Journal of the Korean Institute of Industrial Engineers
Vol. 42, No. 1, pp. 30-37, February 2016.
ISSN 1225-0988 | EISSN 2234-6457
http://dx.doi.org/10.7232/JKIIE.2016.42.1.030

<Original Research Paper>

Analysis of $BMAP^{(r)}/M^{(r)}/N^{(r)}$ Type Queueing System Operating in Random Environment

Chesoong Kim[1†] · Sergey Dudin[2]

[1]Department of Business Administration, Sangji University
[2]Department of Applied Mathematics and Computer Science, Belarusian State University

$BMAP^{(r)}/M^{(r)}/N^{(r)}$ 대기행렬시스템 분석

김제숭[1] · Sergey Dudin[2]

[1]상지대학교 경영학과 / [2]Department of Applied Mathematics and Computer Science, Belarusian State University

A multi-server queueing system with an infinite buffer and impatient customers is analyzed. The system operates in the finite state Markovian random environment. The number of available servers, the parameters of the batch Markovian arrival process, the rate of customers' service, and the impatience intensity depend on the current state of the random environment and immediately change their values at the moments of jumps of the random environment. Dynamics of the system is described by the multi-dimensional asymptotically quasi-Toeplitz Markov chain. The ergodicity condition is derived. The main performance measures of the system are calculated. Numerical results are presented.

Keywords: Multi-Server Queue, Batch Markovian Arrival Process, Random Environment, Stationary Distribution of the System States

1. Introduction

Classical mathematical theory of queues is developed in suggestion that the intervals between customers arrival and service times are defined as random variables with fixed known distributions. However, in real life, these distributions can be essentially changed depending on some random factors (time of day or night, degradation of the transmission thread due to technical reasons, weather conditions, temperature inversions in the atmosphere, frequency interference, noise caused by enterprises and transport, fluctuation of the distance from the nearest base station, parallel service of the customers having more high priority, various malfunctions and equipment failures, etc.). Account of the effect of random factors on the system operation is a very important task in the construction of adequate mathematical models for computation of performance measures of a real world system. In some extent, if the random factors impact only on the instantaneous intensity of customers arrival, this effect can be easily taken into account by means of using more complicated models of the arrival process than the stationary Poisson process, e.g., the batch Markovian arrival process (*BMAP*), see Chakravarthy (2001) and Lucantoni (1991). If the random factors impact on the instantaneous intensity of customers service, this effect can be easily taken into consideration by means of using more complicated distributions of the service time than the exponential one, e.g., a phase-type (*PH*) distribution, see Neuts (1981). However, if the random factors impact on the arrival, service (retrial, breakdown, repair, etc) processes simultaneously, it is necessary to analyze so called queues operating in a random environment (*RE*). The *RE* is assumed to be an external random process with a finite state space, independent of the queuing system. For the fixed state of the *RE*, the sys-

† Corresponding author : Professor Chesoong Kim, Business Administration, Sangji University, Wonju, Kangwon 26339, Korea,
Tel : +82-33-730-0464, Fax : +82-33-743-1115, E-mail : dowoo@sangji.ac.kr
Received June 1, 2015; Revision Received August 30, 2015; Accepted November 30, 2015.

tem operates as a classical queueing system of the corresponding type. But, the parameters of the system (the arrival process, the distribution of the service time, etc.) instantly change their values with the change of the state of the *RE*.

As important early work in the field of queues operating in *RE* the following publications deserve to be mentioned: Gnedenko and Kovalenko (1966), Neuts (1978), Neuts (1981), O'Cinneide and Purdue (1986), Purdue (1974), Yadin and Syski (1979), Yechialy and Naor (1971). A brief history of the development of theory of queues in the *RE*, the reference list and real life examples can be found, e.g., in the papers Cordeiro and Kharoufeh (2012),Wu *et al*. (2011), Yang *et al*. (2013), and Kim *et al*. (2009). In Cordeiro and Kharoufeh (2012),an unreliable M/M/1 retrial queue in a Markovian random environment is analyzed via matrix-analytic methods. Ergodicity condition is proved and approximate distribution of the number of customers in the system is computed. Optimization problem of choosing the arrival and service rates for each environment state is considered. In Wu *et al*. (2011) and Yang *et al*. (2013), the finite source MAP/PH/N retrial queue operating in a random environment is studied. In Wu *et al*. (2011), it is assumed that there is additional MAP arrival process of so called negative customers. The arrival of the negative customer with equal probability goes to any busy server to remove the customer being in service. In Wu *et al*. (2011) and Yang *et al*. (2013), the finite state multi-dimensional Markov chain describing the behavior of the systems is investigated. The algorithms for calculating the stationary state probabilities are elaborated. Main performance measures are obtained and the illustrative numerical examples are presented.

In Kim *et al*. (2009), the BMAP/PH/N/N queue operating in the *RE* is investigated. The arrival flow is described by the batch Markov arrival process (BMAP). The system does not have a buffer. An arriving customer who did not succeed to find a free server upon arrival is lost. Due to possibility of batch arrivals, disciplines of partial admission, complete admission and complete rejection are analyzed. The stationary distribution of the system states and the waiting time distribution are computed. Numerical illustrations are presented. In particular, it is demonstrated that reasonable engineering approximations of performance measures of the system may be very poor. In Kim *et al*. (2007), the BMAP/PH/1/1 queue operating in the *RE* with account of retrials of customers, which do not succeed to get access immediately upon arrival, is investigated. For background information and an overview of the present state of the art in the study of queueing systems operating in the *RE*, the reader is referred also to the papers Wu *et al*. (2011), Yang *et al*. (2013), Kim *et al*. (2009), Kim *et al*. (2007), Krieger *et al*. (2005), and Takine (2005), as well as references therein.

In this paper, we analyze a multi-server queueing system with an infinite buffer operating in the *RE*. The main contribution of our paper is the assumption that not only the parameters of service and arrival process, but also the number of available servers depends on the state of the *RE*. To the best of our knowledge, in all existing papers devoted to analysis of queues operating in the *RE*, the influence of the *RE* on the number of available servers is not considered. The recent paper Kim *et al*. (2014) deals with a very general model of the system in the *RE* with two types of arriving customers and different distributions of service times. However, the number of servers in Kim *et al*. (2014) is assumed to be permanent, not depending on the state of the *RE*, while in this paper we allow such dependence. It is worth to note that the unreliable multi-server queues are a very special example of queues with the random number of available servers, however, to the best of our knowledge, such queues are also not considered in the literature in assumption that the arrival and service processes parameters depend on the *RE*.

In the paper, behavior of the system is described by the continuous time multi-dimensional Markov chain that belongs to the class of asymptotically quasi-Toeplitz Markov chains, see Klimenok and Dudin (2006). This greatly facilitates analysis of the system.

The rest of the paper consists of the following. In the next section, the mathematical model is described in detail. The process of the system states as a multi-dimensional continuous time Markov chain is described in section 3. The generator of this Markov chain is presented. The ergodicity condition for this Markov chain and the problem of computation of its stationary distribution are discussed here. Formulas for computation of some performance measures are obtained in section 4. Some numerical examples showing feasibility of the presented algorithmic results and dependencies of some performance measures on the number of available servers at different states of the *RE* are presented in section 5. Section 6 concludes the paper.

2. Mathematical Model

We analyse a multi-server queueing system with an infinite buffer operating in the Markovian *RE*. The structure of the system under study is represented in <Figure 1>.

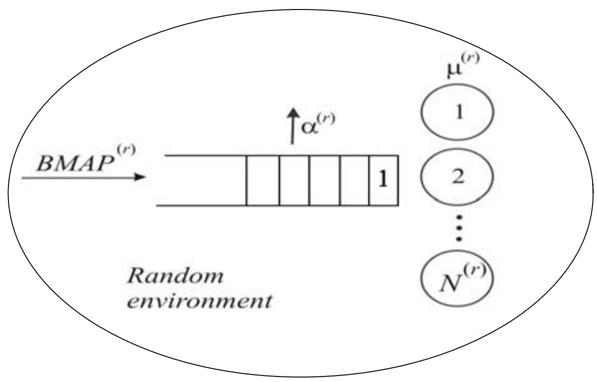

Figure 1. Queueing System under Study

Behavior of the system depends on the state of the *RE*. The *RE* is given by the stochastic process $r_t, t \geq 0$, which is an irreducible regular continuous time Markov chain with the state space $\{1, 2, \cdots, R\}$ and the infinitesimal generator *H*. The vector **d** of the stationary distribution of the *RE* is defined as the unique solution of the system of equations

$$dH = 0, \ \mathrm{de} = 1.$$

Here and in the sequel **e(0)** is a column (row) vector of appropriate size consisting of 1's (0's).

Under the fixed state *r* of the *RE*, the number of available servers is $N^{(r)}, \ r = \overline{1, R}$. Without loss of generality we assume that the states of the *RE* are enumerated in such a way that number of available servers increases with the increase of the number of the state of the *RE*, i.e.,

$$0 \leq N^{(1)} \leq N^{(2)} \leq \cdots \leq N^{(R)}.$$

The customers arrive to the system according to the switching batch Markovian arrival process. The arrival of customers is directed by the stochastic process $v_t, t \geq 0$, with the finite state space $\{0, 1, \cdots, W\}$. Under the fixed state *r* of the *RE*, this process behaves as an irreducible continuous time Markov chain. The intensities of the transitions of the chain $v_t, t \geq 0$, which are not accompanied by arrival, are described by the matrix $D_0^{(r)}$, and the transitions, which are accompanied by arrival of *k*-size batch, are described by the matrix $D_k^{(r)}, \ k \geq 1, \ r = \overline{1, R}$.

Let us denote by $D^{(r)}(z) = \sum_{k=0}^{\infty} D_k^{(r)} z^k, \ |z| \leq 1$, the matrix generating function of the matrices $D_k^{(r)}, \ k \geq 0$. The matrix $D^{(r)}(1)$ for each $r, r = \overline{1, R}$, is the irreducible generator. Under the fixed state *r* of the *RE*, the average intensity $\lambda^{(r)}$ (fundamental rate) of the *BMAP* is defined as

$$\lambda^{(r)} = \mathrm{q}^{(r)} (D^{(r)}(z))'|_{z=1} \mathrm{e}$$

and the intensity $\lambda_b^{(r)}$ of batch arrivals is defined as

$$\lambda_b^{(r)} = \mathrm{q}^{(r)} (-D_0^{(r)}) \mathrm{e}.$$

Here $\mathrm{q}^{(r)}$ is the unique solution to the system $\mathrm{q}^{(r)} D^{(r)}(1) = 0, \ \mathrm{q}^{(r)} \mathrm{e} = 1$. The coefficient of variation $c_{var}^{(r)}$ of intervals between arrival of batches is given by

$$(c_{var}^{(r)})^2 = 2\lambda_b^{(r)} \mathrm{q}^{(r)} (-D_0^{(r)})^{-1} \mathrm{e} - 1,$$

while the correlation coefficient $c_{cor}^{(r)}$ of intervals between successive batch arrivals is calculated as

$$c_{cor}^{(r)} = (\lambda_b^{(r)} \mathrm{q}^{(r)} (-D_0^{(r)})^{-1} (D^{(r)}(1) - D_0^{(r)}) (-D_0^{(r)})^{-1} \mathrm{e} - 1) / (c_{var}^{(r)})^2.$$

Let us introduce the following notations:

- *I* is the identity matrix and *O* is a zero matrix of appropriate dimension;
- $\otimes$ indicates the symbol of Kronecker product of matrices, see Graham (1981);
- $\overline{W} = W + 1$;
- $\widetilde{D}_0 = H \otimes I_{\overline{W}} + \mathrm{diag}\{D_0^{(r)}, \ r = \overline{1, R}\}$;
- $\widetilde{D}_k = \mathrm{diag}\{D_1^{(r)}, \ r = \overline{1, R}\}, \ k > 0$,
- $\mathrm{diag}\{F_r, \ r = \overline{1, R}\}$ is a diagonal matrix with the diagonal entries $F_r, \ r = \overline{1, R}$.

The averaged (over the stationary distribution of the states of the *RE*) intensity λ of input flow of customers is defined as

$$\lambda = \mathrm{q} \sum_{k=1}^{\infty} k \widetilde{D}_k \mathrm{e},$$

and the intensity λ_b of batch arrivals is defined as

$$\lambda_b = \mathrm{q}(-\widetilde{D}_0) \mathrm{e}$$

where the vector **q** is the unique solution to the system

$$\mathrm{q} \sum_{k=0}^{\infty} \widetilde{D}_k = 0, \ \mathrm{qe} = 1.$$

The squared coefficient of variation c_{var} of intervals between successive arrivals is given as

$$c_{var} = 2\lambda \mathrm{q} (-\widetilde{D}_0)^{-1} \mathrm{e} - 1.$$

The coefficient of correlation c_{cor} of two successive intervals between arrivals is given as

$$c_{cor} = (\lambda \mathrm{q} (-\widetilde{D}_0)^{-1} \sum_{k=1}^{\infty} \widetilde{D}_k (-\widetilde{D}_0)^{-1} \mathrm{e} - 1) / c_{var}.$$

We assume that during the epochs of the transitions of the process $r_t, t \geq 0$, the states of the process $\nu_t, t \geq 0$, do not change, only the intensities of the further transitions of this process change.

If there are idle active servers during an arbitrary batch arrival epoch, the customers from the batch occupy the corresponding number of servers. If the number of idle servers is insufficient, then some part of the batch occupies all idle servers while the rest of the batch goes to the buffer. If all servers are busy during an arbitrary batch arrival epoch, all customers from the batch join the buffer.

We suppose that when the transition of the *RE* leads to the reduction of the number of active servers, at first the number of free servers decreases, and if that is not enough, then service of the corresponding number of customers is terminated. We assume that all customers whose service was terminated due to reducing the number of active servers return to the buffer. If the transition of the *RE* leads to the increase of the number of active servers, we suppose that the corresponding number of customers waiting in the buffer, if any, occupies the free additional servers.

The customers in the buffer are assumed to be impatient. Under the fixed state *r* of the *RE*, each waiting customer leaves the buffer due to the lack of service after an exponentially distributed with the parameter $\alpha^{(r)},\ \alpha^{(r)} \geq 0,\ r=\overline{1,R}$ time, independently of other customers.

The probability of the customer service completion in an interval of infinitesimal length $(t,\, t+\Delta t)$ under the fixed state *r* of the *RE* is equal to $\mu^{(r)}\Delta t+o(\Delta t),\ r=\overline{1,R}$.

Our aim is to analyze stationary behavior of the described system.

3. Process of System States and Stationary Distribution

Let $i_t,\, i_t \geq 0$, be the number of customers in the system, $r_t,\, r_t=\overline{1,R}$, be the state of the *RE*, $v_t, v_t=\overline{0,W}$, be the state of the underlying process of the *BMAP* during the epoch $t,\, t\geq 0$.

It is easy to see that the stochastic process $\xi_t=\{i_t,\, r_t,\, v_t\}$, $t\geq 0$, is the regular irreducible continuous time Markov chain.

Let us enumerate the states of the Markov chain ξ_t in the lexicographic order of the components (*i*, *r*, *v*) and let *A* be the generator of this chain.

Lemma 1: *The generator A has the following upper-Hessenbergian structure :*

$$A=\begin{pmatrix} A_{0,0} & A_{0,1} & A_{0,2} & A_{0,3} & \cdots \\ A_{1,0} & A_{1,1} & A_{1,2} & A_{1,3} & \cdots \\ O & A_{2,1} & A_{2,2} & A_{2,3} & \cdots \\ O & O & A_{3,2} & A_{3,3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The non-zero blocks $A_{i,j},\ i,\, j\geq 0$, of generator have the following form :

$$\begin{aligned} &A_{i,i}=(A_{i,i})_{r,r'},\ r,\, r'=\overline{1,R} \\ &(A_{i,i})_{r,r}=(-\mu^{(r)}i+(H)_{r,r})I_{\overline{W}}+D_0^{(r)},\ i=\overline{0,N^{(r)}}, \\ &(A_{i,i})_{r,r}=\left[-\left(\mu^{(r)}N^{(r)}+(i-N^{(r)})\alpha^{(r)}\right)+(H)_{r,r}\right]I_{\overline{W}} \\ &\qquad +D_0^{(r)},\ i>N^{(r)}, \\ &(A_{i,i})_{r,r'}=(H)_{r,r'}I_{\overline{W}},\ r,\, r'=\overline{1,R},\ r\neq r', \\ &A_{i,i+k}=\widetilde{D}_k,\ i\geq 0,\ k\geq 1, \\ &A_{i,i-1}=\mathrm{diag}\left\{\mu^{(r)}i\, I_{\overline{W}},\ r=\overline{1,R}\right\},\ i=\overline{1,N^{(r)}}, \\ &A_{i,i-1}=\mathrm{diag}\left\{\left(\mu^{(r)}N^{(r)}+(i-N^{(r)})\,\alpha^{(r)}\right)I_{\overline{W}},\ r=\overline{1,R}\right\}, \\ &\qquad i>N^{(r)}. \end{aligned}$$

The proof of the lemma is performed by means of careful analysis of possible transitions of the Markov chain during the interval having the infinitesimal length.

Remark 1: It can be verified that the following limits exist

$$\begin{aligned} &Y_0=\lim_{t\to\infty}R_i^{-1}A_{i,i-1},\ Y_1=\lim_{t\to\infty}R_i^{-1}A_{i,i}+I, \\ &Y_k=\lim_{t\to\infty}R_i^{-1}A_{i,i+k-1},\ k>1, \end{aligned}$$

where the matrix R_i is a diagonal matrix with the diagonal entries defined as the moduli of the corresponding diagonal entries of the matrix $A_{i,i},\ i\geq 0$, and the explicit form of matrices $Y_k,\ k\geq 0$, is the following:

$$\begin{aligned} &Y_0=\mathrm{diag}\left\{\widetilde{\Omega}_1,\cdots,\widetilde{\Omega}_R\right\}, \\ &Y_1=\begin{pmatrix} \widetilde{A}_{1,1} & \widetilde{A}_{1,2} & \cdots & \widetilde{A}_{1,R} \\ \widetilde{A}_{2,1} & \widetilde{A}_{2,2} & \cdots & \widetilde{A}_{2,R} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{A}_{R,1} & \widetilde{A}_{R,2} & \vdots & \widetilde{A}_{R,R} \end{pmatrix}, \\ &Y_k=\mathrm{diag}\left\{Z_1^k,\cdots,Z_R^k\right\},\ k>1, \end{aligned}$$

where

$$\begin{aligned} &\widetilde{A}_{r,r'}=\delta_{\alpha^{(r)},0}(K_r(A_{N,N})_{r,r'}+\delta_{r-r',0}I_{\overline{W}}),\ r,\, r'=\overline{1,R}, \\ &Z_r^k=\delta_{\alpha^{(r)},0}K_rD_{k-1}^{(r)},\ r=\overline{1,R},\ k>1, \\ &\widetilde{\Omega}_r=\delta_{\alpha^{(r)},0}N^{(r)}\mu^{(r)}K_r+(1-\delta_{\alpha^{(r)},0})I_{\overline{W}},\ r=\overline{1,R}, \\ &K_r=\left(\mu^{(r)}N^{(r)}I_{\overline{W}}+\Phi_0^{(r)}-(H)_{r,r}\,I_{\overline{W}}\right)^{-1},\ r=\overline{1,R}, \end{aligned}$$

$\delta_{a,b}$ indicates the Kronecker delta, $\Phi_0^{(r)}$ is the diagonal matrix the diagonal entries of which are defined as the corresponding diagonal entries of the matrix $-D_0^{(r)}$.

According to definition of continuous-time asymptotically quasi-Toeplitz Markov chains (*AQTMC*) given in Klimenok and Dudin (2006), existence of the limits $Y_k,\ k>0$, implies that the Markov chain $\xi_t,\ t\geq 0$, belongs to the class of *AQTMC*. So, we can use the results from Klimenok and Dudin (2006) to analyze this chain, in particular, to derive the ergodicity condition for this chain.

As the first step in analysis of the Markov chain $\xi_t,\ t\geq 0$, we have to derive the sufficient condition for the existence of the stationary distribution of this Markov chain. In derivations, we need to separate two cases : the case when $\alpha^{(r)}>0$ at least for one state *r* of the *RE*, $r=\overline{1,R}$, and the case when $\alpha^{(r)}=0$ (customers are absolutely patient) for all $r=\overline{1,R}$. It can be verified that in the former case the Markov chain ξ_t is ergodic for any set of parameters of the queueing system under study.

Let us now consider the latter case $\alpha^{(r)}=0,\ r=\overline{1,R}$. One can see that in this case the blocks of the generator for $i>N^{(R)}$ have the following form :

$$A_1=A_{i,i}=(A_{i,i})_{r,r'},\ r,r'=\overline{1,R},$$
$$(A_{i,i})_{r,r}=\left(-\mu^{(r)}N^{(r)}+(H)_{r,r}\right)I_{\overline{W}}+D_0^{(r)},$$
$$(A_{i,i})_{r,r'}=(H)_{r,r'}I_{\overline{W}},r,r'=\overline{1,R},\ r\neq r',$$
$$A_l=A_{i,i+l-1}=\tilde{D}_{l-1},\ i\geq 0,\ l>1,$$
$$A_0=A_{i,i-1}=\mathrm{diag}\left\{\mu^{(r)}N^{(r)}I_{\overline{W}},\ r=\overline{1,R}\right\},$$

and do not depend on *i*. In this case, as follows from Klimenok and Dudin (2006), the necessary and sufficient condition for the ergodicity of the *QTMC* is the fulfillment of the following inequality

$$\mathbf{x}A_0\mathbf{e}>\mathbf{x}\sum_{l=2}^{\infty}(l-1)A_l\mathbf{e} \tag{1}$$

where the vector **x** is the unique solution to the system

$$\mathbf{x}\sum_{l=0}^{\infty}A_l=\mathbf{0},\ \mathbf{x}\mathbf{e}=1.$$

Theorem 1. If $\alpha^{(r)}=0$ for all states r of the *RE*, $r=\overline{1,R}$, then the Markov chain $\xi_t,\ t\geq 0$, is ergodic if and only if the following inequality holds true :

$$\sum_{r=1}^{R}d_r\mu^{(r)}N^{(r)}>\lambda \tag{2}$$

where $d_r,\ r=\overline{1,R}$, are the components of the vector **d** that defines the stationary distribution of the *RE*.

Proof. Let us consider the matrix $\overline{A}=\sum_{l=0}^{\infty}A_l$ which has the form

$$\overline{A}=H\otimes I_{\overline{W}}+\mathrm{diag}\left\{D_0^{(r)}+\sum_{l=2}^{\infty}D_{l-1}^{(r)},\ r=\overline{1,R}\right\}$$

It is easy to see, that $\overline{A}=\sum_{k=0}^{\infty}\tilde{D}_k$. Hence, the vector **x** coincides with the vector **q**. So, the right side of inequality (1) can be rewritten as

$$\mathbf{x}\sum_{l=2}^{\infty}(l-1)A_le=\mathbf{q}\sum_{l=2}^{\infty}(l-1)A_le=\mathbf{q}\sum_{k=1}^{\infty}k\tilde{D}_ke=\lambda.$$

The left hand side of inequality (1) can be rewritten as

$$\mathbf{q}A_0\mathbf{e}=\mathbf{q}\,\mathrm{diag}\left\{\mu^{(r)}N^{(r)}\otimes I_{\overline{W}},\ r=\overline{1,R}\right\}\mathbf{e}$$
$$=\left(\tilde{\mathbf{q}}_1,\cdots,\tilde{\mathbf{q}}_2\right)\mathrm{diag}\left\{\mu^{(r)}N^{(r)}I_{\overline{W}},\ r=\overline{1,R}\right\}\mathbf{e}$$

where

$$\tilde{\mathbf{q}}_r=\left(\lim_{t\to\infty}P\{r_t=r,\ v_t=0\},\cdots,\lim_{t\to\infty}P\{r_t=r,\ v_t=W\}\right).$$

Using the so called mixed product rule, see Graham (1981), it can be easily verified that

$$\left(\mathrm{diag}\left\{\mu^{(r)}N^{(r)}\otimes I_{\overline{W}},\ r=\overline{1,R}\right\}\right)\mathbf{e}$$
$$=\left(\mu^{(1)}N^{(1)},\cdots\mu^{(R)}N^{(R)}\right)^T\otimes\mathbf{e}_{\overline{W}}.$$

So,

$$\mathbf{q}A_0\mathbf{e}=(\tilde{\mathbf{q}}_1,\cdots,\tilde{\mathbf{q}}_R)\,\mathrm{diag}\left\{\mu^{(r)}N^{(r)}\otimes I_{\overline{W}},\ r=\overline{1,R}\right\}\mathbf{e}$$
$$=\left(\tilde{\mathbf{q}}_1\mu^{(1)}N^{(1)}\mathbf{e}_{\overline{W}}+\cdots+\tilde{\mathbf{q}}_R\mu^{(R)}N^{(R)}\mathbf{e}_{\overline{W}}\right). \tag{3}$$

Taking into account that $\tilde{\mathbf{q}}_r\mathbf{e}=d_r$, the right hand side of equality (3) can be rewritten in the form $\sum_{r=1}^{R}d_r\mu^{(r)}N^{(r)}$. Thus, the theorem is proved.

Remark 2. Inequality (2) is intuitively tractable. The right hand side is the average arrival rate while the left hand side is the average customers departure rate from the overloaded system. It is intuitively clear that the system is stable (the underlying Markov chain is ergodic) if the customers arrival rate is less than the maximal possible service rate.

If the ergodicity condition holds true, the following limits (stationary probabilities) exist :

$$p(i,r,v)=\lim_{t\to\infty}P\{i_t=i,\ r_t=r,\ v_t=v\}$$
$$i\geq 0,\ r=\overline{1,R},\ v=\overline{0,W}.$$

Let us form the row vectors $\mathbf{p}_i$ as follows:

$$\mathbf{p}(i,\ r)=(p(i,\ r,\ 0),\ p(i,\ r,\ 1),\ \cdots,\ p(i,\ r,\ W)),\ r=\overline{1,R},$$
$$\mathbf{p}_i=(\mathbf{p}(i,1),\mathbf{p}(i,2),\cdots,\mathbf{p}(i,R)),\ i\geq 0.$$

It is well known that the probability vectors $\mathbf{p},\ i\geq 0$, satisfy the following system of linear algebraic equations (so called equilibrium or Chapman-Kolmogorov equations) :

$$(\mathbf{p}_0,\mathbf{p}_1,\cdots)A=\mathbf{0},\ (\mathbf{p}_0,\mathbf{p}_1,\cdots)\mathbf{e}=1. \tag{4}$$

System (4) is infinite one and, so, it cannot be directly solved on computer. But the probability vectors $\mathbf{p}_i,\ i\geq 0$, can be computed by means of the numerically stable algorithm that is developed in Klimenok and Dudin (2006) based on the derivation of another system of equations for these vectors with the use of notion of a censored Markov chain.

4. Performance Measures of the System

Having computed the vectors of the stationary probabilities

$\mathbf{p}_i$, $i \geq 0$, it is possible to compute the performance measures of the system.

The stationary distribution of the number of the customers in the system is

$$\lim_{t\to\infty} P\{i_t = i\} = \mathbf{p}_i \mathbf{e},\ i \geq 0.$$

The average number of customers in the system is $L = \sum_{i=1}^{\infty} i\mathbf{p}_i \mathbf{e}$.

The average number of busy servers is

$$N_{server} = \sum_{i=1}^{\infty}\sum_{r=1}^{R} \min\{i,\ N^{(r)}\}\,\mathbf{p}(i,\ r)\mathbf{e}.$$

The average number of customers in the buffer is

$$N_{buffer} = \sum_{i=0}^{\infty}\sum_{r=1}^{R} \max\{0,\ 1 - N^{(r)}\}\,\mathbf{p}(i,\ r)\mathbf{e}$$

The intensity of output flow of customers is

$$\lambda_{out} = \sum_{i=0}^{\infty}\sum_{r=1}^{R} \min\{i,\ N^{(r)}\}\mu^{(r)}\mathbf{p}(i,\ r)\mathbf{e}.$$

The loss probability of an arbitrary customer is

$$P^{(loss)} = 1 - \frac{\lambda_{out}}{\lambda} = \lambda^{-1}\sum_{i=0}^{\infty}\sum_{r=1}^{R} \alpha^{(r)}\max\{0,\ i - N^{(r)}\}\mathbf{p}(i, r)\mathbf{e}.$$

Remark 3. The fact that we have two alternative formulas for computation of the loss probability $P^{(loss)}$ can be helpful at the stage of verification of computer work.

5. Numerical Results

Let us consider the following set of the system parameters. The number of the states of the *RE* is *R* = 2. The generator of the *RE* is given by

$$H = \begin{pmatrix} -0.02 & 0.02 \\ 0.08 & -0.08 \end{pmatrix},$$

so the stationary probability of state 1 is $d_1 = 0.8$ and the stationary probability of state 2 is $d_2 = 0.2$.

To construct the arrival processes, let us fix the matrices

$$B_0 = \begin{pmatrix} -1.35164 & 0 \\ 0 & -0.04387 \end{pmatrix},\ B_1 = \begin{pmatrix} 1.34265 & 0.00899 \\ 0.02443 & 0.01944 \end{pmatrix}.$$

The arrival process is defined by the matrices $D_0^{(1)} = B_0$, $D_k^{(1)} = B_1 q^{k-1}(1-q)/(1-q^K),\ k = \overline{1, K}$, in the first state of the *RE* nd by the matrices $D_0^{(2)} = 3B_0,\ D_k^{(2)} = 3B_1 q^{k-1}(1-q)/(1-q^K),\ k = \overline{1, K}$ in the second state of the *RE*. Here $q = 0.7,\ K = 5$. Arrival processes in both states of the *RE* have the coefficient of variation equal to 12.34 and the coefficient of correlation equal to 0.2. Average intensities are equal to $\lambda^{(1)} = 2.3232$ and $\lambda^{(2)} = 3\lambda^{(1)}$, correspondingly.

The rest of the system parameters under state 1 of the *RE* are as follows :

$$\alpha^{(1)} = 0.05,\ \mu^{(1)} = 0.5.$$

Under state 2 of the *RE*, the parameters are as follows :

$$\alpha^{(2)} = 0.07,\ \mu^{(2)} = 0.7.$$

Let us vary the number of available servers, $N^{(1)}$ and $N^{(2)}$, under the different states of the *RE* in the following way :

$$N^{(2)} \in \{1,\ 50\},\ N^{(1)} \in \{1,\ N^{(2)}\}.$$

<Figure 2>~<Figure 4> illustrate the dependence of some performance measures of the system on the values $N^{(1)}$ and $N^{(2)}$.

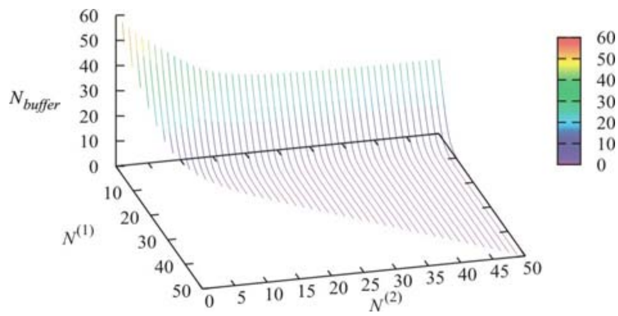

Figure 2. Dependence of the Average Number of Customers in the Buffer on the Numbers of Servers $N^{(1)}$ and $N^{(2)}$

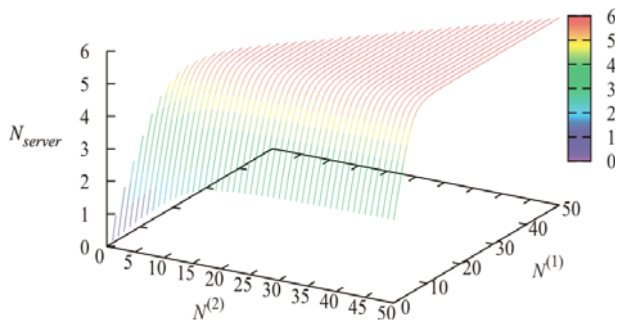

Figure 3. Dependence of the Average Number of Busy Servers on the Numbers of Servers $N^{(1)}$ and $N^{(2)}$

Likely, the most important performance measure of the considered system is the loss probability $P^{(loss)}$ of an arbitrary customer. Because the arrival rate at different states of the *RE* may fluctuate essentially (e.g., in our numerical experiment $\lambda^{(2)} = 3\lambda^{(1)}$) one may think about adjustment of the number of active servers to the current state of the *RE* in such a way as to minimize the weighted over all the states of the *RE* number of active servers under the fixed value ϵ of admissible loss probability $\boldsymbol{P}^{(loss)}$.

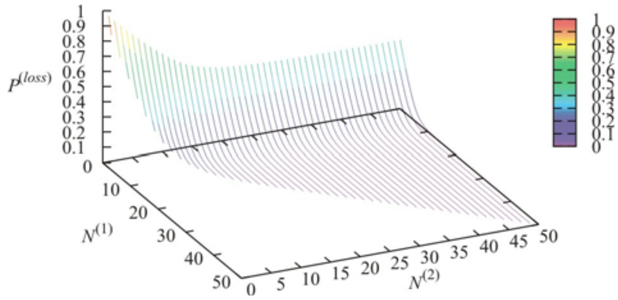

Figure 4. Dependence of the Loss Probability of Customers on the Numbers of Servers $N^{(1)}$ and $N^{(2)}$

Thus, the following optimization problem arises in a natural way:

$$\overline{N} = N^{(1)}d_1 + N^{(2)}d_2 \rightarrow \min$$

subject to restriction

$$\boldsymbol{P}^{(loss)} < \epsilon.$$

In application to real life systems, the necessity of minimization of the value $\overline{N}$ is motivated by consideration of salary payment to operators (in contact center) or energy saving (in cloud computing system).

Let us fix $\epsilon = 0.0001$. Using the presented above results of computation of the values of performance measures of the system, we can compute the optimal values, $N_*^{(1)}$ and $N_*^{(2)}$, of the numbers $N^{(1)}$ and $N^{(2)}$ of active servers at different states of the *RE* as $N_*^{(1)} = 20$ and $N_*^{(2)} = 30$. The optimal value $\overline{N_*}$ of the averaged number $\overline{N}$ of active servers is equal to 22. So, if we would like to provide loss probability in the system be less than $\epsilon = 0.0001$, we have to use, in average, 22 servers, including 20 servers at periods of low load of the system (under state 1 of the *RE*) and 30 servers at periods of high load of the system.

Under the fixed above value 0.7 of the parameter $\boldsymbol{q}$ defining the geometrical distribution of the number of customers in an arbitrary arriving batch, the average batch size is equal to 3.333. The intensities $\lambda^{(1)}$ and $\lambda^{(2)}$ are equal to $\lambda^{(1)} = 2.3232$ and $\lambda^{(2)} = 6.9696$, respectively. The averaged arrival rate, λ, is equal to 3.2525.

Let now change the value of $\boldsymbol{q}$ to 0.6. In this case, the average batch size is equal to 2.5. The intensities $\lambda^{(1)}$ and $\lambda^{(2)}$ are equal to $\lambda^{(1)} = 2.07$ and $\lambda^{(2)} = 6.21$, respectively. The averaged arrival rate, λ, is equal to 2.9. The optimal values of $\overline{N_*}$, $N_*^{(1)}$ and $N_*^{(2)}$ are now equal to 20, 18 and 28, respectively.

6. Conclusion

In this paper, a multi-server queueing system with an infinite buffer operating in the Markovian *RE* is analyzed. The number of active servers, as well as the parameters defining the arrival process, the rate of customers service, and the impatience intensity depend on the state of the *RE*. Using the results for the asymptotically quasi-Toeplitz Markov chains we derived the ergodicity condition in a nice analytical form. The formulas for computation of the main performance measures of the system are presented. Results can be applied for capacity planning and performance evaluation of a variety of real life systems, including computer networks, e.g., for analysis of intellectual transportation communication systems. The number of available servers in sequential areas along the route of a vehicle defines the state of the *RE* and the changes of the *RE* are caused by the motion of a vehicle.

References

Chakravarthy, S. R. (2001), The batch Markovian arrival process : a review and future work Advances, *In : Krishnamoorthy A et al. (Eds.), Probability Theory and Stochastic Processes*, NJ : Notable Publications, 21-49.

Choi, D. I., Kim, B. K., and Lee, D. H. (2014), A Note on the M/G/1/K Queue with Two-Threshold Hysteresis Strategy of Service Intensity Switching, *Journal of the Korean Operations Research and Management Science Society*, **39**(3), 1-5.

Cordeiro, J. D. and Kharoufeh, J. P. (2012), The unreliable M/M/1 retrial queue in a random environment, *Stochastic Models*, **28**, 29-48.

Gnedenko, B. V. and Kovalenko, I. N. (1966), *Introduction to queueing theory*, Science, Moscow.

Graham, A. (1981), *Kronecker products and matrix calculus with applications*, Ellis Horwood, Cichester.

Kim, C. S., Dudin, A. N., Klimenok, V. I., and Khramova, V. V. (2009), Erlang loss queueing system with batch arrivals operating in a random environment, *Computers and Operations Research*, **36**, 674-967.

Kim, C. S., Dudin, A., Dudin, S., and Dudina, O. (2014), Analysis of $MMAP/PH_1, PH_2/N/\infty$ queueing system operating in a random environment, *International Journal Applied Mathematics and Computer Science*, **24**(3), 485-501.

Kim, C. S., Klimenok, V. I., Sang, C. L., and Dudin, A. N. (2007), The $BMAP/PH/1$ retrial queueing system operating in random environment, *Journal of Statistical Planning and Inference*, **137**, 3904-3916.

Kim, C. S. and Lyakhov, A. (2008), Study of Dynamic Polling in the IEEE 802.11 PCF, *Journal of the Korean Institute of Industrials Engineers*, **34**(2), 140-150.

Klimenok, V. I. and Dudin, A. N. (2006), Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory, *Queueing Systems*, **54**, 245-259.

Krieger, U., Klimenok, V. I., Kazimirsky, A. V., Breuer, L., and Dudin, A. N. (2005), A $BMAP/PH/1$ queue with feedback operating in a random environment, *Mathematical and Computer Modelling*, **41**, 867-882.

Lucantoni, D. (1991), New results on the single server queue with a batch Markovian arrival process, *Communications in Statistics-Stochastic Models*, **7**, 1-46.

Neuts, M. F. (1978), The M/M/1 queue with randomly varying arrival and service rates, *Operations Research*, **15**, 139-157.

Neuts, M. F. (1981), *Matrix-geometric solutions in stochastic models*, The Johns Hopkins University Press, Baltimore.

O'Cinneide, C. and Purdue, P. (1986), The M/M/\oo queue in a random environment, *Journal of Applied Probability*, **23**, 175-184.

Purdue, P. (1974), The M/M/1 queue in a Markovian environment, *Operations Research*, **22**, 562-569.

Takine, T. (2005), Single-server queues with Markov-modulated arrivals and service speed, *Queueing Systems*, **49**, 7-22.

Wu, J., Liu, Z., Yang, G. (2011), Analysis of the finite source *MAP/PH/N* retrial G-queue operating in a random environment, *Applied Mathematical Modelling*, **35**, 1184-1193.

Yadin, M. and Syski, R. (1979), Randomization of intensities in a Markov chain, *Advances in Applied Probability*, **11**, 397-421.

Yang, G., Yao, L. G., and Ouyang, Z. S. (2013), The *MAP/PH/N* retrial queue in a random environment, *Acta Mathematicae Applicatae Sinica*, **29**, 725-738.

Yechialy, U. and Naor, P. (1971), Queueing Problems with Heterogeneous Arrivals and Service, *Operations Research*, **19**, 722-734.