Operations Research Perspectives 7 (2020) 100139

Contents lists available at ScienceDirect

Operations Research Perspectives

journal homepage: www.elsevier.com/locate/orp

Competitive queueing systems with comparative rating dependent arrivals

A.N. Dudin*,[a,b], S.A. Dudin[a,b], O.S. Dudina[a,b], K.E. Samouylov[b]

[a] Department of Applied Mathematics and Computer Science, Belarusian State University, 4, Nezavisimosti Ave., Minsk, 220030, Belarus
[b] Applied Mathematics and Communications Technology Institute, Peoples Friendship University of Russia (RUDN University), 6, Miklukho-Maklaya st., Moscow, 117198, Russia

ARTICLE INFO

Keywords:
Queueing systems
Competition
Comparative rating
Marked Markovian arrival process
Optimization

ABSTRACT

We consider the joint operation of two multi-server queueing systems. Both systems provide the same type of service and compete for customers. It is assumed that the first system is controlled by ourselves while the second one is controlled by our competitor. The arriving customers are shared between the systems depending on the comparative rating of our system. The stationary behavior of the considered model under the fixed parameters of both systems and the fixed mechanism of calculation of the rating is described by a continuous-time multi-dimensional Markov chain. The stationary distribution of this Markov chain is computed. Expressions for the key performance measures of the systems are derived. Obtained results provide an opportunity to analyse possible consequences of various managerial actions aiming to maximize the profit of our system. A numerical experiment illustrates the application of the results for making a decision about the rationality of establishing and maintaining a service system while an alternative system providing the same type of service already exists.

1. Introduction

Situations when the same (or at least similar) type of service can be provided to customers (clients, users, etc) by different systems and these alternative systems compete for the customers are typical in real life. We can mention competition of various producers, shops, online stores, banks, insurance companies, retailers, airlines, fast food services, etc. Typically, in many such systems there are two types of customers. One of the types is the indifferent customers. If such a customer needs to obtain service and has several options of choosing a service provider, he/she makes a choice randomly, probably depending on the current availability and crowding of the corresponding service facilities. Another type of customers are the non-indifferent customers. Such customers have certain a priori preferences and make their choice depending on their taste, previous own experience or information about the providers received from the Internet. The latter information is frequently presented at some specialized web pages integrating information about the similar services. In many such information systems, the registered providers of service dynamically obtain likes or dislikes from the customers and the weighted resulting grade of each provider in some scale (e.g., in the range from 1 to 5 or from 1 to 10) is presented on the web page. The non-indifferent customer can take into account these grades in making the decision regarding the choice of the suitable service provider. Thus, the average arrival rate of such customers to the concrete system may essentially vary depending on its rating.

In our paper, we consider a situation when the number of competing providers (systems) is equal to two. Without loss of generality, instead of separate ratings of the providers we consider a comparative rating of one provider (we call it as System-1) with respect to another one (System-2). It is natural that the customer that did not succeed to receive service, e.g., he/she did not succeed to enter System-1 because it is overloaded or he/she left the queue due to a very long waiting time, may give negative evaluation of quality of service (which may cause the decrease of relative rating of System-1). Analogously, the failure of a customer in attempt to receive service at System-2 may imply the increase of the comparative rating of System-1. In this paper, we assume a concrete specific count of the comparative rating of System-1 and derive formulas for computation of the main performance measures of the systems under any fixed set of the systems parameters. Results can be used for managerial goals aiming to increase the profit of System-1. These results can be also used for searching the equilibrium points when System-1 and System-2 will cooperate.

Models with competing queues are discussed in the literature. For some survey of the state-of-art, see the recent papers [8] and [14]. In [14], a competition of two single-server queues arising in modeling two-tier healthcare service system consisting of two different service providers is analysed. The service in the system is provided by two different providers. One of them is a public service provider and the

* Corresponding author.
E-mail addresses: dudin@bsu.by (A.N. Dudin), dudin85@mail.ru (S.A. Dudin), dudina@bsu.by (O.S. Dudina), ksam@sci.pfu.edu.ru (K.E. Samouylov).

https://doi.org/10.1016/j.orp.2020.100139
Received 2 October 2019; Received in revised form 13 December 2019; Accepted 6 January 2020
Available online 14 January 2020

second is a private one. One service provider offers a subsidy or charges a premium while the other maintains the fixed service fee.

The important contributions of our paper over the existing ones are: (i) the consideration of a more general queueing model, assuming a more general, state dependent, potentially bursty and correlated arrival process, multi-server service systems, possible abandonment and jockeying of customers from one system to another; (ii) introducing the relative rating as the natural mechanism defining the choice made by the customers depending on the comparison of the current levels of the quality of service in the competing systems and illustration of the work of this mechanism. As a quite general model of the heterogeneous customers arrival, the so-called marked Markovian arrival process (*MMAP*), see [10], is considered in the existing literature. This arrival flow is the extension of the well-known Markovian arrival process (*MAP*), see [3,12,13], that is now very popular for modeling real-world arrival flows to the case when arriving customers are heterogeneous. But, due to the necessity to account of the comparative rating, which dynamically changes the distribution of types of arriving customers, the arrival flows of customers entering System-1 and System-2 are essentially more complicated than the *MAP* or *MMAP*. Such arrival processes were not considered in the existing literature. In this paper, we introduce a new heterogeneous arrival process with rating dependent arrivals which is a significant and important generalization of the *MMAP*. It is worth to stress that arrival processes entering both systems are strongly dependent and the exact mathematical analysis of the considered queueing model via its decomposition into two separate systems is not possible.

In some sense, the considered model is close to the queueing networks with parallel servers, see, e.g., [9]. However, in consideration of the networks with alternative routing of the customers it is usually assumed that the choice of the route does not dynamically depend on the state of the network. In our model, if it would be interpreted as the queueing network with two parallel multi-server stations, the choice of the station by each non-indifferent customer permanently depends on the evaluation (relative rating) of the quality of service by the customers at the alternative stations. This implies, in particular, the dependence of the stochastic processes describing the operation of the competitive service systems. In turn, this does not allow simplification of the required analysis via decomposition of the system and obtaining a product form solution. Note that queueing networks with the *MAP* arrival process are insufficiently explored in the existing literature. For references, see the recent papers [5,11] and references therein.

The remainder of the paper consists of the following. In Section 2, the mathematical model is described in detail. In Section 3, the behavior of the considered system is described by a continuous-time multidimensional Markov chain. The generator of this chain is presented and its derivation is explained. Formulas for computation of the key performance measures of the system are presented in Section 4. Section 5 contains the numerical results illustrating application of the results of the paper for making a decision about the reasonability of establishing and maintaining the service system in presence of already existing alternative system (by another owner) providing the same type of service. Section 6 concludes the paper.

2. Mathematical model

We consider a queueing model consisting of two competitive queueing systems. The first system (System-1) represents an N_1-server queue with a finite buffer of capacity $N - N_1$. The second system (System-2) is an R_1-server queue with a finite buffer of capacity $R - R_1$. The structure of the model is depicted in Fig. 1.

It is assumed that both queueing systems provide the same type of service and compete for customers. The service time of a customer at System-l has an exponential distribution with the parameter μ_l, $l = 1, 2$. System-1 operates under our control and we aim to maximize the profit of System-1 by means of the optimal choice of its

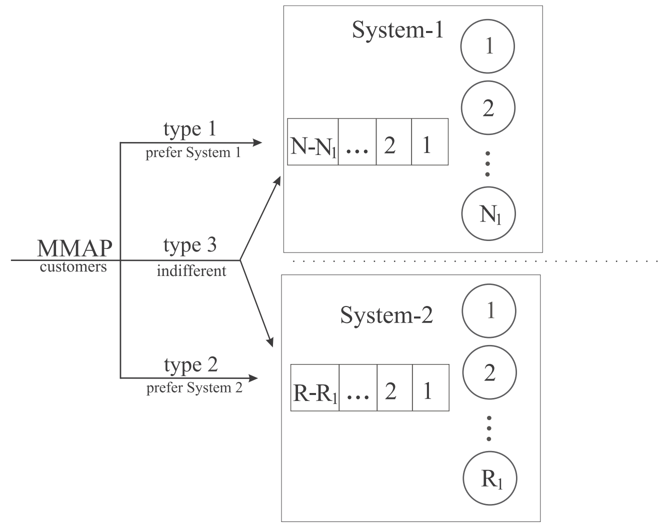

Fig. 1. Structure of the model under study.

parameters. We assume that we know the parameters of System-2, however, we cannot control this system.

The arriving customers are divided into two kinds and the arrivals are defined by the marked Markovian arrival process (*MMAP*), see He [10]. Let us denote the underlying process of the *MMAP* as $\nu_t, t \geq 0$. This process is an irreducible continuous-time Markov chain with a finite state space $\{0, 1, ..., W\}$. The behaviour of the *MMAP* is described by the matrices D_0, D, D_3. The non-diagonal entries of the matrix D_0 define the intensities of transitions of the Markov chain ν_t which are not accompanied by customer arrival. The modules of the negative diagonal entries of the matrix D_0 define the parameters of the exponential distribution of the sojourn time of the Markov chain ν_t in the corresponding states. The entries of the non-negative matrix D define the intensities of transitions of the Markov chain ν_t that are accompanied by arrival of non-indifferent customers. These customers decide to which system to enter depending on the rating of System-1 as it is explained below. The entries of the non-negative matrix D_3 define the intensities of transitions of the Markov chain ν_t that are accompanied by arrival of indifferent customers.

The intensities of transitions of this Markov chain are defined by the generator $D(1) = D_0 + D + D_3$. The average rate $\lambda_{non-indif}$ of non-indifferent customers is calculated as $\lambda_{non-indif} = \theta D\mathbf{e}$ where $\boldsymbol{\theta}$ is the row vector of the stationary state probabilities of the process ν_t defined as the unique solution of the system $\theta D(1) = \mathbf{0}$, $\theta\mathbf{e} = 1$. Here, $\mathbf{e}$ is a column vector of an appropriate size consisting of 1's and $\mathbf{0}$ is a row vector of an appropriate size consisting of zeroes.

The average rate λ_3 of indifferent customers is calculated as $\lambda_3 = \theta D_3\mathbf{e}$. The total rate λ of customers' arrival is defined as $\lambda = \lambda_{non-indif} + \lambda_3$. The squared coefficient of variation c_{var} of the intervals between successive arrivals is given by $c_{var} = 2\lambda\theta(-D_0)^{-1}\mathbf{e} - 1$. The coefficient of correlation c_{cor} of two successive intervals between arrivals is given by $c_{cor} = (\lambda\theta(-D_0)^{-1}D(1)(-D_0)^{-1}\mathbf{e} - 1)/c_{var}$.

The non-indifferent customers are divided into two classes. Type-1 customers arrive to System-1 and type-2 customers arrive to System-2. The arrival intensities of customers of different types to the corresponding systems depend on the current value of the *comparative rating* of System-1. We assume that the comparative rating admits values in the set $\{1, \cdots, K\}$. If the comparative rating of System-1 at an arbitrary moment is equal to k, the intensities of the flow of type-1 customers are given by the entries of the non-negative matrix $D_1^{(k)}$ and the intensities of the flow of type-2 customers are given by the entries of the non-negative matrix $D_2^{(k)} = D - D_1^{(k)}$. Let $\lambda_l^{(k)}$ be the average arrival rate of type-l customers, $l = 1, 2$, when the comparative rating of System-1 is equal to k, $k = \overline{1, K}$. Here and in the sequel, the notation like $k = \overline{1, K}$

means that the parameter k admits the values from the set $\{1, \cdots, K\}$. The rates $\lambda_l^{(k)}$ are computed by formula $\lambda_l^{(k)} = \theta D_l^{(k)}\mathbf{e}$ and $\lambda_1^{(k)} + \lambda_2^{(k)} = \lambda_{non-indif}$. We assume that the splitting of the matrix D into two matrices $D_1^{(k)}$ and $D_2^{(k)}$, $D = D_1^{(k)} + D_2^{(k)}$ for all values of k, $k = \overline{1, K}$, provides the fulfillment of inequalities

$$\lambda_1^{(k)} \le \lambda_1^{(k+1)},\ \lambda_2^{(k)} \ge \lambda_2^{(k+1)},\ k = \overline{1, K-1},$$

i.e., the higher value of the rating implies the higher (at least, not lower) rate of customers directing to System-1 (correspondingly, the lower rate of customers directing to System-2).

If System-l, $l = 1, 2$, is not full during the arrival epoch of type-l customer, the customer joins this system. If the buffer of System-l is full, the customer tries to enter System-$\hat{l}$, $\hat{l} = 1, 2$, $\hat{l} \ne l$, with probability p_l and with the complimentary probability this customer leaves the system (it is assumed to be lost). If the customer chooses to enter System-$\hat{l}$ but this system is full, the customer is also lost.

The indifferent (type-3) customers do not have any preference in making a choice of a system. If there are free servers in both systems at the type-3 customer arrival epoch, the customer chooses the system equiprobable. If all servers in one of the systems are busy while another system has idle servers, the customer occupies one of these idle servers. If all the servers in System-1 and System-2 are busy but buffers are not full, the type-3 customer joins the system with the shortest queue length. If the numbers of customers in the buffers are equal and there are free places in each buffer, the customer chooses the system equiprobable. If one of the systems is full but the other system has free places, the customer is admitted to the system that is not full. If at the arrival epoch of a type-3 customer both systems are full, the customer is lost.

We assume that customers are impatient. Impatience of customers, i.e., abandonment of the service in case of too long waiting in the queue, is the inherent feature of many real-world systems, see, e.g., [4]. We assume that the customer staying in the buffer of System-l leaves this system after an exponentially distributed waiting time described by the parameter α_l, $0 \le \alpha_l < \infty$, $l = 1, 2$. After leaving System-l, the customer tries to obtain service in System-$\hat{l}$, $\hat{l} = 1, 2$, $\hat{l} \ne l$, with probability p_l and with the complimentary probability this customer is assumed to be lost. If the customer chooses to enter System-$\hat{l}$, but this system is full, the customer is lost.

In our model, an important role is played by the way of calculating the comparative rating of System-1. The rating represents the estimation of comparative quality of service in System-1 and System-2 and reflects the attractiveness of System-1 over System-2 in the eyes of non-indifferent customers. Although some more complicated ways for computing the rating are possible (e.g., presence of some delay in reaction of customers to the quality of service, smoothing the rating variations, etc.) and deserve to be analysed, here we assume the following mechanism of establishing the rating. Let the comparative rating of System-1 be equal to k, $k = \overline{1, K}$. If, at an arbitrary moment a loss of a customer (due to the system overflow or impatience) occurs in System-2, with probability y the rating of System-1 increases by one, i.e., it becomes equal to $\min\{k+1, K\}$. With the complimentary probability $1 - y$ the rating does not change. If, at an arbitrary moment a loss of a customer occurs in System-1, with probability x the rating of System-1 decreases by one, i.e., it becomes equal to $\max\{1, k-1\}$. With the complimentary probability $1 - x$ the rating does not change. In potential real-world applications, the parameters y and x may reflect, e.g., the probability that a lost customer will leave a negative review in specialized web pages. We assume that if the customer is lost in both systems at the same time (a customer arrives when the systems are full or tries to join the full system after the leaving the buffer of another system), the comparative rating does not change.

It is evident that the owner or the manager of System-1 can have essential impact on the rating of his/her system. E.g., the average value of the rating of System-1 can be increased by means of the corresponding investments via the increase of the number of servers and the buffer capacity, the service rate, etc. It is natural that the increase of the comparative rating of System-1 has to lead to the reduction of the arrival flow to System-2. Therefore, the manager of System-2 can take certain actions to improve the quality of service in System-2 and implicitly to decrease the rating of System-1. In this paper, we analyse the operation of the queueing model under the fixed sets of the parameters of both systems. Based on these results, various problem formulations in terms of game theory are possible. E.g., it is possible to check whether or not it is possible to devastate the competing system or what is the value of the guaranteed profit of Server-1, what is the equilibrium point, how to reach Pareto equilibrium, etc.

3. The process of the system states

It is easily observed that the behavior of the system under study can be described in terms of the following regular irreducible continuous-time Markov chain

$$\xi_t = \{n_t,\ r_t,\ k_t,\ \nu_t\},\ t \ge 0,$$

where, at moment $t, t \ge 0$,

- n_t is the number of customers in System-1, $n_t = \overline{0, N}$;
- r_t is the number of customers in System-2, $r_t = \overline{0, R}$;
- k_t is the value of comparative rating, $k_t = \overline{1, K}$;
- ν_t is the state of the underlying process of the *MMAP*, $\nu_t = \overline{0, W}$.

Since this Markov chain is irreducible and has a finite state space, the stationary probabilities of the system states

$$\pi(n, r, k, \nu) = \lim_{t\to\infty} P\{n_t = n,\ r_t = r,\ k_t = k,\ \nu_t = \nu\},\ n = \overline{0, N},$$
$$r = \overline{0, R},\ k = \overline{1, K},\ \nu = \overline{0, W},$$

exist for any values of the system parameters.

Let us enumerate the states of the process $\xi_t, t \ge 0$, in the direct lexicographic order of the components $\{r_t, k_t, \nu_t\}$. Corresponding to this enumeration, we form the row vectors $\boldsymbol{\pi}_n$, $n = \overline{0, N}$, as follows:

$$\pi(n, r, k) = (\pi(n, r, k, 0),\ \pi(n, r, k, 1), \ldots, \pi(n, r, k, W)),\ n = \overline{0, N},$$
$$r = \overline{0, R},\ k = \overline{1, K},$$

$$\boldsymbol{\pi}(n, r) = (\boldsymbol{\pi}(n, r, 1),\ \boldsymbol{\pi}(n, r, 2), \ldots, \boldsymbol{\pi}(n, r, K)),\ n = \overline{0, N},\ r = \overline{0, R},$$

$$\boldsymbol{\pi}_n = (\boldsymbol{\pi}(n, 0),\ \boldsymbol{\pi}(n, 1), \ldots, \boldsymbol{\pi}(n, R)),\ n = \overline{0, N}.$$

It is well known that the probability vectors $\boldsymbol{\pi}_n$, $n = \overline{0, N}$, satisfy the following system of linear algebraic equations:

$$(\boldsymbol{\pi}_0,\ \boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_N)\mathbf{G} = \mathbf{0},\quad (\boldsymbol{\pi}_0,\ \boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_N)\mathbf{e} = 1 \qquad (1)$$

where $\mathbf{G}$ is the generator of the Markov chain $\xi_t, t \ge 0$.

Theorem 1. *The generator* $\mathbf{G}$ *of the Markov chain* $\xi_t, t \ge 0$, *has the following block-tridiagonal matrix structure:*

$$\mathbf{G} = \begin{pmatrix} G_{0,0} & G_{0,1} & O & \ldots & O & O \\ G_{1,0} & G_{1,1} & G_{1,2} & \ldots & O & O \\ O & G_{2,1} & G_{2,2} & \ldots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \ldots & G_{N,N-1} & G_{N,N} \end{pmatrix}.$$

The non-zero blocks have the following form:

$$G_{n,n}=\begin{pmatrix}G_{n,n}^{0,0} & G_{n,n}^{0,1} & O & \dots & O & O\\ G_{n,n}^{1,0} & G_{n,n}^{1,1} & G_{n,n}^{1,2} & \dots & O & O\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ O & O & O & \dots & G_{n,n}^{R-1,R-1} & G_{n,n}^{R-1,R}\\ O & O & O & \dots & G_{n,n}^{R,R-1} & G_{n,n}^{R,R}\end{pmatrix} \quad (2)$$

where

$$G_{n,n}^{r,r}=I_K\otimes D_0 - \left(\alpha_1\delta_{n>N_1}(n-N_1)+\alpha_2\delta_{r>R_1}(r-R_1)+\min\{n,N_1\}\mu_1+\min\{r,R_1\}\mu_2\right)I_{K\bar{W}},$$

$r=\overline{0,R-1},\ n=\overline{0,N-1},$

$$G_{n,n}^{R,R}=I_K\otimes D_0 - \left(\alpha_1\delta_{n>N_1}(n-N_1)+\alpha_2(R-R_1)+\min\{n,N_1\}\mu_1+R_1\mu_2\right)I_{K\bar{W}}+$$

$$+(1-p_2)\mathcal{D}_2((yE_K^+ +(1-y)I_K)\otimes I_{\bar{W}}),\ n=\overline{0,N-1},$$

$$G_{N,N}^{r,r}=I_K\otimes D_0 - \left(\alpha_1(N-N_1)+\alpha_2\delta_{r>R_1}(r-R_1)+N_1\mu_1+\min\{r,R_1\}\mu_2\right)I_{K\bar{W}}+$$

$$+(1-p_1)\mathcal{D}_1((xE_K^- +(1-x)I_K)\otimes I_{\bar{W}}),\ r=\overline{0,R-1},$$

$$G_{N,N}^{R,R}=I_K\otimes D_0-(\alpha_1(N-N_1)+\alpha_2(R-R_1)+N_1\mu_1+R_1\mu_2)I_{K\bar{W}}+$$

$$+(1-p_1)\mathcal{D}_1((xE_K^- +(1-x)I_K)\otimes I_{\bar{W}})+p_1\mathcal{D}_1+$$

$$+(1-p_2)\mathcal{D}_2((yE_K^+ +(1-y)I_K)\otimes I_{\bar{W}})+p_2\mathcal{D}_2+I_K\otimes D_3,$$

$$G_{n,n}^{r,r+1}=\mathcal{D}_2+(0.5\delta_{(r<R_1)\cap(n<N_1)}+\delta_{(r<R_1)\cap(n\ge N_1)}+$$

$$+0.5\delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1=n-N_1)}+\delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1<n-N_1)})I_K\otimes D_3,$$
$r=\overline{0,R-1},\ n=\overline{0,N-1},$

$$G_{N,N}^{r,r+1}=\mathcal{D}_2+I_K\otimes D_3+p_1\mathcal{D}_1((xE_K^- +(1-x)I_K)\otimes I_{\bar{W}}),$$
$r=\overline{0,R-1},\ n=\overline{0,N-1},$

$$G_{n,n}^{r,r-1}=\min\{r,R_1\}\mu_2I_{K\bar{W}}+\delta_{r>R_1}\alpha_2(r-R_1)(1-p_2)(yE_K^+ +(1-y)I_K)\otimes I_{\bar{W}},\ r=\overline{1,R},$$
$n=\overline{0,N-1},$

$$G_{N,N}^{r,r-1}=\min\{r,R_1\}\mu_2I_{K\bar{W}}+\delta_{r>R_1}\alpha_2(r-R_1)((1-p_2)(yE_K^+ +(1-y)I_K)+p_2I_K)\otimes I_{\bar{W}},$$
$r=\overline{1,R}.$

$$G_{n,n+1}=\begin{pmatrix}G_{n,n+1}^{0,0} & O & O & \dots & O & O\\ G_{n,n+1}^{1,0} & G_{n,n+1}^{1,1} & O & \dots & O & O\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ O & O & O & \dots & G_{n,n+1}^{R-1,R-1} & O\\ O & O & O & \dots & G_{n,n+1}^{R,R-1} & G_{n,n+1}^{R,R}\end{pmatrix} \quad (3)$$

where

$$G_{n,n+1}^{r,r}=\mathcal{D}_1+(0.5\delta_{(r<R_1)\cap(n<N_1)}+\delta_{(r\ge R_1)\cap(n<N_1)}+$$

$$+0.5\delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1=n-N_1)}+\delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1>n-N_1)})I_K\otimes D_3,$$
$r=\overline{0,R-1},\ n=\overline{0,N-1},$

$$G_{n,n+1}^{R,R}=\mathcal{D}_1+I_K\otimes D_3+p_2\mathcal{D}_2((yE_K^+ +(1-y)I_K)\otimes I_{\bar{W}}),$$
$n=\overline{0,N-1},$

$$G_{n,n+1}^{r,r-1}=\delta_{r>R_1}\alpha_2p_2(r-R_1)(yE_K^+ +(1-y)I_K)\otimes I_{\bar{W}},\ r=\overline{1,R},$$
$n=\overline{0,N-1}.$

$$G_{n,n-1}=\begin{pmatrix}G_{n,n-1}^{0,0} & G_{n,n-1}^{0,1} & O & \dots & O & O\\ O & G_{n,n-1}^{1,1} & G_{n,n-1}^{1,2} & \dots & O & O\\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots\\ O & O & O & \dots & G_{n,n-1}^{R-1,R-1} & G_{n,n-1}^{R-1,R}\\ O & O & O & \dots & O & G_{n,n-1}^{R,R}\end{pmatrix} \quad (4)$$

where

$$G_{n,n-1}^{r,r}=\min\{n,N_1\}\mu_1I_{K\bar{W}}+\delta_{n>N_1}\alpha_1(n-N_1)(1-p_1)(xE_K^- +(1-x)I_K)\otimes I_{\bar{W}},$$
$r=\overline{0,R-1},\ n=\overline{1,N},$

$$G_{n,n-1}^{R,R}=\min\{n,N_1\}\mu_1I_{K\bar{W}}+\delta_{n>N_1}\alpha_1(n-N_1)((1-p_1)((xE_K^- +(1-x)I_K)+p_1I_K)\otimes I_{\bar{W}},$$
$n=\overline{1,N}.$

$$G_{n,n-1}^{r,r+1}=\delta_{n>N_1}\alpha_1p_1(n-N_1)(xE_K^- +(1-x)I_K)\otimes I_{\bar{W}},\ r=\overline{0,R-1},$$
$n=\overline{1,N}.$

Here,

- *I is the identity matrix of an appropriate dimension,*
- *O is a zero matrix of an appropriate dimension,*
- *$\otimes$ indicates the symbol of the Kronecker product of matrices, see [7],*
- $\bar{W}=W+1$,
- *$\delta_{condition}=1$ if condition is true, and $\delta_{condition}=0$ otherwise,*
- *$\mathcal{D}_l$ is the block-diagonal matrix with the diagonal blocks $D_l^{(k)},\ k=\overline{1,K},\ l=1,2,$*
- *E^- and E^+ are the square matrices of size K which are used for the description of the changes of the rating when it decreases and increases, correspondingly. They are defined as follows:*

$$E^- = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad E^+ = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

Proof. of Theorem 1 is implemented by means of analysis of possible transitions of the Markov chain during an interval of the infinitesimal length.

The generator **G** has a block-tridiagonal structure (i.e. $G_{n_1,n_2} = O$ if $|n_1 - n_2| > 1$) because the probability that more than one customer arrives or leaves System-1 during an infinitesimally small interval is negligible.

- The matrices $G_{n,n}$, $n = \overline{0, N}$, have block-tridiagonal form (2) because the probability that more than one customer arrives or leaves System-2 during an infinitesimally small interval is negligible. The diagonal entries of the matrix $G_{n,n}^{r,r}$, $r = \overline{0, R}$, $n = \overline{0, N}$, are negative. They define, up to the sign, the intensities of leaving the corresponding states of the Markov chain. To compute the values of these diagonal entries, we have to analyse the existing reasons of leaving a state of the Markov chain:
  1) The underlying processes of the *MMAP* makes a transition. Basically, the corresponding intensities are defined by the diagonal entries of the matrix $I_K \otimes D_0$. However, it is necessary to take into account that not all transitions lead to the departure from the corresponding state of the Markov chain. We list several cases when the transition of the underlying processes from some state to the same state occurs with generation of customers but this does not imply the exit from the corresponding state of the Markov chain. If $r = R$, $n < N$, an arrival of a type 2 customer with probability $(1 - p_2)$ leads to the loss of this customer what does not imply the change of the rating of System-1 with probability $(1 - y)$ if the rating k is more than 1 and with probability 1 if the rating k is equal to 1. In this case, we have to add to the diagonal entries of the matrix $I_K \otimes D_0$ the diagonal entries of the matrix $(1 - p_2)\mathcal{D}_2((yE_K^+ + (1 - y)I_K) \otimes I_{\bar{W}})$. By the analogy, in the case $n = N$, $r < R$, we have to add to the diagonal entries of the matrix $I_K \otimes D_0$ the diagonal entries of the matrix $(1 - p_1)\mathcal{D}_1((xE_K^- + (1 - x)I_K) \otimes I_{\bar{W}})$. In the case $n = N$ and $r = R$, we have to add to the diagonal entries of the matrix $I_K \otimes D_0$ the diagonal entries of the matrix $(1 - p_1)\mathcal{D}_1((xE_K^- + (1 - x)I_K) \otimes I_{\bar{W}}) + p_1\mathcal{D}_1 + (1 - p_2)\mathcal{D}_2((yE_K^+ + (1 - y)I_K) \otimes I_{\bar{W}}) + p_2\mathcal{D}_2 + I_K \otimes D_3$ which define the intensities of arrival of an arbitrary type customer without changing the state of the *MMAP* underlying process and further loss of the customer without changing the rating.
  2) A customer leaves the buffer of System-1 due to impatience. Corresponding intensities are given by the entries of the matrix $\alpha_1\delta_{n>N_1}(n - N_1)I_{K\bar{W}}$.
  3) A customer leaves the buffer of System-2 due to impatience. Corresponding intensities are given by the entries of the matrix $\alpha_2\delta_{r>R_1}(r - R_1)I_{K\bar{W}}$.
  4) Service of a customer is completed at System-1. Corresponding intensities are given by the entries of the matrix $\min\{n, N_1\}\mu_1 I_{K\bar{W}}$.
  5) Service of a customer is completed at System-2. Corresponding intensities are given by the entries of the matrix $\min\{r, R_1\}\mu_2 I_{K\bar{W}}$.

  As the results of these considerations, we obtain the values of the diagonal entries of the matrix $G_{n,n}^{r,r}$.

  Now, let us compute the non-diagonal entries of the matrices $G_{n,n}^{r,r}$, $r = \overline{0, R}$, $n = \overline{0, N}$, that define the intensities of transitions that do not lead to the change of the number of customers in System-1 and System-2, but imply the change of the rating of the system and (or) of the state of the underlying processes of the *MMAP*. The events, which cause such transitions, are the following:
  1) A transition of the underlying process of the *MMAP* without generation of a customer occurs. The intensities of such transitions are defined by the non-diagonal entries of the matrix $I_K \otimes D_0$.
  2) A transition of the underlying process of the *MMAP* to another state with generation of a customer and its further loss or a transition to the same state with generation of a customer and its further loss with changing the rating. In the case $r = R$, $n < N$, the intensities of such transitions are given by the non-diagonal entries of the matrix $(1 - p_2)\mathcal{D}_2((yE_K^+ + (1 - y)I_K) \otimes I_{\bar{W}})$. In the case $n = N$, $r < R$, the intensities of such transitions are given by the non-diagonal entries of the matrix $(1 - p_1)\mathcal{D}_1((xE_K^- + (1 - x)I_K) \otimes I_{\bar{W}})$. In the case $n = N$ and $r = R$, the intensities of such transitions are given by the non-diagonal entries of the matrix $(1 - p_1)\mathcal{D}_1((xE_K^- + (1 - x)I_K) \otimes I_{\bar{W}}) + p_1\mathcal{D}_1$
  $+ (1 - p_2)\mathcal{D}_2((yE_K^+ + (1 - y)I_K) \otimes I_{\bar{W}}) + p_2\mathcal{D}_2 + I_K \otimes D_3$.

  The entries of the matrices $G_{n,n}^{r,r+1}$, $r = \overline{0, R-1}$, $n = \overline{0, N}$, define the intensities of transitions that do not lead to the change of the number n of customers in System-1 while lead to the increase of the number r of customers in System 2. The events that cause such transitions are the following:
  1) An arrival of a type-2 customer. The corresponding transition intensities are defined by the entries of the matrix $\mathcal{D}_2$.
  2) An arrival of a type-3 customer which decides to join System-2. The corresponding intensities are defined by the entries of the matrix $I_K \otimes D_3$ multiplied by $(0.5\delta_{(r<R_1)\cap(n<N_1)} + \delta_{(r<R_1)\cap(n\ge N_1)} + 0.5\delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1=n-N_1)} + \delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1<n-N_1)})$ in the case $n < N$ and by 1 in the case $n = N$.
  3) An arrival of a type-1 customer when $n = N$ and the transition of this customer to System-2. The corresponding transition intensities are defined by the entries of the matrix $p_1\mathcal{D}_1((xE_K^- + (1 - x)I_K) \otimes I_{\bar{W}})$.

  The entries of the matrices $G_{n,n}^{r,r-1}$, $r = \overline{1, R}$, $n = \overline{0, N}$, define the intensities of transitions that do not lead to the change of the number n of customers in System-1 but lead to the decrease of the number r of customers in System-2. The events that cause such transitions are the following:
  1) Service completion at System-2. The corresponding transition intensities are defined by the entries of the matrix $\min\{r, R_1\}\mu_2 I_{K\bar{W}}$.
  2) A customer leaves System-2 due to impatience and does not join System-1. The corresponding transition intensities are defined by the entries of the matrix $\delta_{r>R_1}\alpha_2(r - R_1)(1 - p_2)(yE_K^+ + (1 - y)I_K) \otimes I_{\bar{W}}$ in the case $n < N$ and by the entries of the matrix $\delta_{r>R_1}\alpha_2(r - R_1)((1 - p_2)(yE_K^+ + (1 - y)I_K) + p_2 I_K) \otimes I_{\bar{W}}$ in the case $n = N$.
- The matrix $G_{n,n+1}$, $n = \overline{0, N-1}$, defines the intensities of the transitions that lead to the increase of the number of customers in System-1 and has form (3).

  The entries of the matrices $G_{n,n+1}^{r,r}$, $r = \overline{0, R}$, $n = \overline{0, N-1}$, define the intensities of the transitions that lead to the increase of the number of customers in System-1 without the change of the number of customers in System-2. The events that cause such transitions are the following:
  1) An arrival of a type-1 customer. The corresponding intensities are defined by the entries of the matrix $\mathcal{D}_1$.
  2) An arrival of a type-3 customer which decides to join System-1. The corresponding intensities are defined by the entries of the matrix $I_K \otimes D_3$ multiplied by $(0.5\delta_{(r<R_1)\cap(n<N_1)} + \delta_{(r\ge R_1)\cap(n<N_1)} + 0.5\delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1=n-N_1)} + \delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1>n-N_1)})$ if $r < R$ and by 1 otherwise.
  3) An arrival of a type-2 customer when $r = R$ and immediate

transition of this customer to System-2. The corresponding intensities are defined by the entries of the matrix $p_2\mathcal{D}_2((yE_K^+ + (1-y)I_K) \otimes I_{\bar{W}})$.
The entries of the matrices $G_{n,n+1}^{r,r-1}$, $r=\overline{1,R}$, $n=\overline{0,N-1}$, define the intensities of the transitions that lead to the increase in the number of customers in System-1 and decrease in the number of customers in System-2. This can happen if a customer leaves System-2 and joins System-1. The corresponding intensities are defined by the entries of the matrix $\delta_{r>R_1}\alpha_2 p_2(r-R_1)(yE_K^+ + (1-y)I_K) \otimes I_{\bar{W}}$.

- The matrix $G_{n,n-1}$, $n=\overline{1,N}$, defines the intensities of transitions that lead to the decrease of the number of customers in System-1 and has form (4).
The entries of the matrices $G_{n,n-1}^{r,r}$, $r=\overline{0,R}$, $n=\overline{1,N}$, define the intensities that lead to the decrease in the number of customers in System-1 and no change of the number of customers in System-2. The events that cause such transitions are the following:
1) Service completion at System-1. The corresponding transition intensities are defined by the entries of the matrix $\min\{n, N_1\}\mu_1 I_{K\bar{W}}$.
2) A customer leaves System-1 due to impatience and does not join System-2. The corresponding intensities are defined by the entries of the matrix $\delta_{n>N_1}\alpha_1(n-N_1)(1-p_1)(xE_K^- + (1-x)I_K) \otimes I_{\bar{W}}$ in the case $r<R$ and by the entries of the matrix $\delta_{n>N_1}\alpha_1(n-N_1)((1-p_1)((xE_K^- + (1-x)I_K) + p_1 I_K) \otimes I_{\bar{W}}$ in the case $r=R$.
The entries of the matrices $G_{n,n-1}^{r,r+1}$, $r=\overline{0,R-1}$, $n=\overline{1,N}$, define the intensities of transitions that lead to the decrease in the number of customers in System-1 and the increase in the number of customers in System-2. This can happen if a customer leaves System-1 and joins System-2. The corresponding intensities are defined by the entries of the matrix $\delta_{n>N_1}\alpha_1 p_1(n-N_1)(xE_K^- + (1-x)I_K) \otimes I_{\bar{W}}$.

This completes the proof. □

Markov chains having the block-tridiagonal structure of the generator $\mathbf{G}$ are called in the literature as Level Dependent Quasi-Birth-and-Death processes. The number of equations of system (1) may be high. Therefore, to solve this system, we recommend to use some algorithm that effectively uses the sparse structure of the generator $\mathbf{G}$. In particular, we recommend the algorithm from [1]. Note that it is possible to compute the listed below important performance measures of the system without preliminary computation of the stationary distribution of the Markov chain. To this end, a memory-efficient method developed in [2] can be applied.

4. Performance measures

As soon as the vectors $\boldsymbol{\pi}_n$, $n=\overline{0,N}$, have been computed, we can determine various performance measures of the queueing systems under consideration.

The average number of customers in System-1 is

$$N_{sys}^{(1)} = \sum_{n=1}^{N} n\boldsymbol{\pi}_n\mathbf{e}.$$

The average number of customers in System-2 is

$$N_{sys}^{(2)} = \sum_{n=0}^{N}\sum_{r=1}^{R} r\boldsymbol{\pi}(n,r)\mathbf{e}.$$

The average number of busy servers in System-1 is

$$N_{serv}^{(1)} = \sum_{n=1}^{N} \min\{n, N_1\}\boldsymbol{\pi}_n\mathbf{e}.$$

The average number of busy servers in System-2 is

$$N_{sys}^{(2)} = \sum_{n=0}^{N}\sum_{r=1}^{R} \min\{r, R_1\}\boldsymbol{\pi}(n,r)\mathbf{e}.$$

The average number of customers in the buffer of System-l is

$$N_{buffer}^{(l)} = N_{sys}^{(l)} - N_{serv}^{(l)},\ l=1,2.$$

The average rating $\bar{K}$ of System-1 is

$$\bar{K} = \sum_{n=0}^{N}\sum_{r=0}^{R}\sum_{k=1}^{K} k\boldsymbol{\pi}(n,r,k)\mathbf{e}.$$

The output intensity of successfully serviced customers from System-l is

$$\mu_{out}^{(l)} = \mu_l N_{serv}^{(l)},\ l=1,2.$$

The average intensity of type-l customers is

$$\lambda_l = \sum_{n=0}^{N}\sum_{r=0}^{R}\sum_{k=1}^{K} \boldsymbol{\pi}(n,r,k)D_l^{(k)}\mathbf{e},\ l=1,2.$$

The probability that a type-3 customer will be lost upon arrival is

$$P_3^{arr-loss} = \lambda_3^{-1}\boldsymbol{\pi}(N,R)(I_K \otimes D_3)\mathbf{e}.$$

The probability that an arriving type-3 customer decides to join System-1 is

$$P_{1-best} = \sum_{n=0}^{N-1}\Big(\sum_{r=0}^{R-1}(0.5\delta_{(r<R_1)\cap(n<N_1)} + \delta_{(r\ge R_1)\cap(n<N_1)} + 0.5\delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1=n-N_1)} + \delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1>n-N_1)})\boldsymbol{\pi}(n,r)\mathbf{e} + \boldsymbol{\pi}(n,R)\mathbf{e}\Big).$$

The probability that an arriving type-3 customer decides to join System-2 is

$$P_{2-best} = \sum_{r=0}^{R-1}\Big(\sum_{n=0}^{N-1}(0.5\delta_{(r<R_1)\cap(n<N_1)} + \delta_{(r<R_1)\cap(n\ge N_1)} + 0.5\delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1=n-N_1)} + \delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1<n-N_1)})\boldsymbol{\pi}(n,r)\mathbf{e} + \boldsymbol{\pi}(N,r)\mathbf{e}\Big).$$

The probability that System-1 is full

$$P_{busy}^{(1)} = \boldsymbol{\pi}_N\mathbf{e}.$$

The probability that System-2 is full

$$P_{busy}^{(2)} = \sum_{n=0}^{N}\boldsymbol{\pi}(n,R)\mathbf{e}.$$

The average intensity of customer arrivals to System-1 is

$$\lambda_{arrival}^{(1)} = \lambda_1 + \alpha_2 p_2 N_{buffer}^{(2)} + p_2\sum_{n=0}^{N}\sum_{k=1}^{K}\boldsymbol{\pi}(n,R,k)D_2^{(k)}\mathbf{e} + \Big(\sum_{n=0}^{N-1}\Big(\sum_{r=0}^{R-1}(0.5\delta_{(r<R_1)\cap(n<N_1)} + \delta_{(r\ge R_1)\cap(n<N_1)} + 0.5\delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1=n-N_1)} + \delta_{(r\ge R_1)\cap(n\ge N_1)\cap(r-R_1>n-N_1)})\boldsymbol{\pi}(n,r) + \boldsymbol{\pi}(n,R)\Big) + \boldsymbol{\pi}(N,R)\Big)(I_K \otimes D_3)\mathbf{e}.$$

The average intensity of customer arrivals to System-2 is

$$\lambda_{arrival}^{(2)} = \lambda_2 + \alpha_1 p_1 N_{buffer}^{(1)} + p_1\sum_{r=0}^{R}\sum_{k=1}^{K}\boldsymbol{\pi}(N,r,k)D_1^{(k)}\mathbf{e} + \Big(\sum_{r=0}^{R-1}\Big(\sum_{n=0}^{N-1}(0.5\delta_{(r<R_1)\cap(n<N_1)} + \delta_{(r<R_1)\cap(n\ge N_1)} +$$

$$+\ 0.5\delta_{(r\geq R_1)\cap(n\geq N_1)\cap(r-R_1=n-N_1)} + \delta_{(r\geq R_1)\cap(n\geq N_1)\cap(r-R_1<n-N_1)})\pi(n,\ r) + \pi(N,\ r)) + \pi(N,\ R))(I_K \otimes\ D_3)\mathbf{e}.$$

Note, that we assume that the indifferent customer who arrives when both systems are busy is rejected in each system.

The loss probability of an arbitrary customer in System-l is

$$P_{loss}^{(l)} = 1 - \frac{\mu_{out}^{(l)}}{\lambda_{arrival}^{(l)}},\ l = 1,\ 2.$$

The loss probability of an arbitrary customer in System-l due to impatience is

$$P_{imp-loss}^{(l)} = \frac{\alpha_l N_{buffer}^{(l)}}{\lambda_{arrival}^{(l)}},\ l = 1,\ 2.$$

The loss probability of an arbitrary customer upon arrival to System-l is

$$P_{ent-loss}^{(l)} = P_{loss}^{(l)} - P_{imp-loss}^{(l)},\ l = 1,\ 2.$$

5. Numerical examples

The goals the numerical experiment are to demonstrate the feasibility of the presented algorithms for computation of performance measures of the system and illustrate dependencies of some measures on the number of servers in the competing systems. In this numerical experiment, we consider the question whether or not it is reasonable to open a small business (store, cafeteria, hairdressing salon, etc) given that we know that a similar business (the potential competitor) is already working in the same region. First of all, we should formulate an economic criterion of our future operation. Then, we have to choose the parameters defining the operation of our business (system) that provide the best value of the cost criterion under the known fixed values of the parameters of the competitor's system. However, to make a correct decision, we must take in mind that the competitor can react on the start of operation of our system by changing some parameters of his/her system and our initial choice of the parameters of our system may become non-optimal and our profit will become less than its expected value. Therefore, in order to make a decision about the reasonability of opening our business, it is necessary to evaluate our guaranteed potential profit as the maximum of the value of the cost criterion under all possible choices of the parameters defining operation of the system owned by our competitor.

Let a competitor has the premises where there is an opportunity to accommodate $R_1,\ R_1 \in [1, \cdots, 10]$, servers and there are $R - R_1 = 12$ places for waiting the service. Let we have an opportunity to accommodate $N_1,\ N_1 \in [1, \cdots, 15]$, servers and there are $N - N_1 = 10$ places for waiting. In this example, we assume that we and our competitor can control only the number of servers.

We assume that an economic criterion of quality of the parameters choice for our system, that defines our expected profit during a unit of time, is as follows:

$$J_1(N_1,\ R_1) = a_1\mu_{out}^{(1)} - b_1 N_1 - c_1(N - N_1)$$

where a_1 is the profit obtained from service of one customer, b_1 is the charge paid for using one server per unit of time, and c_1 is the charge paid for one place in the buffer per unit of time.

The criterion of operation of the competitor is similar:

$$J_2(N_1,\ R_1) = a_2\mu_{out}^{(2)} - b_2 R_1 - c_2(R - R_1)$$

where the meaning of the costs a_2, b_2, c_2 is the same as the meaning of the costs a_1, b_1, c_1 defined above.

Note that both $J_1(N_1,\ R_1)$ and $J_2(N_1,\ R_1)$ depend on N_1 and R_1 because the output intensities $\mu_{out}^{(l)}$ of successfully serviced customers from System-l, $l = 1,\ 2$, depend on N_1 and R_1.

Our purpose is to find the value N_1^* such that either $J_2(N_1^*,\ R_1) < 0$ for any R_1 (this means that the profit of the competitor at the unit of time is negative and his/her business may be ruined due to the start of our business) or which provides maximum to the value $\min\limits_{R_1} J_1(N_1,\ R_1)$. The value $\max\limits_{N_1}\min\limits_{R_1} J_1(N_1,\ R_1)$ defines our optimal guaranteed profit (irrespectively on the activity of the competitor). It is natural that the minimum of this value is taken only among the values of R_1 such that $J_2(N_1,\ R_1) > 0$, i.e., such a value of R_1 under the fixed value of N_1 provides the positive profit to our competitor. Otherwise, he/she does not able to use such a number R_1 of servers.

We did not apply the results to a real system and choose the parameters of the systems based on common sense reasonings. Let us assume that the parameters of the systems are defined as follows. The *MMAP* arrival flow of customers is defined by the matrices

$$D_0 = \begin{pmatrix} -\ 5.40656 & 0 \\ 0 & -\ 0.17552 \end{pmatrix},\ D = \begin{pmatrix} 4.02796 & 0.02696 \\ 0.07332 & 0.05832 \end{pmatrix},$$

$$D_3 = \begin{pmatrix} 1.34264 & 0.009 \\ 0.02444 & 0.02444 \end{pmatrix}.$$

This arrival flow has the coefficient of correlation of successive inter-arrival times $c_{cor} = 0.2$, and the coefficient of variation of inter-arrival times $c_{var} = 12$. The total customers' arrival rate is $\lambda = 4$, the average rate of indifferent customers having no preferences is $\lambda_3 = 1$, and the average rate of arrival of non-indifferent customers is $\lambda_{non-indif} = 3$.

We assume that the comparative rating of our system can change over the interval $[1,\ \cdots, K]$ where $K = 20$. Let the probabilities q_k define the dependence $D_1^{(k)} = q_k D$ of the matrices $D_1^{(k)}$ that contain the intensities of the transitions of the underlying process ν_t, which are accompanied by a customers arrival to our system, when the rating of the system is equal to k. We assume that these probabilities are given by $q_k = 0.3 + \frac{0.4}{19}(k - 1),\ k = \overline{1,\ 20}$. This implies that when our system has the lowest rating equal to 1, only 30 % of non-indifferent customers prefer to go to our system, while 70 % of non-indifferent customers prefer to go to our competitor's system. When our system has the highest rating equal to 20, then 70 % of non-indifferent customers prefer to go to our system.

The other systems' parameters are fixed as follows. The intensities of customers abandonment from the queues in System-1 and System-2 are defined by $\alpha_1 = 0.06$ and $\alpha_2 = 0.07$. The service rates are given by $\mu_1 = 0.5$ and $\mu_2 = 0.55$, correspondingly. The probabilities that a customer arrives to the system when it is full and tries to enter the alternative system are given by $p_1 = 0.65$ and $p_2 = 0.7$, respectively. The probabilities x and y are chosen as $x = 0.01,\ y = 0.01$. This means that the loss of 100 customers in each system implies the decrease or increase (by one) of the comparative rating of our system.

Below we present the results of computation of some performance measures of the systems and the values of cost criteria for ourselves and our competitor for all 150 pairs $(N_1,\ R_1)$ where $R_1 = \overline{1,\ 10}$ and $N_1 = \overline{1,\ 15}$. The total time required to compute the stationary distribution of the states and the main performance measures for all pairs (N_1, R_1) of the considered model on PC with an Intel Core i7-8700 CPU and 16 GB RAM is about 22 minutes.

Figs. 2 and 3 define the dependencies of the output intensities $\mu_{out}^{(1)}$ and $\mu_{out}^{(2)}$ of successfully serviced customers from our system (System-1) and System-2 on the values of R_1 and N_1.

These figures confirm the intuitively obvious facts that the output intensity from each system increases when the number of servers at this system increases. This increase is more essential when the competing system has less servers and, as a consequence, the smaller rating and, due to this, smaller customers arrival rate.

Figs. 4 and 5 illustrate the dependencies of the probability $P_{loss}^{(1)}$ of loss of an arbitrary customer in System-1 and the probability $P_{loss}^{(2)}$ of loss of an arbitrary customer in System-2 on the values of R_1 and N_1.

Figs. 4 and 5 show that the loss probability in each system is pretty

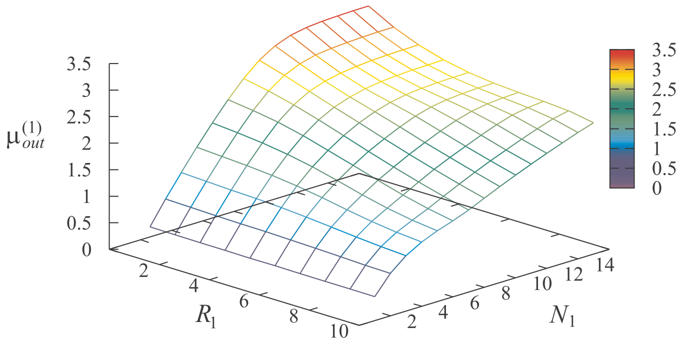

Fig. 2. Dependence of the output intensity $\mu_{out}^{(1)}$ of successfully serviced customers from System-1 on the values of R_1 and N_1

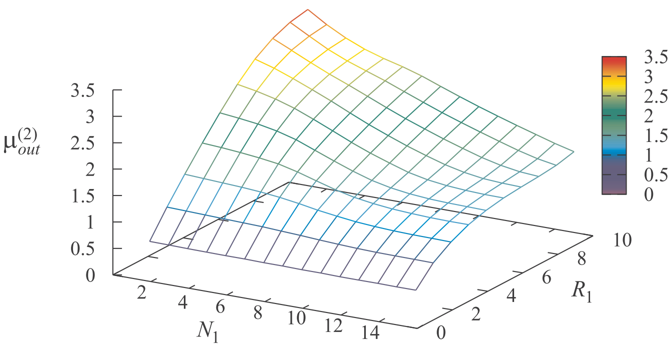

Fig. 3. Dependence of the output intensity $\mu_{out}^{(2)}$ of successfully serviced customers from System-2 on the values of R_1 and N_1

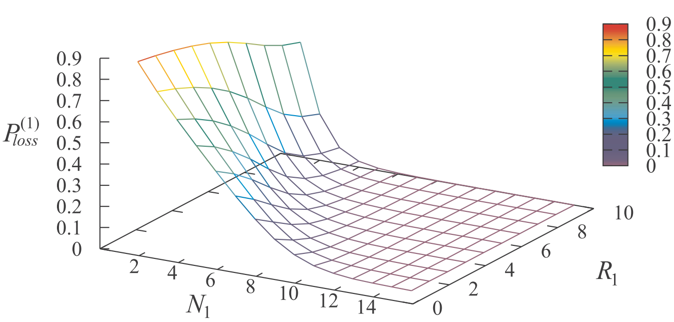

Fig. 4. Dependence of the probability $P_{loss}^{(1)}$ of loss of an arbitrary customer in System 1 on the values of R_1 and N_1

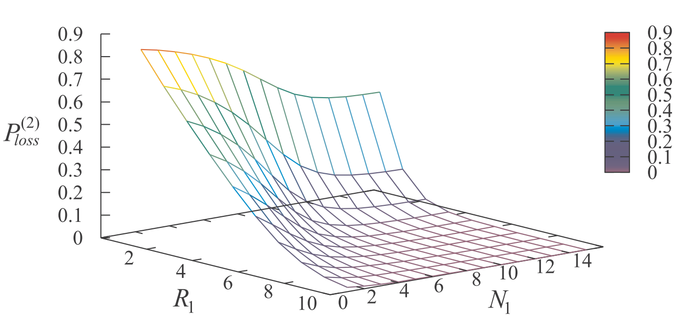

Fig. 5. Dependence of the probability $P_{loss}^{(2)}$ of loss of an arbitrary customer in System 2 on the values of R_1 and N_1

high when the number of servers in this system is small. Then this probability quickly decreases to a very small value with grows of the number of servers. The loss probability in each system also significantly depends on the number of servers in the competing system and decreases when this number grows.

Figs. 6 and 7 define the dependencies of the average rating $\bar{K}$ and the probability P_{1-best} that an arriving indifferent (type-3) customer prefers to join System-1 on the values of R_1 and N_1.

The average rating $\bar{K}$ of System-1 is about the maximum (about 20) when the number of servers at this system is large, and when the number of servers at System-2 is small. This rating is about the minimum (about 1) when the number of servers in our system is small while the number of servers in System-2 is large. More or less the same behavior as the average rating exhibits the probability P_{1-best}. However, the shape of the surfaces is a bit different. The curves given by the cuts of these surfaces, which are made parallel to axis N_1, have different intervals where they are convex or concave.

It is worth to note that essentially the *qualitative* behavior of performance measures given on Figs. 2-7 is more or less intuitively clear. The importance of the analysis presented in this paper stems from the fact that our results allow to characterise the behavior of these measures *quantitatively*.

Now let us consider the formulated above optimization problem. Let the cost coefficients in the economic criteria for System-1 and System-2 be fixed as $a_1 = 5,\ b_1 = 0.85,\ c_1 = 0.08,\ a_2 = 4.2,\ b_2 = 0.9,\ c_2 = 0.08.$

Figs. 8 and 9 show the dependence of the economic criterion $J_1(N_1, R_1)$ of System-1 and the economic criterion $J_2(N_1, R_1)$ of System-2 on the values of R_1 and N_1.

It can be verified that there is no N^* such as $J_2(N_1^*, R_1) < 0$ for any R_1 and the optimal guaranteed profit of System-1, i.e. $\max_{N_1}\min_{R_1} J_1(N_1, R_1)$ is achieved when $N_1 = 10$ and $R_1 = 6$ and is equal to $J^* = 3.06565$.

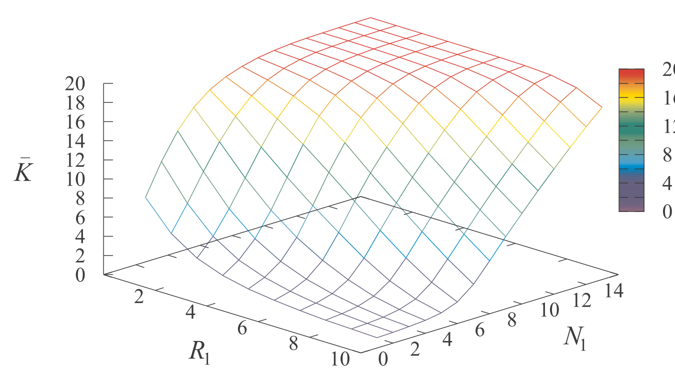

Fig. 6. Dependence of the average rating $\bar{K}$ on the values of R_1 and N_1

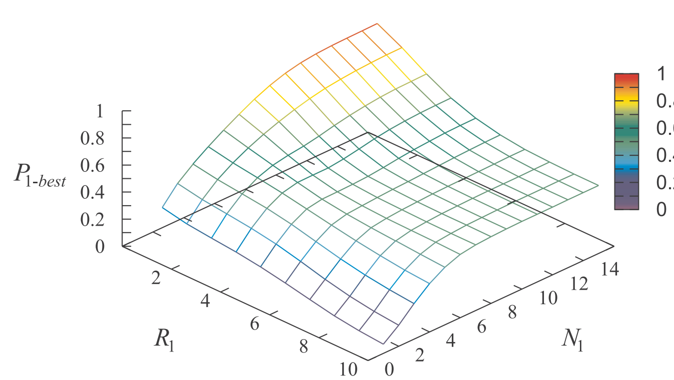

Fig. 7. Dependence of the probability P_{1-best} that an arriving type-3 customer prefers to join System-1 on the values of R_1 and N_1

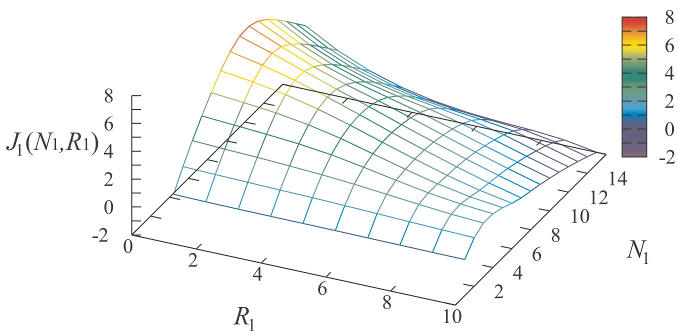

Fig. 8. Dependence of the economic criterion $J_1(N_1, R_1)$ of System-1 on the values of R_1 and N_1

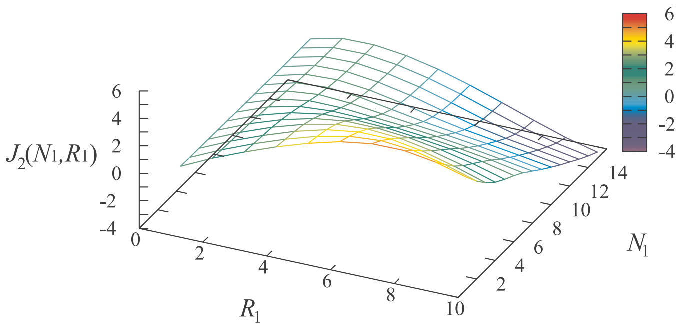

Fig. 9. Dependence of the economic criterion $J_2(N_1, R_1)$ of System-2 on the values of R_1 and N_1

Taking into account this value of the optimal guaranteed profit of our system per unit of time, the planned period of time of our system operation and the initial investments required to start our business we can make a decision whether or not this business is reasonable.

6. Conclusion

We analysed a queueing model consisting of two multi-server queues with finite buffers and having a common arrival process. These queues compete with each other. Some part of arriving customers makes a choice of the queue, in which they will try to get service, almost independently of the quality of operation of the queues. Another part makes a choice depending on the relative rating of System-1. The rating varies depending on the share of lost (due to the buffer overfull or due to impatience) customers at each queue. The stationary distribution of the states of both queues is computed. This allowed to address the problem of the optimal choice of the number of servers of System-1 under any possible choice of the number of servers of the competitive System-2. Results are illustrated by a numerical example.

The results can be extended to the case of more than two competitive queues and the systems with retrials of customers by following Dudin and Dudina [6] and Yang et al. [15].

CRediT authorship contribution statement

A.N. Dudin: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Project administration. **S.A. Dudin:** Conceptualization, Methodology, Software, Formal analysis, Writing - original draft. **O.S. Dudina:** Software, Validation, Formal analysis. **K.E. Samouylov:** Validation, Investigation, Writing - review & editing, Project administration, Supervision.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

The publication has been prepared with the support of the “RUDN University Program 5-100”.

References

[1] Baumann H, Sandmann W. Numerical solution of level dependent quasi-birth-and-death processes. Procedia Comput Sci 2010;1:1561–9.
[2] Baumann H, Sandmann W. Multi-server tandem queue with Markovian arrival process, phase-type service times, and finite buffers. Eur J Oper Res 2017;256:187–95.
[3] Chakravarthy SR. The batch Markovian arrival process: a review and future work. In: Krishnamoorthy A, Raju N, Ramaswami V, editors. Advances in probability theory and stochastic processes. New Jersey: Notable Publications Inc.; 2001. p. 21–9.
[4] Chang FM, Liu TH, Ke JC. On an unreliable-server retrial queue with customer feedback and impatience. Appl Math Modell 2018;55:171–82.
[5] Dhingra V, Kumawat GL, Roy D, de Koster R. Solving semi-open queuing networks with time-varying arrivals: an application in container terminal landside operations. Eur J Oper Res 2018;267:855–76.
[6] Dudin S, Dudina O. Retrial multi-server queueing system with PHF service time distribution as a model of a channel with unreliable transmission of information. Appl Math Modell 2019;65:676–95.
[7] Graham A. Kronecker products and matrix calculus with applications. Ellis Horwood: Cichester,UK; 1981.
[8] Guo X, Wang H, Zhang X. Analysis of two-tier healthcare service systems under value-based competition in China. Procedia Comput Sci 2018;126:1810–8.
[9] Hassin R. Rational queueing. CRC Press; 2016.
[10] He QM. Queues with marked customers. Adv Appl Probab 1996;28:567–87.
[11] Kim J, Dudin A, Dudin S, Kim C. Analysis of a semi-open queueing network with Markovian arrival process. Perform Eval 2018;120:1–19.
[12] Lucantoni DM. New results on the single server queue with a batch Markovian arrival process. Commun Stat 1991;7:1–46.
[13] Ramaswami V. The $n/g/1$ queue and its detailed analysis. Adv Appl Probab 1980;12:222–61.
[14] Wan G, Wang Q. Two-tier healthcare service systems and cost of waiting for patients. Appl Stoch Models BusInd 2017;33:167–83.
[15] Yang DY, Chang FM, Ke JC. On an unreliable retrial queue with general repeated attempts and J optional vacations. Appl Math Modell 2016;40(4):3275–88.