $S \geq 60$. The if one puts $d_{k+1} = 1.02^k$ for $k \geq 1$ then using results from subsection 3.2 we have that $g_k > 1.02^k$, $k \geq 2$ and the inequality (8) is satisfied for $\alpha^* = 2$, $M^* \leq e^2$ and $\Lambda \approx 100$. Hence the *Theorem* 2 gives us the following bound for the value of S:

$$\limsup_{t\to\infty} \Pr(X(t) < S) \geq 1 - \frac{10^3}{2 \cdot 1.02^S}. \qquad (17)$$

Surprisingly, as in Example 3, here $S \geq 450$ guarantees that more than 90% of incoming requests will be served without waiting. But the true value is $S = 53$. The behaviour of the probabilities of zero waiting time for $S = 52$ and $S = 53$ is depicted in Fig. 7 and 8.

Example 5

Consider the $M_t/M_t/S$ queue with breakdowns and batch arrivals when empty. Let the arrival and service rates be as in Examples 1,2. Assume that the breakdown and batch arrival intensities are state-dependent and equal to $\beta_n(t) = 2 + \cos 2\pi t + \frac{1}{n}$ and $r_n(t) = \frac{1-\sin 2\pi t}{4^n}$, $n \geq 1$, respectively. Notice that breakdowns may happen only when system is not empty.

Put $\beta^*(t) = 1$ and $d_k = \left(\frac{4}{3}\right)^k$ for $k \geq 1$. Then using results from the subsection 3.3 we have that $\beta^{**}(t) \geq \frac{1}{3}$ and (12), (13) hold for $M = 1$, $a^* = \frac{1}{3}$ and $\Theta = 1$. Hence the *Theorem* 3 gives us the following bound for the value of S:

$$\limsup_{t\to\infty} \Pr(X(t) < S) \geq 1 - \frac{3^{S+1}}{4^S}. \qquad (18)$$

One can see that $S \geq 12$ guarantees that more than 90% of incoming requests will be served without waiting. It turns out that this bound is very crude. Indeed from the direct computation of the limiting distribution it follows that in fact already one server ($S = 1$) is enough to guarantee 90% of no-wait. (see Fig. 9).

Example 6

Consider again the $M_t/M_t/S$ queue with breakdowns and batch arrivals when empty but with arrival and service intensities as in Example 3. Assume that the intensities of breakdowns and batch arrivals are $\beta_n(t) = 2 + \cos 2\pi t + \frac{1}{n}$ and $r_n(t) = \frac{1-\sin 2\pi t}{4^n}$, $n \geq 1$, respectively. Again breakdowns may happen only when system is not empty.

Put $\beta^*(t) = 1$, and $d_k = \left(\frac{201}{200}\right)^k$ for $k \geq 1$. Then using results from the subsection 3.3 we have that $\beta^{**}(t) \geq \frac{2}{5}$ and (12), (13) hold for $M = 1$, $a^* = \frac{2}{5}$ and $\Theta = 1$. Hence

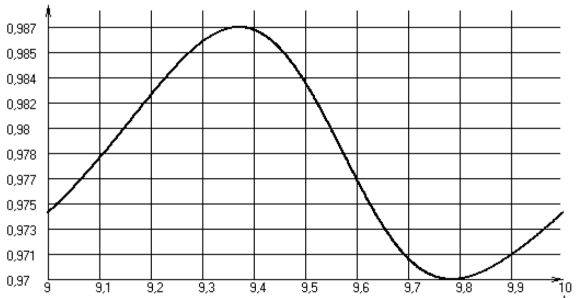

Figure 9: Probability of immediately serving of 90% incoming requests $\Pr(X(t) < 1)$ on $[9, 10]$, with error less then 10^{-6}, one can choose $n = 100$.

the *Theorem* 3 gives us the following bound for the value of S:

$$\limsup_{t\to\infty} \Pr(X(t) < S) \geq 1 - \frac{5 \cdot 200^S}{2 \cdot 201^S}. \qquad (19)$$

One can see that $S \geq 650$ guarantees more than 90% of incoming requests will be served without waiting. But from numerical computation of limiting probabilities it follows that in fact $S = 13$ is enough (see Fig. 10 and 11).

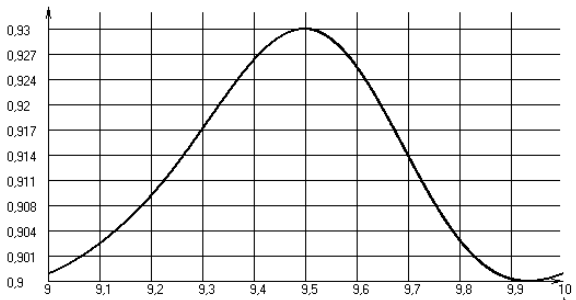

Figure 10: Probability of immediately serving of 88% incoming requests $\Pr(X(t) < 12)$ on $[9, 10]$, with error less then 10^{-6}, one can choose $n = 600$.

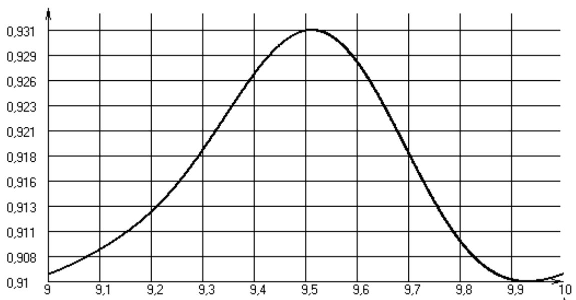

Figure 11: Probability of immediately serving of 90% incoming requests $\Pr(X(t) < 13)$ on $[9, 10]$, with error less then 10^{-6}, one can choose $n = 600$.

5 CONCLUSION

As one can see from the examples the (upper) bounds given by the theorems remain good as long as the arrival intensity is low. As the latter grows the bounds become very inaccurate and additional analysis is needed. Clearly two directions of further research are visible: elaboration of similar bounds for more complex systems and the development of methodology for the estimation of bounds for probabilities given by (3) in cases when the arrival rates are high.

Acknowledgement

This work was supported by Russian Scientific Foundation (Grant No. 14-11-00397).

REFERENCES

Whitt, W., Jennings, B., Mandelbaum, A., Massey, W. 1996. Server Staffing to Meet Time-Varying Demand. Management Science. Vol. 42. No. 10. Pp. 1383-1394.

Whitt, W. 2006. Staffing a Call Center with Uncertain Arrival Rate and Absenteeism. Production and Operations Management. Vol. 15. No. 1. Pp. 88–102.

Whitt, W. 2002. Stochastic Models for the Design and Management of Customer Contact Centers: Some Research Directions. Working paper.

Whitt, W., Song-Hee, K. 2014. Are Call Center and Hospital Arrivals Well Modeled by Nonhomogeneous Poisson Processes? Manufacturing and Service Operations Management. Vol. 16. No. 3. Pp. 464–480.

Whitt, W., Liu, Y. 2012. 'Stabilizing Customer Abandonment in Many-Server Queues with Time-Varying Arrivals. Operations Research. Vol. 60. No. 6. Pp. 1551–1564.

Dalecki, Ju., Krein, M. Stability of solutions of differential equations in Banach space. American Mathematical Society Translations. Vol. 43. 386 p.

Di Crescenzo, A., Giorno, V., Nobile, A., Ricciardi, L. 2008. A note on birth-death processes with catastrophes. Statististical Probabability Letters. Vol. 78. Pp. 2248–2257.

Dudin, A., Karolik, A. 2001. BMAP/SM/1 queue with Markovian input of disasters and non-instantaneous recovery. Performance Evaluation. Vol. 45. Pp. 19–32.

Dudin, A., Nishimura, S. 1999. A BMAP/SM/1 queueing system with Markovian arrival input of disasters. Journal of Applied Probability. Vol. 36. Pp. 868–881.

Dudin, A., Semenova, O. 2004. Stable algorithm for stationary distribution calculation for a BMAP/SM/1 queueing system with markovian input of disasters. Journal of Applied Probability. Vol. 42. Pp. 547–556.

Satin, Ya., Zeifman, A., Korotysheva, A. 2013. On the rate of convergence and truncations for a class of Markovian queueing systems. Theory. Prob. Appl. Vo. 57. Pp. 529C-539.

Zeifman A. 1995. Upper and lower bounds on the rate of convergence for nonhomogeneous birth and death processes. Stochastic Procecces and Applications. Vol. 59. Pp. 157–173.

Zeifman, A., Leorato, S., Orsingher, E., Satin, Ya., Shilova, G. 2006. Some universal limits for nonhomogeneous birth and death processes. Queueing Systems. Vol. 52. Pp. 139–151.

Zeifman, A., Satin, Ya., Korolev, V., Shorgin, S. 2014. On truncations for weakly ergodic inhomogeneous birth and death processes. International Journal of Applied Mathematics and Computer Science. Vol. 24. No. 3. Pp. 503–518.

Zeifman, A., Korotysheva, A. 2012. Perturbation bounds for $M_t|M_t|N$ queue with catastrophes. Stochastic models. Vol. 28. Pp. 49–62.

Zeifman, A., Satin, Ya., Korotysheva, A., Tereshina N. 2009. On the limiting characteristics for $M(t)|M(t)|S$ queue with catastrophes. Informatics and its Applcations. Vol. 3. No. 3. Pp. 16–22. (in Russian)

Zeifman, A., Satin, Ya., Shilova, G., Korolev, V., Bening, V., Shorgin, S. 2013. On $M_t|M_t|S$ type queue with group services. Proceedings of the 27th European Conference on Modelling and Simulation. Pp. 604–609.

Zeifman, A., Korotysheva, A., Satin, Ya., Korolev, V., Bening, V. 2014. Perturbation bounds and truncations for a class of Markovian queues. Queueing Systems. Vol. 76. Pp. 205–221.

Zeifman, A., Korotysheva, A., Satin, Ya., Korolev, V., Shorgin, S., Razumchik, R. 2015. Ergodicity and perturbation bounds for inhomogeneous birth and death processes with additional transitions from and to origin. Int. J. Appl. Math. Comput. Sci. Vol. 25. No. 4. Pp. 787–802.

Zeifman, A., Sipin, A., Korolev, V., Bening, V. 2015. Estimates of some characteristics of multidimensional birth-and-death processes. Doklady Mathematics. Vol. 92. Pp. 695–697.

Zeifman, A., Satin, Ya., Korotysheva, A., Korolev, V., Bening, V. 2016. On a class of Markovian queues with particularities in zero. Doklady Mathematics. Submitted.

Zeifman, A., Korotysheva, A., Razumchik, R., Korolev, V., Shorgin, S. 2015. Some Results For Inhomogeneous Birth-And-Death Process With Application To Staffing Problem In Telecommunication Service Systems. Proceedings of the 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops. Pp. 243–246.

Engblom, S., Pender, J. 2014. Approximations for the Moments of Nonstationary and State Dependent Birth-Death Queues. eprint arXiv:1406.6164.

Margoulis, B. 2007. Transient and periodic solution to the time-inhomogeneous quasi-birth death process. Queueing Systems. Vol. 56. No. 3. Pp. 183–194.

AUTHOR BIOGRAPHIES

ALEXANDER ZEIFMAN is Doctor of Science in physics and mathematics; professor, Heard of Department of Applied Mathematics, Vologda State University; senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences; principal scientist, Institute of Socio-Economic Development of Territories, Russian Academy of Sciences. His email is `a_zeifman@mail.ru` and his personal webpage at `http://uni-vologda.ac.ru/ zai/eng.html`.

YACOV SATIN is Candidate of Science (PhD) in physics and mathematics, associate professor, Vologda State University. His email is `yacovi@mail.ru`.

ANNA KOROTYSHEVA is Candidate of Science (PhD) in physics and mathematics, associate professor, Vologda State University. Her email is `a_korotysheva@mail.ru`.

GALINA SHILOVA is Candidate of Science (PhD) in physics and mathematics, associate professor, Heard of Department of Mathematics, Vologda State University. Her email is `shgn@mail.ru`.

VICTOR KOROLEV is Doctor of Science in physics and mathematics, professor, Head of Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University; leading scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences. His email is `victoryukorolev@yandex.ru`.

SERGEY SHORGIN is Doctor of Science in physics and mathematics, professor, Deputy Director of the Institute of Informatics Problems of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences. His email is `sshorgin@ipiran.ru`.

ROSTISLAV V. RAZUMCHIK received his Ph.D. degree in Physics and Mathematics in 2011. Since then, he has worked as a senior research fellow at Institute of Informatics Problems of the Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (FRC CSC RAS). Currently he holds the position of Head of the Information and Telecommunication System Modelling section at the FRC CSC RAS and associate professor position at Peoples' Friendship University of Russia. His current research activities focus on queueing theory and its applications in performance evaluation of stochastic systems. His email address is `rrazumchik@ipiran.ru`.