Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2015, Article ID 180481, 12 pages
http://dx.doi.org/10.1155/2015/180481

Research Article

Analysis of Multiserver Retrial Queueing System with Varying Capacity and Parameters

Alexander N. Dudin[1] and Olga S. Dudina[2]

[1]*Department of Applied Mathematics and Cybernetics, Tomsk State University, 36 Lenina Avenue, Tomsk, Russia*
[2]*Department of Applied Mathematics and Computer Science, Belarusian State University, 4 Nezavisimosti Avenue, 220030 Minsk, Belarus*

Correspondence should be addressed to Alexander N. Dudin; dudin-alexander@mail.ru

Received 19 April 2015; Revised 29 June 2015; Accepted 5 July 2015

Academic Editor: Francisco Alhama

A multiserver queueing system, the dynamics of which depends on the state of some external continuous-time Markov chain (random environment, RE), is considered. Change of the state of the RE may cause variation of the parameters of the arrival process, the service process, the number of available servers, and the available buffer capacity, as well as the behavior of customers. Evolution of the system states is described by the multidimensional continuous-time Markov chain. The generator of this Markov chain is derived. The ergodicity condition is presented. Expressions for the key performance measures are given. Numerical results illustrating the behavior of the system and showing possibility of formulation and solution of optimization problems are provided. The importance of the account of correlation in the arrival processes is numerically illustrated.

1. Introduction

Queueing theory is widely used for decision making about the resources needed to provide service in a variety of real life systems including contact centers, intelligent transportation systems, telecommunication networks, manufacturing and administrative systems, and banking. Due to the increasing technical and logical complexity of real life systems and diversity of the provided services, queueing models of these systems and their fragments become more and more involved. One of the essential features of modern systems, which should be accounted in queueing models, is that the parameters of the system operation may vary, for example, depending on time of a day or night, available amount of required resources, and possible sharing with some other systems. This fact gave rise to the progress in the study of the so-called queues operating in the random environment (RE).

In contrast to the classical queueing models, where the parameters and the distributions characterizing arrival, service, and other processes describing dynamics of the system are assumed to be fixed in advance, queues operating in the RE presuppose that some or all parameters may dynamically vary due to influence of some external medium called RE. Early research in this topic was done by Gnedenko and Kovalenko [1], Yechiali and Naor [2], Yadin and Syski [3], O'Cinneide and Purdue [4], Purdue [5], Neuts [6, 7], and others. A brief history of the development of theory of queues in the RE, the reference list, and real life examples of queues in the RE can be found, for example, in the papers [8–12]. In [8], 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 [9, 10], the finite source MAP/PH/N retrial queue operating in a random environment is studied. The arrival flow is described by the Markov arrival process (MAP) (for definition, more details, and usefulness in modelling in telecommunications, see [13–15]), and the service time has a phase-type (PH) distribution (for definition and more details, see [7]). It is assumed that the parameters defining the MAP process and PH distribution depend on the state

of the RE that is a continuous-time Markov chain with a finite state space. In [9], it is assumed that there is additional MAP arrival process of negative customers. The arrival of the negative customer with equal probability goes to any busy server to remove the customer being in service. In [9, 10], the finite state multidimensional 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 [11], 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 [12], the BMAP/PH/N retrial queue operating in the RE is investigated. As in the previously described model, the system does not have a buffer. But the customer who did not succeed to find a free server upon arrival is not lost, but he proceeds to the retrial orbit, which is a location from which customers may attempt to gain or regain service. The intensities of repeating the attempts are also assumed to be dependent on the state of the RE. In [12], the sufficient condition of stability of the system is proved, the stationary distribution of the system states is computed. The presented numerical examples illustrate a poor quality of the approximation of the main performance measures of the system by means of the simpler queueing models. An effect of possible smoothing the traffic and an impact of the retrial intensity are shown. Analysis presented in [12] is more complicated than the analyses in [9, 10] due to two evident reasons: (i) the state space of the multidimensional Markov chain describing the behavior of the system is finite in [9, 10] and is infinite in [12]; (ii) the generator of the multidimensional Markov chain has tridiagonal structure in [9, 10] (i.e., the chain is level dependent quasi-birth-and-death process) while it has more general upper-Hessenbergian structure in [12].

The main advantages of the model considered in this paper, compared to the ones analyzed in [11, 12] and to all the other papers devoted to multiserver queues operating in the RE, are as follows.

(i) The model under study has more flexible service discipline that combines the features of loss systems, systems with a finite buffer and systems with retrials. The queueing systems that incorporate both normal queues and retrial orbits are called hybrid retrial queues. The importance of their investigation stems from the fact that many modern technologies of customers random access assume the existence of some places where the customers, who did not succeed to get access upon arrival, may be temporarily kept (in registers for handover customers in cells of mobile communication networks, in IVR (Interactive Voice Response) machines in call centers, etc.). Such hybrid systems were considered, for example, in [16, 17]. Queueing systems in RE with customers loss considered in [11] and with retrials considered in [12] are obtained as special cases of the hybrid retrial queue in RE considered in this paper.

(ii) We suppose that the capacity of the service area of the system (the number of servers and places in the buffer) may depend on the state of the RE, while it is usually assumed that the capacity is constant and the RE influences only the parameters of the arrival and service processes. Note that the changes in service capacity may be considered as server breakdowns, among other possibilities.

(iii) Customers balking, impatience (abandonment), and nonpersistence are taken into account. This allows us to account psychology of customers, use of visible queue option, information obsolescence during the waiting time, customer's mobility, service provider's competition, and so forth.

Due to the generality of the considered queueing model, it has a lot of possible applications for the investigation of a variety of real life systems. Besides the examples of application to the performance evaluation, capacity planning, and optimization of some systems, which coincide with the ones described in [12] (application to modelling hot spot in airport and wireless local area network), it is worth to mention important potential applications for work force management in call-centers. Fluctuation of the intensities of customer arrivals and retrials occurs here, for example, due to the well known existence of hours of low, middle, and peak load of the center. The servers correspond to the operators of the call-center and places in the buffer correspond to IVR machines. Impatience of customers staying in the buffer reflects the possibility to get the requested information directly from IVR without contacting to an operator. It is necessary to create the schedule of the work of the operators and IVRs to dynamically fit the number of the active operators and IVRs to the current level of the load of the call-center in such a way as to guarantee the best quality of the customer's service under the existing restrictions imposed on the total number of available operators and their working schedule and on energy consumption by IVRs.

The rest of the paper is organized as follows. In Section 2, the mathematical model is described. The process of the system states is defined; its generator in a block matrix form is presented. In Section 3, the stability condition is derived and the problem of computation of the stationary probabilities of the system states is touched on. Formulas for computation of the steady state performance measures of the system are presented in Section 4. Numerical illustrations are given and briefly discussed in Section 5. Section 6 concludes the paper.

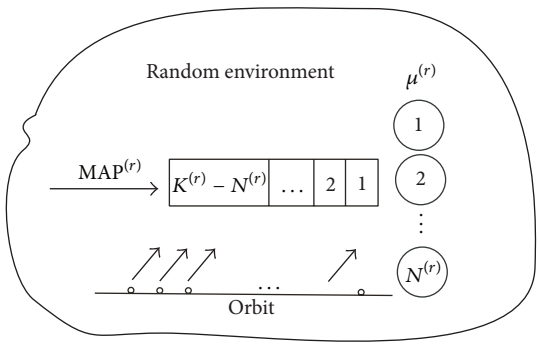

Figure 1: Queueing system under study.

2. Mathematical Model

We consider a retrial multiserver queueing system with varying capacity and behavior of the customers. The structure of the system under study is presented in Figure 1.

The dynamics of the system depends on the state of the RE. The RE is defined by the stochastic process r_t, $t \geq 0$, which is an irreducible continuous-time Markov chain with the state space $\{1, 2, \ldots, R\}$ and the infinitesimal generator H.

Under the fixed state r of the RE, capacity of the system is equal to $K^{(r)}$ including $N^{(r)}$ servers, $0 \leq N^{(r)} \leq K^{(r)}$, and a waiting room (buffer) of size $K^{(r)} - N^{(r)}$, $r \in \{1, \ldots, R\}$. Without the loss of generality, we assume that the states of the RE are enumerated in ascending order of the capacity of the system; that is, $0 \leq K^{(1)} \leq K^{(2)} \leq \cdots \leq K^{(R)}$. We also suggest that $0 \leq N^{(1)} \leq N^{(2)} \leq \cdots \leq N^{(R)}$. We call the servers and the buffer as the service area of the system.

Arrival of customers is defined by the Markovian arrival process (MAP). The underlying process of this MAP is $\{r_t, \nu_t\}$, $t \geq 0$, where r_t is the state of the RE and the process ν_t has a finite state space $\{0, 1, \ldots, W\}$. Under the fixed state r of the RE the process ν_t behaves as irreducible continuous-time Markov chain. The sojourn time of this chain in the state ν is exponentially distributed with the positive finite parameter $\lambda_\nu^{(r)}$. When the sojourn time in the state ν expires, with probability $p_0^{(r)}(\nu, \nu')$, the process ν_t jumps to the state ν' without generation of a customer, $\nu, \nu' \in \{0, \ldots, W\}$, $\nu \neq \nu'$, $r \in \{1, \ldots, R\}$. With probability $p_1^{(r)}(\nu, \nu')$, the process ν_t jumps to the state ν' with generation of a customer, $\nu, \nu' \in \{0, \ldots, W\}$, $r \in \{1, \ldots, R\}$.

The behavior of the arrival process under the fixed state r of the RE is completely characterized by the matrices $D_0^{(r)}$ and $D_1^{(r)}$ defined by the entries

$$\begin{aligned}
&\left(D_0^{(r)}\right)_{\nu,\nu} = -\lambda_\nu^{(r)}, \quad \nu \in \{0, \ldots, W\},\\
&\left(D_0^{(r)}\right)_{\nu,\nu'} = \lambda_\nu^{(r)} p_0^{(r)}\left(\nu, \nu'\right),\\
&\qquad\qquad \nu, \nu' \in \{0, \ldots, W\}, \ \nu \neq \nu', \quad (1)\\
&\left(D_1^{(r)}\right)_{\nu,\nu'} = \lambda_\nu^{(r)} p_1^{(r)}\left(\nu, \nu'\right),\\
&\qquad\qquad \nu, \nu' \in \{0, \ldots, W\}, \ r \in \{1, \ldots, R\}.
\end{aligned}$$

The square matrix $D^{(r)}(1) = D_0^{(r)} + D_1^{(r)}$ of dimension $\overline{W} = W + 1$ represents the generator of the process ν_t, $t \geq 0$, under the fixed state r, $r \in \{1, \ldots, R\}$.

The average arrival rate $\lambda^{(r)}$ under the fixed state r of the RE is given as

$$\lambda^{(r)} = \boldsymbol{\theta}^{(r)} D_1^{(r)} \mathbf{e}, \quad (2)$$

where $\boldsymbol{\theta}^{(r)}$ is the invariant vector of the stationary distribution of the Markov chain ν_t, $t \geq 0$, under the fixed state r. The vector $\boldsymbol{\theta}^{(r)}$ is the unique solution to the system

$$\begin{aligned}
\boldsymbol{\theta}^{(r)} D^{(r)}(1) &= \mathbf{0},\\
\boldsymbol{\theta}^{(r)} \mathbf{e} &= 1.
\end{aligned} \quad (3)$$

The squared coefficient of variation $c_{\text{var}}^{(r)}$ of intervals between successive arrivals under the fixed state r of the RE is given as

$$c_{\text{var}}^{(r)} = 2\lambda^{(r)} \boldsymbol{\theta}^{(r)} \left(-D_0^{(r)}\right)^{-1} \mathbf{e} - 1, \quad r \in \{1, \ldots, R\}. \quad (4)$$

The coefficient of correlation $c_{\text{cor}}^{(r)}$ of two successive intervals between arrivals under the fixed state r of the RE is given as

$$c_{\text{cor}}^{(r)} = \frac{\left(\lambda^{(r)} \boldsymbol{\theta}^{(r)} \left(-D_0^{(r)}\right)^{-1} \left(D^{(r)}(1) - D_0^{(r)}\right) \left(-D_0^{(r)}\right)^{-1} \mathbf{e} - 1\right)}{c_{\text{var}}^{(r)}}, \quad r \in \{1, \ldots, R\}. \quad (5)$$

Let us introduce the following matrices:

$$\begin{aligned}
\widetilde{D}_1 &= \operatorname{diag}\left\{D_1^{(r)}, \ r \in \{1, \ldots, R\}\right\},\\
\widetilde{D}_0 &= H \otimes I_{\overline{W}} + \operatorname{diag}\left\{D_0^{(r)}, \ r \in \{1, \ldots, R\}\right\},\\
\widetilde{D}(1) &= \widetilde{D}_0 + \widetilde{D}_1.
\end{aligned} \quad (6)$$

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

$$\lambda = \boldsymbol{\theta} \widetilde{D}_1 \mathbf{e}, \quad (7)$$

where the vector $\boldsymbol{\theta}$ is the unique solution of the system

$$\begin{aligned}
\boldsymbol{\theta} \widetilde{D}(1) &= \mathbf{0},\\
\boldsymbol{\theta} \mathbf{e} &= 1.
\end{aligned} \quad (8)$$

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

$$c_{\text{var}} = 2\lambda \boldsymbol{\theta} \left(-\widetilde{D}_0\right)^{-1} \mathbf{e} - 1. \quad (9)$$

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

$$c_{\text{cor}} = \frac{\left(\lambda \boldsymbol{\theta} \left(-\widetilde{D}_0\right)^{-1} \widetilde{D}_1 \left(-\widetilde{D}_0\right)^{-1} \mathbf{e} - 1\right)}{c_{\text{var}}}. \quad (10)$$

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

If, at an arbitrary arrival moment, some server is free, an arriving customer starts the service. If all servers are busy, but the buffer is not full, the customer is placed to the buffer. If the buffer is full and the state of the RE is r, the customer balks (leaves the system permanently) with probability $1 - p_1^{(r)}$ or, with the complementary probability, moves to orbit and tries again later on.

If the RE transits from the state r to the state r', where $r' < r$, the capacity of the system (the number of available servers and/or places in the buffer) decreases. If the current number of customers in service and/or in the buffer exceeds the available number of servers and/or places in the buffer under the state r' of the RE, the redundant customers leave the service area. The customers leaving the service area go, independently of each other, into orbit with probability $p_2^{(r)}$ or leave the system permanently with the complementary probability.

When the state of the RE is r, $r \in \{1, \ldots, R\}$, each customer staying in orbit repeats the attempts to reach the service area after an exponentially distributed time described by the parameter $\alpha^{(r)}$, $\alpha^{(r)} \geq 0$. Customers retry independent of each other and are not absolutely persistent. If the attempt is successful, the customer leaves orbit and moves to service or to the buffer. Otherwise, the customer returns to orbit with probability $p_3^{(r)}$. With the complementary probability, the customer leaves the system permanently.

Customers in orbit are impatient. When the state of the RE is r, $r \in \{1, \ldots, R\}$, each customer may leave the orbit and the system after an exponentially distributed time described by the parameter $\gamma^{(r)}$, $\gamma^{(r)} > 0$. If the customers staying in the orbit are patient, we put $\gamma^{(r)} = 0$.

We assume that the probability of the service completion of a customer during the time interval $(t, t + \Delta t)$ is equal to $\mu^{(r)}\Delta t + o(t)$ when the state of the RE is r, $r \in \{1, \ldots, R\}$.

If the RE transits from the state r to the state r', where $r' > r$, the capacity of the system increases and the corresponding number of customers from the buffer, if any, occupies additional $N^{(r')} - N^{(r)}$ servers.

Mention that in the described queueing model the discipline of customers admission may be different under the various states of the RE. For example, if $p_1^{(r)} = 1$ and $\alpha^{(r)} = 0$, no customers are admitted to the system if the buffer is full; that is, the system operates as the system with a finite buffer. If, additionally, $K^{(r)} = N^{(r)}$, the system operates as the system with customers loss (Erlang loss model). If $p_1^{(r)} = 0$, the system operates as the usual retrial system where all arriving customers, who meet the service area full, go into orbit. In the presented analysis, it is accounted that $N^{(r)}$ can be equal to 0, that is, for some states of the RE service to customers are not provided at all.

In order to improve the readability of the paper, we collect and already introduced some new notation in Notation.

Consider the following:

(i) let i_t, $i_t \geq 0$, be the number of customers in orbit,

(ii) let r_t, $r_t \in \{1, \ldots, R\}$, be the state of the RE,

(iii) let n_t, $n_t \in \{0, \ldots, K^{(r)}\}$, be the number of customers in the service area,

(iv) let ν_t, $\nu_t \in \{0, \ldots, W\}$, be the state of the second component of the underlying process of customers arrivals

at the moment t, $t \geq 0$.

It is easy to see that the process $\xi_t = \{i_t, r_t, n_t, \nu_t\}$, $t \geq 0$, is the four-dimensional irreducible Markov chain. Let us enumerate the states of the Markov chain ξ_t, $t \geq 0$, in the direct lexicographic order of the components (i, r, n, ν). We call the set of the states having value (i, r) of the two first components of the Markov chain the macrostate (i, r).

Let Q be the generator of the Markov chain ξ_t, $t \geq 0$. It is formed by the blocks $Q_{i,j}$, consisting of the matrices $(Q_{i,j})_{r,r'}$ that define the intensities of the transitions of the Markov chain ξ_t, $t \geq 0$, from the macrostate (i, r) to the macrostate (j, r'), $r, r' \in \{1, \ldots, R\}$. The diagonal entries of the matrix $Q_{i,i}$ are negative. The modulus of each entry defines the intensity of departing from the corresponding state of the Markov chain ξ_t, $t \geq 0$.

Lemma 1. *The generator Q has the following block structure:*

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & Q_{0,2} & \cdots & Q_{0,\overline{K}} & O & O & \cdots \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & \cdots & Q_{1,\overline{K}} & Q_{1,\overline{K}+1} & O & \cdots \\ O & Q_{2,1} & Q_{2,2} & \cdots & Q_{2,\overline{K}} & Q_{2,\overline{K}+1} & Q_{2,\overline{K}+2} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{11}$$

where $\overline{K} = \max\{\max\{K^{(R)} - K^{(1)}, N^{(R)} - N^{(1)}\}, 1\}$.

The nonzero blocks $Q_{i,j}$, $i, j \geq 0$, are defined as follows. Consider

(i)

$$Q_{i,i} = \left(Q_{i,i}\right)_{r,r'}, \quad r, r' \in \{1, \ldots, R\}, \tag{12}$$

where

$$\begin{aligned} \left(Q_{i,i}\right)_{r,r} = &- \left(\mu^{(r)} C_r + i\left(\alpha^{(r)} + \gamma^{(r)}\right) I_{K^{(r)}+1}\right. \\ &+ \mu^{(r)} C_r E^{-}_{K^{(r)}} + i p_3^{(r)} \alpha^{(r)} \widehat{I}_{K^{(r)}} \Big) \otimes I_{\overline{W}} \\ &+ \left(\left(1 - p_1^{(r)}\right) \widehat{I}_{K^{(r)}} + E^{+}_{K^{(r)}}\right) \otimes D_1^{(r)} + I_{K^{(r)}+1} \otimes D_0^{(r)} \\ &+ (H)_{r,r}\, I_{(K^{(r)}+1)\overline{W}}, \end{aligned} \tag{13}$$

$$\left(Q_{i,i}\right)_{r,r'} = (H)_{r,r'}\, M^{(0)}_{r,r'} \otimes I_{\overline{W}}, \quad r' < r, \tag{14}$$

$$\left(Q_{i,i}\right)_{r,r'} = (H)_{r,r'}\, M^{+}_{r,r'} \otimes I_{\overline{W}}, \quad r' > r,\ i \geq 0. \tag{15}$$

Hereinafter, the entry $(M_{r,r'}^{(n)})_{k,k'}$ *of the matrix* $M_{r,r'}^{(n)}$, $r \in \{1,\dots,R\}$, $r' \in \{1,\dots,r-1\}$, *of dimension* $(K^{(r)}+1)\times(K^{(r')}+1)$ *defines the probability that when the state of the RE changes from* r *to* r' *(and the capacity of the service area decreases from* $K^{(r)}$ *to* $K^{(r')}$*) the number of customers in the service area changes from* k *to* k' *and* n *customers move to orbit. The matrices* $M_{r,r'}^{(n)}$ *have the following nonzero entries. Consider*

$$\begin{aligned}
&\left(M_{r,r'}^{(0)}\right)_{k,k} = 1, \quad k \in \left\{0,\dots,N^{(r')}\right\},\\
&\left(M_{r,r'}^{(n)}\right)_{k,N^{(r')}} = p^{(r)}\left(n, k-N^{(r')}\right),\\
&\qquad k \in \left\{N^{(r')}+1,\dots,N^{(r)}\right\}.\\
&\left(M_{r,r'}^{(n)}\right)_{k,N^{(r')}+k-N^{(r)}} = p^{(r)}\left(n, N^{(R)}-N^{(r')}\right),\\
&k \in \left\{N^{(r)}+1,\dots,\min\left\{K^{(r')}+N^{(r)}-N^{(r')},K^{(r)}\right\}\right\},\\
&\left(M_{r,r'}^{(n)}\right)_{k,K^{(r')}} = p^{(r)}\left(n, k-K^{(r')}\right),\\
&k \in \left\{\min\left\{K^{(r')}+N^{(r)}-N^{(r')},K^{(r)}\right\}+1,\dots,K^{(r)}\right\}.,
\end{aligned} \tag{16}$$

The probability $p^{(r)}(n,k)$ *that* n *customers go into orbit in the case when* k *customers leave the service area is given by*

$$p^{(r)}(n,k) = \begin{cases} C_k^n\left(1-p_2^{(r)}\right)^n\left(p_2^{(r)}\right)^{k-n}, & n \le k,\\ 0, & n > k. \end{cases} \tag{17}$$

The matrix $M_{r,r'}^{+}$, $r \in \{1,\dots,R-1\}$, $r' \in \{r+1,\dots,R\}$, *of dimension* $(K^{(r)}+1)\times(K^{(r')}+1)$, *which defines the transition probabilities of the component* n_t *when the state of the RE changes from* r *to* r' *(and the capacity of the service area increases from* $K^{(r)}$ *to* $K^{(r')}$*), is equal to the matrix* $(I_{K^{(r)}+1} \mid O)$. *Consider*

(ii)

$$Q_{i,i+k} = \begin{pmatrix} Z_{1,1}^{(k)} & O & O & \dots & O & O\\ Z_{2,1}^{(k)} & Z_{2,2}^{(k)} & O & \dots & O & O\\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots\\ Z_{R-1,1}^{(k)} & Z_{R-1,2}^{(k)} & Z_{R-1,3}^{(k)} & \ddots & Z_{R-1,R-1}^{(k)} & O\\ Z_{R,1}^{(k)} & Z_{R,2}^{(k)} & Z_{R,3}^{(k)} & \dots & Z_{R,R-1}^{(k)} & Z_{R,R}^{(k)} \end{pmatrix}, \tag{18}$$
$$i \ge 0,\ k \in \left\{1,\dots,\overline{K}\right\},$$

where the entries of the matrices $Z_{r,r'}^{(k)}$ *define the intensities of the transitions of components* $\{r_t, n_t, \nu_t\}$ *of the Markov chain* ξ_t *leading to the arrival of* k *customers into the orbit. These matrices are defined by the following formulas:*

$$\begin{aligned}
&Z_{r,r}^{(1)} = p_1^{(r)}\widehat{I}_{K^{(r)}} \otimes D_1^{(r)},\\
&Z_{r,r}^{(k)} = O, \quad k > 1,\\
&Z_{r,r'}^{(k)} = (H)_{r,r'}\, M_{r,r'}^{(k)} \otimes I_{\overline{W}},\\
&\qquad k \ge 1,\ r' < r,\ r, r' \in \{1,\dots,R\}.
\end{aligned} \tag{19}$$

Consider

(iii)

$$\begin{aligned}
Q_{i,i-1} = \operatorname{diag}\Big\{ & i\left(\alpha^{(r)}\left[E_{K^{(r)}}^{+} + \left(1-p_3^{(r)}\right)\widehat{I}_{K^{(r)}}\right] + \gamma^{(r)}\right.\\
&\left.\otimes I_{K^{(r)}+1}\right) \otimes I_{\overline{W}},\ r \in \{1,\dots,R\}\Big\}, \quad i \ge 1.
\end{aligned} \tag{20}$$

Proof of the lemma is performed by means of the analysis of the intensities of all possible transitions of the Markov chain ξ_t during the time interval having infinitesimal length. The generator Q is block upper-Hessenbergian matrix; that is, all the blocks below the first subdiagonal are equal to zero blocks. This stems from the fact that, at any moment, no more than one customer may leave the orbit, so the blocks $Q_{i,j}$ are equal to zero when $j < i-1$. The number $\overline{K}$ defines the maximal number of customers that can simultaneously move to the orbit at an arbitrary moment. We recall that one customer may arrive to the orbit if the buffer is full at the customer's arrival moment and at most $\max\{K^{(R)} - K^{(1)}, N^{(R)} - N^{(1)}\}$ customers may arrive to the orbit if the RE jumps from state R to state 1 when all servers are busy and the buffer is full. So, the blocks $Q_{i,j}$ are equal to zero when $j > \overline{K}$. This explains structure (11) of generator Q.

The negative diagonal entries of the matrix $(Q_{i,i})_{r,r}$ define, up to the sign, intensities of the exit of the Markov chain ξ_t from the macrostate (i,r). This exit can happen due to a customer service completion, attempt of a customer from the orbit to reach the service area, departure of a customer from the orbit due to impatience, change of the state of the RE, and change of the state of the underlying process of MAP arrivals. The nondiagonal entries of the matrix $(Q_{i,i})_{r,r}$ define intensities of transition of the Markov chain ξ_t inside of the macrostate (i,r). These transitions can happen due to the customer service completion, change of the state of the RE, and change of the state of the underlying process of MAP arrivals. This explains form (13) of the block $(Q_{i,i})_{r,r}$.

Transition from the macrostate (i,r) to the macrostate (i,r'), $r' < r$, can happen when the state of the RE changes from r to r', $r' < r$. Such a change of the state of the RE implies the possible reduction of the number of available servers and places in the buffer. This may cause departure of n customers from the service to the orbit. Matrices $M_{r,r'}^{(n)}$, $r \in \{1,\dots,R\}$, $r' \in \{1,\dots,r-1\}$, define the probability that when the state of the RE changes from r to r' (and the capacity of the service area decreases from $K^{(r)}$ to $K^{(r')}$) the number of customers in the service area changes from k to k' and

n customers move to the orbit. This explains formula (14). Careful analysis of probabilities, which form the matrices $M_{r,r'}^{(n)}$, $r \in \{1,\dots,R\}$, $r' \in \{1,\dots,r-1\}$, leads to formulas (16).

Transition from the macrostate (i,r) to the macrostate (i,r'), $r' > r$, can happen when the state of the RE changes from r to r', $r' > r$. Such a change of the state of the RE implies the possible increase of the number of available servers and places in the buffer. This explains formula (15).

The increase of the value of the first components of the Markov chain ξ_t from i to $i+k$, $k \in \{1,\dots,\overline{K}\}$, may happen if one customer arrives to the orbit from outside (for $k=1$) or k customers leave the service area due to the reduction of the available space there. This explains formulas (18)-(19).

Finally, the decrease of the value of the first components of the Markov chain ξ_t from i to $i-1$ may happen if a customer makes an attempt from the orbit when the service area is not full or a customer leaves the orbit due to impatience. This explains formula (20). Lemma is proved.

Remark 2. It is easy to verify that the following limits exist:

$$\begin{aligned} Y_0 &= \lim_{i\to\infty} R_i^{-1} Q_{i,i-1}, \\ Y_1 &= \lim_{i\to\infty} R_i^{-1} Q_{i,i} + I, \\ Y_k &= \lim_{i\to\infty} R_i^{-1} Q_{i,i+k-1}, \\ & \qquad k \in \left\{2,\dots,\overline{K}+1\right\}, \end{aligned} \tag{21}$$

where R_i is a diagonal matrix with the diagonal entries which are defined as the moduli of the corresponding diagonal entries of the matrix $Q_{i,i}$, $i \geq 0$. The matrices Y_k, $k \in \{0,\dots,\overline{K}+1\}$, have the following form:

$$Y_0 = \operatorname{diag}\left\{\widetilde{\Omega}_1,\dots,\widetilde{\Omega}_R\right\},$$

$$Y_1 = \begin{pmatrix} \widetilde{Q}_{1,1} & \widetilde{Q}_{1,2} & \dots & \widetilde{Q}_{1,R} \\ \widetilde{Q}_{2,1} & \widetilde{Q}_{2,2} & \dots & \widetilde{Q}_{1,R} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{Q}_{R,1} & \widetilde{Q}_{R,2} & \dots & \widetilde{Q}_{R,R} \end{pmatrix},$$

$$Y_k = \begin{pmatrix} \widetilde{Z}_{1,1}^{(k-1)} & O & O & \dots & O & O \\ \widetilde{Z}_{2,1}^{(k-1)} & \widetilde{Z}_{2,2}^{(k-1)} & O & \dots & O & O \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \widetilde{Z}_{R-1,1}^{(k-1)} & \widetilde{Z}_{R-1,2}^{(k-1)} & \widetilde{Z}_{R-1,3}^{(k-1)} & \ddots & \widetilde{Z}_{R-1,R-1}^{(k-1)} & O \\ \widetilde{Z}_{R,1}^{(k-1)} & \widetilde{Z}_{R,2}^{(k-1)} & \widetilde{Z}_{R,3}^{(k-1)} & \dots & \widetilde{Z}_{R,R-1}^{(k-1)} & \widetilde{Z}_{R,R}^{(k-1)} \end{pmatrix}, \quad k > 1, \tag{22}$$

where

$$\widetilde{\Omega}_r = E_{K^{(r)}}^{+} \otimes I_{\overline{W}}, \quad \text{if } p_3^{(r)} = 1,\ \alpha^{(r)} > 0,\ \gamma^{(r)} = 0,$$

$$\widetilde{\Omega}_r = \left(\frac{\gamma^{(r)}}{\gamma^{(r)}+\alpha^{(r)}} \left(I_{K^{(r)}+1} - \widehat{I}_{K^{(r)}} \right) + \frac{\alpha^{(r)}}{\gamma^{(r)}+\alpha^{(r)}} E_{K^{(r)}}^{+} + \widehat{I}_{K^{(r)}} \right) \otimes I_{\overline{W}} \quad \text{if } \left(p_3^{(r)} \neq 1,\ \alpha^{(r)} > 0 \right) \text{ or } \gamma^{(r)} \neq 0,$$

$$\widetilde{\Omega}_r = O, \quad \text{if } \alpha^{(r)} = 0,\ \gamma^{(r)} = 0,\ r \in \{1,\dots,R\}, \tag{23}$$

$$\widetilde{Q}_{r,r'} = \begin{cases} R_1^{(r)} \left(Q_{0,0}\right)_{r,r'} + \delta_{r-r',0} \widehat{I}_{K^{(r)}} \otimes I_{\overline{W}}, & \text{if } p_3^{(r)} = 1,\ \alpha^{(r)} > 0,\ \gamma^{(r)} = 0, \\ R_2^{(r)} \left(Q_{0,0}\right)_{r,r'} + \delta_{r-r',0} I_{K^{(r)}} \otimes I_{\overline{W}}, & \text{if } \alpha^{(r)} = 0,\ \gamma^{(r)} = 0, \\ O, & \text{if } \left(p_3^{(r)} \neq 1,\ \alpha^{(r)} > 0 \right) \text{ or } \gamma^{(r)} \neq 0, \end{cases} \quad r, r' \in \{1,\dots,R\},$$

$\delta_{i,j}$ indicates the Kronecker delta,

$$\begin{aligned} R_1^{(r)} &= \widehat{I}_{K^{(r)}} \otimes \left(\left(\mu^{(r)} N^{(r)} - (H)_{r,r} \right) I_{\overline{W}} + \Sigma_0^{(r)} \right. \\ &\quad \left. - \left(1 - p_1^{(r)}\right) \Sigma_1^{(r)} \right)^{-1}, \quad r \in \{1,\dots,R\}, \\ R_2^{(r)} &= \left(\mu^{(r)} C_r I_{\overline{W}} + I_{K^{(r)}} \otimes \left(\Sigma_0^{(r)} - (H)_{r,r} I_{\overline{W}} \right) - \widehat{I}_{K^{(r)}} \right. \\ &\quad \left. \otimes \left(1 - p_1^{(r)}\right) \Sigma_1^{(r)} \right)^{-1}, \quad r \in \{1,\dots,R\}, \end{aligned}$$

$$\widetilde{Z}_{r,r'}^{(k)} = \begin{cases} R_1^{(r)} Z_{r,r'}^{(k)}, & \text{if } p_3^{(r)} = 1,\ \alpha^{(r)} > 0,\ \gamma^{(r)} = 0, \\ R_2^{(r)} Z_{r,r'}^{(k)}, & \text{if } \alpha^{(r)} = 0,\ \gamma^{(r)} = 0, \\ O, & \text{if } \left(p_3^{(r)} \neq 1,\ \alpha^{(r)} > 0 \right) \text{ or } \gamma^{(r)} \neq 0, \end{cases} \quad r, r' \in \{1,\dots,R\}. \tag{24}$$

Here, $\Sigma_0^{(r)}$ and $\Sigma_1^{(r)}$ are the diagonal matrices, the diagonal entries of which are defined as the corresponding diagonal entries of the matrices $-D_0^{(r)}$ and $D_1^{(r)}$, respectively.

According to the definition given in [18], the existence of the limits Y_k, $k \in \{0, \dots, \overline{K}+1\}$, means that the Markov chain ξ_t, $t \geq 0$, belongs to the class of continuous-time asymptotically quasi-Toeplitz Markov chains (AQTMC). This fact allows us to use the results of [18] for derivation of the sufficient condition of the ergodicity of the Markov chain ξ_t and computation of its steady state distribution.

3. System Stability and Stationary Distribution

As follows from [18], the sufficient condition for the ergodicity of AQTMC is the fulfillment of the inequality

$$\mathbf{y} Y_0 \mathbf{e} > \mathbf{y} \sum_{k=2}^{\overline{K}+1} (k-1) Y_k \mathbf{e}, \tag{25}$$

where the vector $\mathbf{y}$ is the unique solution to the system

$$\begin{aligned} \mathbf{y} \sum_{k=0}^{\overline{K}+1} Y_k &= \mathbf{y}, \\ \mathbf{y}\mathbf{e} &= 1. \end{aligned} \tag{26}$$

Thus, to check whether or not the Markov chain ξ_t is ergodic, it is necessary to substitute the matrices Y_k, $k \in \{0, \dots, \overline{K}+1\}$, to system (26), solve this system, and verify the fulfillment of inequality (25). If this inequality is fulfilled, the Markov chain under study is ergodic. The solution of a finite system of (26) on computer does not meet any essential problems.

Mention that there exists an important particular case (the customers are impatient or nonpersistent at least for one state of the RE) when ergodicity of the Markov chain ξ_t can be established more easy than via solution of system (26) and verification of inequality (25). For this case, the following statement is true.

Theorem 3. *If customers are impatient or nonpersistent ($\gamma^{(r)} \neq 0$ or $p_3^{(r)} \neq 1$) at least for one state r of the RE, then the Markov chain ξ_t is ergodic for any set of parameters of the queueing system under study.*

Proof. Let $L = \{l_1, l_2, \dots, l_S\}$ be the set of the states of the RE for which at least one of inequalities $p_3^{(r)} \neq 1$ or $\gamma^{(r)} \neq 0$, is fulfilled, $r \in L$. It can be verified that, in this case, by means of coordinated perturbations of the block rows and block columns, the matrix $Y = \sum_{k=0}^{\overline{K}+1} Y_k$ can be transformed to the canonical normal form

$$Y = \begin{pmatrix} Y_{1,1} & Y_{1,2} \\ O & Y_{2,2} \end{pmatrix}. \tag{27}$$

Here, $Y_{1,1}$ is the matrix obtained from the matrix Y by means of removing the block rows and block columns with numbers l, $l \in L$, $Y_{1,2}$ is the matrix obtained from the matrix Y by means of removing the block rows with numbers l, $l \in L$, and the block columns with numbers r, $r \in \{1, \dots, R\} \setminus L$, and $Y_{2,2} = \operatorname{diag}\{\widetilde{\Omega}_l, l \in L\}$. This means that the matrix Y is reducible. So, it follows, from Theorem 2 in [18], that the sufficient condition for the ergodicity of the Markov chain ξ_t can be rewritten in the form

$$\mathbf{z} \overline{Y}_0 \mathbf{e} > \mathbf{z} \sum_{k=2}^{\overline{K}+1} (k-1) \overline{Y}_k \mathbf{e}, \tag{28}$$

where the vector $\mathbf{z}$ is the unique solution to the system

$$\begin{aligned} \mathbf{z} Y_{2,2} &= \mathbf{z}, \\ \mathbf{z}\mathbf{e} &= 1, \end{aligned} \tag{29}$$

and the matrices $\overline{Y}_0$ and $\overline{Y}_k$ are obtained from the matrices Y_0 and Y_k by means of removing the block rows and block columns with numbers r, $r \in \{1, \dots, R\} \setminus L$. It is easy to check that $\overline{Y}_k = O$, $k \in \{1, \dots, \overline{K}+1\}$ and $\overline{Y}_0$ is the stochastic matrix. Consequently,

$$\mathbf{z} \overline{Y}_0 \mathbf{e} = \mathbf{z}\mathbf{e} = 1 > 0 = \mathbf{z} \sum_{k=2}^{\overline{K}+1} (k-1) \overline{Y}_k \mathbf{e}. \tag{30}$$

Theorem is proved. □

In the sequel, we assume that the ergodicity condition is fulfilled. Then, the following stationary probabilities exist:

$$\begin{aligned} &\pi(i, r, n, \nu) = \lim_{t \to \infty} P\{i_t = i,\ r_t = r,\ n_t = n,\ \nu_t = \nu\}, \\ &i \geq 0,\ r \in \{1, \dots, R\},\ n \in \left\{0, \dots, K^{(r)}\right\},\ \nu \in \{0, \dots, W\}. \end{aligned} \tag{31}$$

Let us form the row-vectors $\boldsymbol{\pi}_i$ as follows:

$$\begin{aligned} &\boldsymbol{\pi}(i, r, n) \\ &\quad = (\pi(i, r, n, 0), \pi(i, r, n, 1), \dots, \pi(i, r, n, W)), \\ &\qquad\qquad n \in \left\{0, \dots, K^{(r)}\right\}, \\ &\boldsymbol{\pi}(i, r) = \left(\boldsymbol{\pi}(i, r, 0), \boldsymbol{\pi}(i, r, 1), \dots, \boldsymbol{\pi}\left(i, r, K^{(r)}\right)\right), \\ &\qquad\qquad r \in \{1, \dots, R\}, \\ &\boldsymbol{\pi}_i = (\boldsymbol{\pi}(i, 1), \boldsymbol{\pi}(i, 2), \dots, \boldsymbol{\pi}(i, R)), \quad i \geq 0. \end{aligned} \tag{32}$$

It is well known that the vectors $\boldsymbol{\pi}_i$, $i \geq 0$, satisfy the system

$$\begin{aligned} (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \dots) Q &= \mathbf{0}, \\ (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \dots) \mathbf{e} &= 1, \end{aligned} \tag{33}$$

where Q is the generator of ξ_t, $t \geq 0$. System (33) is infinite and cannot be directly solved on computer. However, it can be successfully solved by means of the numerically stable algorithms developed in [18, 19].

4. Performance Measures of the System

Having computed the vectors of the stationary probabilities $\boldsymbol{\pi}_i$, $i \geq 0$, it is possible to compute a variety of the steady state performance parameters of the system.

The distribution of the number of the customers in orbit is

$$\lim_{t \to \infty} P\{i_t = i\} = \boldsymbol{\pi}_i \mathbf{e}, \quad i \geq 0. \tag{34}$$

The average number of customers in the service area is

$$L = \sum_{i=0}^{\infty} \sum_{r=1}^{R} \sum_{k=0}^{K^{(r)}} k \boldsymbol{\pi}(i,r,k)\, \mathbf{e}. \tag{35}$$

Here and throughout this section, formulas for the main performance measures contain the infinite sums. However, this does not create essential difficulties in computer implementation. It is well known that if the ergodicity condition is fulfilled, the stationary probability vectors $\boldsymbol{\pi}_i$ converge in norm to zero vector when i approaches infinity. So, computation of an infinite sum may be stopped if the summand becomes less than some preassigned value ε (e.g., $\varepsilon = 10^{-10}$).

The average number of customers in the buffer is

$$N_{\text{buffer}} = \sum_{i=0}^{\infty} \sum_{r=1}^{R} \sum_{k=N^{(r)}+1}^{K^{(r)}} \left(k - N^{(r)}\right) \boldsymbol{\pi}(i,r,k)\, \mathbf{e}. \tag{36}$$

The average number of busy servers is

$$N_{\text{server}} = \sum_{i=0}^{\infty} \sum_{r=1}^{R} \sum_{k=1}^{K^{(r)}} \min\left\{k, N^{(r)}\right\} \boldsymbol{\pi}(i,r,k)\, \mathbf{e}. \tag{37}$$

The average number of customers in orbit is

$$L_{\text{orbit}} = \sum_{i=1}^{\infty} i \boldsymbol{\pi}_i \mathbf{e}. \tag{38}$$

The intensity of output of customers is

$$\lambda_{\text{out}} = \sum_{i=0}^{\infty} \sum_{r=1}^{R} \sum_{k=1}^{K^{(r)}} \min\left\{k, N^{(r)}\right\} \mu^{(r)} \boldsymbol{\pi}(i,r,k)\, \mathbf{e}. \tag{39}$$

The probability that a customer arrives at the system when the buffer is full and leaves the system is

$$P^{(\text{loss-ent})} = \lambda^{-1} \sum_{i=0}^{\infty} \sum_{r=1}^{R} \left(1 - p_1^{(r)}\right) \boldsymbol{\pi}\left(i,r,K^{(r)}\right) D_1^{(r)} \mathbf{e}. \tag{40}$$

The probability that a customer arrives at the system when the buffer is full and goes into orbit is

$$P^{(\text{orb-ent})} = \lambda^{-1} \sum_{i=0}^{\infty} \sum_{r=1}^{R} p_1^{(r)} \boldsymbol{\pi}\left(i,r,K^{(r)}\right) D_1^{(r)} \mathbf{e}. \tag{41}$$

The loss probability of a customer is

$$P^{(\text{loss})} = 1 - \frac{\lambda_{\text{out}}}{\lambda}. \tag{42}$$

The probability of customers loss due to the decrease of the number of servers caused by change of the state of the RE is

$$P^{(\text{RE-loss})} = \frac{1}{\lambda} \sum_{i=0}^{\infty} \sum_{r=2}^{R} \sum_{r'=1}^{r-1} \left(1 - p_2^{(r)}\right) (H)_{r,r'} \cdot \left[\sum_{k=N^{(r')}+1}^{N^{(r)}} \left(k - N^{(r')}\right) \boldsymbol{\pi}(i,r,k) + \left(N^{(r)} - N^{(r')}\right) \sum_{k=N^{(r)}+1}^{\min\{K^{(r')}+N^{(r)}-N^{(r')},K^{(r)}\}} \boldsymbol{\pi}(i,r,k) + \sum_{k=K^{(r')}+N^{(r)}-N^{(r')}+1}^{K^{(r)}} \left(k - K^{(r')}\right) \boldsymbol{\pi}(i,r,k) \right] \cdot \mathbf{e}. \tag{43}$$

This formula is quite transparent. The right hand side of this formula represents the fraction. The denominator of this fraction is the average rate λ of customers arrival to the system. The numerator is the average rate of customers loss due to the decrease of the number of servers caused by the change of the state of the RE. Such a loss occurs, with probability $1 - p_2^{(r)}$, every time when the state r, $r \in \{2, \ldots, R\}$, of the RE changes to the state r', $r' \in \{1, \ldots, r-1\}$, while the number k of customers in service area is greater than $N^{(r')}$. The number of the simultaneously lost customers is equal to $(k - N^{(r')})$ if $k \in \{N^{(r')} + 1, \ldots, N^{(r)}\}$, equal to $(N^{(r)} - N^{(r')})$ if $k \in \{N^{(r)} + 1, \ldots, \min\{K^{(r')} + N^{(r)} - N^{(r')}, K^{(r)}\}\}$, and equal to $(k - K^{(r')})$ if $k \in \{K^{(r')} + N^{(r)} - N^{(r')} + 1, \ldots, K^{(r)}\}$. Applying the formula of total probability, we get formula (43).

The probability of an arbitrary customer loss from orbit is

$$P^{(\text{loss-from-orbit})} = P^{(\text{loss})} - P^{(\text{loss-ent})} - P^{(\text{RE-loss})}. \tag{44}$$

The probability that an arbitrary customer from orbit will make an attempt to receive service when the system is full and return to orbit is

$$P^{(\text{return-to-orbit})} = \alpha^{-1} \sum_{i=1}^{\infty} \sum_{r=1}^{R} i \alpha^{(r)} p_3^{(r)} \boldsymbol{\pi}\left(i,r,K^{(r)}\right) \mathbf{e}, \tag{45}$$

where $\alpha = \sum_{i=1}^{\infty} \sum_{r=1}^{R} i \alpha^{(r)} \boldsymbol{\pi}(i,r) \mathbf{e}$.

The probability that an arbitrary customer from orbit makes an attempt to receive service when the system area is full and leaves the system without service is

$$P^{(\text{loss-non-persistence})} = \alpha^{-1}\sum_{i=1}^{\infty}\sum_{r=1}^{R} i\alpha^{(r)}\left(1-p_3^{(r)}\right)\boldsymbol{\pi}\left(i,r,K^{(r)}\right)\mathbf{e}. \tag{46}$$

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

$$H = \begin{pmatrix} -0.03 & 0.03 \\ 0.07 & -0.07 \end{pmatrix}, \tag{47}$$

so the stationary probability of state 1 is $\varphi_1 = 0.7$ and the stationary probability of state 2 is $\varphi_2 = 0.3$. These probabilities are the components of the vector $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)$ which is computed as the unique solution to the system $\boldsymbol{\varphi}H = \mathbf{0}$, $\boldsymbol{\varphi}\mathbf{e} = 1$.

We assume that the arrival flow under state 1 of the RE is defined by the matrices

$$D_0^{(1)} = \begin{pmatrix} -2.7032 & 0 \\ 0 & -0.0877 \end{pmatrix}, \quad D_1^{(1)} = \begin{pmatrix} 2.6853 & 0.0179 \\ 0.0488 & 0.0389 \end{pmatrix}, \tag{48}$$

and under state 2 of the RE it is defined by the matrices

$$D_0^{(2)} = \begin{pmatrix} -6.7582 & 0 \\ 0 & -0.2193 \end{pmatrix}, \quad D_1^{(2)} = \begin{pmatrix} 6.7133 & 0.0449 \\ 0.1221 & 0.0972 \end{pmatrix}. \tag{49}$$

For both arrival processes, the coefficient of correlation is $c_{\text{cor}}^{(r)} = 0.2$ and the coefficient of variation is $c_{\text{var}}^{(r)} = 12.34$, $r = 1, 2$. The average arrival rate $\lambda^{(1)}$ of customers under state 1 of the RE is 2, and the average arrival rate of customers $\lambda^{(2)}$ under state 2 of the RE is 5.

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

$$\begin{aligned} K^{(1)} &= 25, \\ p_1^{(1)} &= 0.95, \\ p_2^{(1)} &= 1, \\ p_3^{(1)} &= 0.9, \\ \gamma^{(1)} &= 0.1, \\ \alpha^{(1)} &= 0.3, \\ \mu^{(1)} &= 0.4. \end{aligned} \tag{50}$$

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

$$\begin{aligned} K^{(2)} &= 25, \\ p_1^{(1)} &= 0.9, \\ p_2^{(1)} &= 1, \\ p_3^{(1)} &= 0.9, \\ \gamma^{(2)} &= 0, \\ \alpha^{(2)} &= 0.5, \\ \mu^{(2)} &= 0.5. \end{aligned} \tag{51}$$

Because arrival rate $\lambda^{(2)}$ is 2.5 times higher than $\lambda^{(1)}$ while the service rate $\mu^{(2)}$ is only 1.25 times higher than $\mu^{(1)}$, we may interpret state 1 as a normal mode of the system operation while we interpret state 2 as a congestion mode. To provide good quality of customers service, we should properly adjust the number of available servers to the mode of system. From economical considerations, let us assume that the average number of servers operating at an arbitrary moment of time should not exceed the predefined number $\widetilde{N}$. The average number of servers operating at an arbitrary moment evidently is computed by formula $N^{(1)}\varphi_1 + N^{(2)}\varphi_2$. Suppose that we have an opportunity to arbitrarily assign the numbers of available servers $N^{(1)}$ and $N^{(2)}$ that satisfy inequality

$$N^{(1)}\varphi_1 + N^{(2)}\varphi_2 \le \widetilde{N} \tag{52}$$

aiming to provide good quality of customers service. It is intuitively obvious that the number $N^{(2)}$ of available servers under operation in the congestion mode should be not less than the number of available servers $N^{(1)}$ in the normal mode.

Let us vary the number of available servers under the different states of the RE, $N^{(1)}$ and $N^{(2)}$, in such a way that $N^{(1)} \le N^{(2)}$ and inequality (52) holds good.

Figures 2–6 illustrate the dependence of some performance measures of the system for values $N^{(1)}$ and $N^{(2)}$ varying in the described set when $\widetilde{N} = 9$.

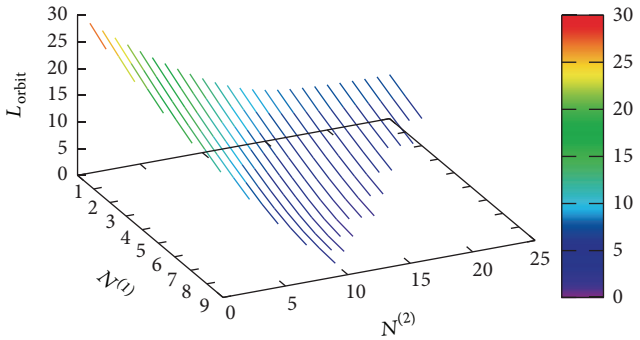

Figure 2: Dependence of the average number of customers in orbit on the numbers of servers $N^{(1)}$ and $N^{(2)}$.

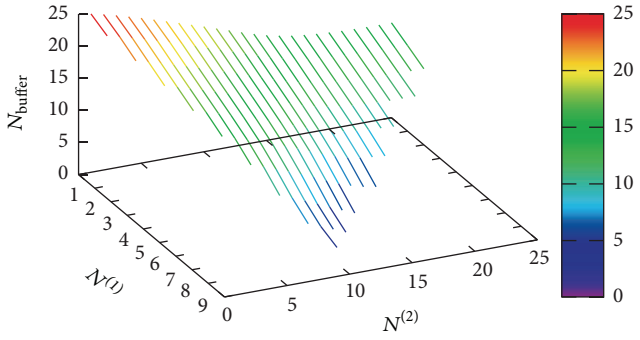

Figure 3: Dependence of the average number of customers in the buffer on the numbers of servers $N^{(1)}$ and $N^{(2)}$.

It is seen, from Figure 2, that the average number of customers in orbit significantly changes (from about 0 to about 30) when $N^{(1)}$ and $N^{(2)}$ vary, and this number is maximal when both $N^{(1)}$ and $N^{(2)}$ are small and is minimal when about 9 servers are active in both modes.

Similar observations can be made about the average number N_{buffer} of busy servers based on Figure 3. However, the value of N_{buffer} is less sensitive with respect to the change of $N^{(2)}$ compared to L_{orbit}.

It is seen, from Figure 4, that the average number of busy servers also significantly changes (from about 0 to about 7) when $N^{(1)}$ and $N^{(2)}$ vary, and this number is minimal when both $N^{(1)}$ and $N^{(2)}$ are small and maximal when about 9 servers are active in both modes.

Because the capacity of service area (the number of servers plus the number of places in a buffer) under both the states of the RE is pretty large, 25, the probability that a customer arrives at the system when the buffer is full and leaves the system is not very high (about 0.08) even when the number of active servers is small; see Figure 5. This may seem surprising. But, this effect is easily explained by Figure 6. Due to the impatience of customers, the loss probability of customers is very high (close to 1) when the number of active servers is small. So, the system is never overcrowded. It is worth noting, based on Figure 6, that the loss probability of customers due to impatience essentially decreases with growth of $N^{(2)}$ and, especially, growth of $N^{(1)}$.

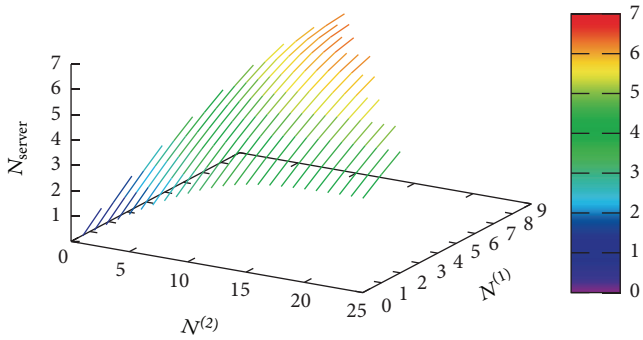

Figure 4: Dependence of the average number of busy servers on the numbers of servers $N^{(1)}$ and $N^{(2)}$.

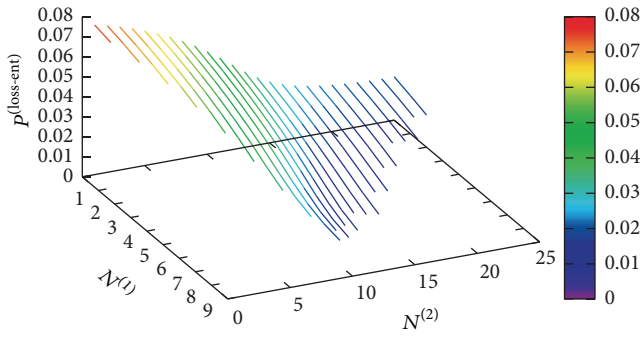

Figure 5: Dependence of the probability that a customer arrives at the system when the buffer is full and leaves the system on the numbers of servers $N^{(1)}$ and $N^{(2)}$.

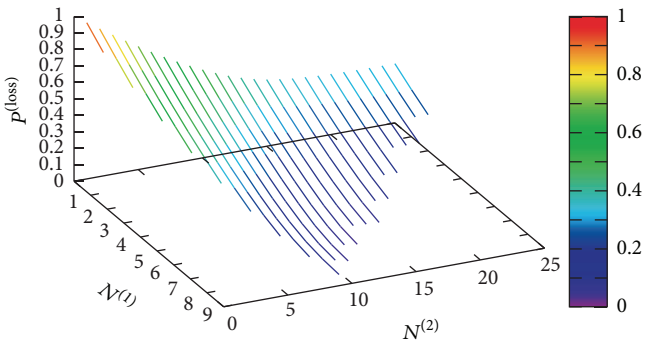

Figure 6: Dependence of the loss probability of a customer on the numbers of servers $N^{(1)}$ and $N^{(2)}$.

All the mentioned performance measures of the system are quite important. However, likely, the most important measure from the point of view of the system manager is the loss probability $P^{(loss)}$ of an arbitrary customer because the loss of a customer implies the loss of a potential profit from his/her service. So, let us consider, namely, the loss probability $P^{(loss)}$ as criterion of the quality of the system operation and find the optimal number of the servers under the different states of the RE, $N^{(1)}$, and $N^{(2)}$, under the condition that inequality (52) should be fulfilled.

Table 1: Stationary probabilities of the system states under $N^{(1)} = 7$ and $N^{(2)} = 13$.

$i\backslash r$	$r = 1$	$r = 2$
$i = 0$	0.56478	0.23043
$i = 1$	0.04455	0.01657
$i = 2$	0.02590	0.01007
$i = 3$	0.01818	0.00749
$i = 4$	0.01352	0.00597
$i = 5$	0.01027	0.00491
$i = 6$	0.00786	0.00409
$i = 7$	0.00414	0.00344
$i = 8$	0.00296	0.00290
$i = 9$	0.00214	0.00245
$i = 10$	0.00154	0.00206
$i = 11$	0.00111	0.00173
$i = 12$	$8.06E-4$	0.00145
$i = 13$	$5.86E-4$	0.00121
$i = 14$	$4.28E-4$	0.00100
$i = 15$	$3.15E-4$	$8.25E-4$

Because, to build Figure 6, we have computed the probability $P^{(\text{loss})}$ for all feasible $N^{(1)}$ and $N^{(2)}$, the solution of optimization problem is trivial. The optimal (minimal) value of the loss probability is $P^{(\text{loss})} = 0.02519$ when $N^{(1)} = 7$ and $N^{(2)} = 13$. The values of stationary probabilities $\boldsymbol{\pi}(i, r)$ under the optimal values $N^{(1)} = 7$ and $N^{(2)} = 13$ are presented in Table 1. Probabilities $\boldsymbol{\pi}(i, r)$ for $i > 15$ are less than 0.001 and they are omitted here.

It is worth noting that if $N^{(1)} = N^{(2)} = 9$, that is, the number of active servers is constant and is not adjusted to the alternation of the normal and congestion mode, the value of the loss probability is about three times higher: $P^{(\text{loss})} = 0.067389$. So, the adjustment of the number of active servers to the current load of the system definitely makes sense.

To briefly illustrate the profound effect of correlation and variance of interarrival times in the arrival process, let us assume now that, instead of the MAP flows with the coefficient of correlation 0.2 considered above, arrival flows are described by the stationary Poisson processes with the same mean arrival rates. These processes are defined by

$$\begin{aligned} D_0^{(1)} &= (-2), \\ D_1^{(1)} &= (2), \\ D_0^{(2)} &= (-5), \\ D_1^{(2)} &= (5). \end{aligned} \tag{53}$$

Let us repeat the experiment described above with these arrival processes having zero correlation and coefficient of variation equal to 1. We obtain that the minimal value of the loss probability ($P^{(\text{loss})} = 0.005102$) is achieved when $N^{(1)} = 8$ and $N^{(2)} = 11$. At the same time, under these values of $N^{(1)}$ and $N^{(2)}$, $P^{(\text{loss})}$ is equal to =0.037619 for arrival flows with $c_{\text{cor}}^{(r)} = 0.2$, $r = 1, 2$. So, the ignorance of the possible correlation in the arrival process gives relative error in the prediction of the value of $P^{(\text{loss})}$ equal to $(0.037619 - 0.005102)/0.005102 \times 100 = 637$ percent. This is not admissible in the performance evaluation and capacity planning of real life systems. Note that our experience of investigation of arrival flows in contact centers of several banks shows that these flows are indeed correlated.

6. Conclusion

We considered a multiserver queuing system, the dynamics of which depends on the state of the RE. Change of the state of the RE may cause variation of the parameters of arrival, service, retrial, impatience processes, number of available servers, and available buffer capacity, as well as the behavior of the customers. Evolution of the system is described by the multidimensional continuous-time Markov chain. The generator of this Markov chain is derived in a block matrix form. The ergodicity condition is presented. In particular, it is shown that if the customers are impatient or nonpersistent at least in one state of the RE, then the Markov chain is ergodic for any set of the system parameters. Expressions for the key performance measures are given. Numerical results illustrating the behavior of the system and showing the possibility of formulation and solution of optimization problems are provided. Positive effect of adjusting the number of active servers to the current load of the system is demonstrated. The importance of the account of correlation in the arrival processes is numerically illustrated.

Notation

$K^{(r)}$: The service area capacity under the state r of the RE

$N^{(r)}$: The number of servers under the state r of the RE

$D_0^{(r)}, D_1^{(r)}$: The square matrices of size $\overline{W} = W + 1$ that characterize MAP under the state r of the RE

$\lambda^{(r)}$: The average arrival intensity of customers under the state r of the RE

λ: The averaged arrival intensity

$\alpha^{(r)}$: The retrial intensity under the state r of the RE

$\gamma^{(r)}$: The intensity of impatience of customers from orbit under the state r of the RE

$\mu^{(r)}$: The service intensity under the state r of the RE

$p_1^{(r)}$: The probability that a customer goes to the orbit in the case of its arrival when the buffer is full under the state r of the RE

$p_2^{(r)}$: The probability that a customer goes to orbit in the case of leaving the service area due to the decrease of its capacity under the state r of the RE

$p_3^{(r)}$: The probability that a customer returns to orbit after unsuccessful attempt under the state r of the RE

$\mathbf{e}$: A column vector of appropriate size consisting of 1's

$\mathbf{0}$: A row vector of appropriate size consisting of zeroes

I: The identity matrix of the corresponding dimension

O: A zero matrix

$\oplus$ and $\otimes$: The Kronecker sum and product of matrices, respectively (see, e.g., [20, 21])

$\mathrm{diag}\{A_1, \ldots, A_l\}$: The block diagonal matrix with the diagonal entries $A_1, \ldots, A_l$

C_r: The square matrix of dimension $K^{(r)} + 1$ defined by formula $C_r = \mathrm{diag}\{0, 1, \ldots, N^{(r)}, \ldots, N^{(r)}\}$

E_l^-: The square matrix of dimension $l + 1$ with all zero entries except the entries $(E_l^-)_{k,k-1}$, $k \in \{1, \ldots, l\}$, that are equal to 1

E_l^+: The square matrix of dimension $l + 1$ with all zero entries except the entries $(E_l^+)_{k,k+1}$, $k \in \{0, \ldots, l-1\}$, that are equal to 1

$\widehat{I}_l$: The square matrix of dimension $l + 1$ with all zero entries except the entry $(\widehat{I}_l)_{l,l} = 1$.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The work is supported by Tomsk State University Competitiveness Improvement Program.

References

[1] B. V. Gnedenko and I. N. Kovalenko, *Introduction to Queueing Theory*, Science, Moscow, Russia, 1966 (Russian).

[2] U. Yechiali and P. Naor, "Queuing problems with heterogeneous arrivals and service," *Operations Research*, vol. 19, no. 3, pp. 722–734, 1971.

[3] M. Yadin and R. Syski, "Randomization of intensities in a Markov chain," *Advances in Applied Probability*, vol. 11, no. 2, pp. 397–421, 1979.

[4] C. O'Cinneide and P. Purdue, "The $M/M/\infty$ queue in a random environment," *Journal of Applied Probability*, vol. 23, no. 1, pp. 175–184, 1986.

[5] P. Purdue, "The $M/M/1$ queue in a Markovian environment," *Operations Research*, vol. 22, no. 3, pp. 562–569, 1974.

[6] M. F. Neuts, "The $M/M/1$ queue with bandomly varying arrival and service rates," *Opsearch*, vol. 15, pp. 139–157, 1978.

[7] M. Neuts, *Matrix-Geometric Solutions in Stochastic Models*, The Johns Hopkins University Press, Baltimore, Md, USA, 1981.

[8] J. D. Cordeiro and J. P. Kharoufeh, "The unreliable *M/M/1* retrial queue in a random environment," *Stochastic Models*, vol. 28, no. 1, pp. 29–48, 2012.

[9] J. Wu, Z. Liu, and G. Yang, "Analysis of the finite source *MAP/PH/N* retrial G-queue operating in a random environment," *Applied Mathematical Modelling*, vol. 35, no. 3, pp. 1184–1193, 2011.

[10] G. Yang, L.-G. Yao, and Z.-S. Ouyang, "The *MAP/PH/N* retrial queue in a random environment," *Acta Mathematicae Applicatae Sinica*, vol. 29, pp. 725–738, 2013.

[11] C. S. Kim, A. N. Dudin, V. I. Klimenok, and V. V. Khramova, "Erlang loss queueing system with batch arrivals operating in a random environment," *Computers & Operations Research*, vol. 36, no. 3, pp. 674–697, 2009.

[12] C. S. Kim, V. Klimenok, V. Mushko, and A. Dudin, "The *BMAP/PH/N* retrial queueing system operating in Markovian random environment," *Computers & Operations Research*, vol. 37, no. 7, pp. 1228–1237, 2010.

[13] S. R. Chakravarthy, "The batch Markovian arrival process: a review and future work," in *Advances in Probability Theory and Stochastic Processes*, A. Krishnamoorthy, N. Raju, and V. Ramaswami, Eds., pp. 21–49, Notable Publications, Branchburg, NJ, USA, 2001.

[14] S. R. Chakravarthy, "Markovian arrival processes," in *Wiley Encyclopedia of Operations Research and Management Science*, J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, and J. C. Smith, Eds., John Wiley & Sons, 2010.

[15] D. M. Lucantoni, "New results on the single server queue with a batch Markovian arrival process," *Communications in Statistics. Stochastic Models*, vol. 7, no. 1, pp. 1–46, 1991.

[16] B. D. Choi and Y. Chang, "MAP_1, $MAP_2/M/c$ retrial queue with the retrial group of finite capacity and geometric loss," *Mathematical and Computer Modelling*, vol. 30, no. 3-4, pp. 99–113, 1999.

[17] O. Dudina, C. Kim, and S. Dudin, "Retrial queuing system with Markovian arrival flow and phase-type service time distribution," *Computers & Industrial Engineering*, vol. 66, no. 2, pp. 360–373, 2013.

[18] V. Klimenok and A. Dudin, "Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory," *Queueing Systems*, vol. 54, no. 4, pp. 245–259, 2006.

[19] C. S. Kim, S. Dudin, O. Taramin, and J. Baek, "Queueing system MAP|PH|N|$N + R$ with impatient heterogeneous customers as a model of call center," *Applied Mathematical Modelling*, vol. 37, no. 3, pp. 958–976, 2013.

[20] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, Ellis Horwood, Chichester, UK, 1981.

[21] M. H. Lee, *Jacket Matrices: Construction and Its Applications for Fast Cooperative Wireless Signal Processing*, Lambert, Saarbrücken, Germany, 2012.