

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



## Performance analysis of multi-server queueing system operating under control of a random environment

Che Soong Kim<sup>1</sup>, Alexander Dudin<sup>2</sup>, Valentina Klimenok<sup>2</sup>,  
and Valentina Khramova<sup>2</sup>

<sup>1</sup>*Sangji University  
Korea*

<sup>2</sup>*Belarusian State University  
Belarus*

### 1. Introduction

Since the early 1900th the Erlang multi-server queueing systems with losses ( $B$ -model or  $M/M/N/0$  system) and with an infinite size buffer ( $C$ -model or  $M/M/N$  system) provided good mathematical tools for capacity planning and performance evaluation in the classic telephone networks for many years. Good quality of the loss probability forecasting in real world networks based on the formulas obtained for the  $M/M/N/0$  system and the delay prediction based on the formula obtained for the  $M/M/N$  system was a rather surprising because the requirement that inter-arrival and service times have an exponential distribution, which is imposed in the  $M/M/N/0$  and  $M/M/N$  models, seems to be too strict. The interest of mathematicians to the fact of good matching of the calculated under debatable assumptions characteristics to their measured value in real world systems have lead to the following two results.

By efforts of many mathematicians (A. Ya. Khinchin and B. I. Grigelionis first of all), it was proved that the superposition of a large number of independent flows having uniformly small intensity approaches to the stationary Poisson input when the number of the superposed inputs tends to infinity. It explains the fact that the flows in classic telephone networks (where flows are composed by small individual flows from independent subscribers) have the exponentially distributed inter-arrival times.

Concerning the service time distribution, situation was more complicated. The real-life measurements have shown that the service (conversation) time can not be well approximated by means of the exponentially distributed random variable. So, due to the good matching of results obtained for the  $M/M/N/0$  queue performance characteristics to characteristics of real systems modelled by such a queue, the hypothesis has arisen that the stationary state distribution in the  $M/M/N/0$  queue is the same as the one in the  $M/G/N/0$

queue conditional that the average service times in both models coincide. This property was called the invariant (or insensitivity) property of the model with respect to the service time distribution. The work [10] by B. A. Sevastjanov is the first one where this property was proven strictly.

So, the question why the Erlang's models give very good results for a practice was highlighted. The special books containing the tables for a loss probability under the given values of the number  $N$  of channels and intensities of the input and service exist. Different design problems (for a fixed value of permissible loss probability, to find the maximal intensity of the flow, which can be served by the line consisting of a fixed number of channels under the fixed average service time, or to find the necessary number of channels sufficient for transmission of the flow with a fixed intensity, etc) are solved by means of these tables.

However, the flows in the modern telecommunication networks have lost the nice properties of their predecessors in the old classic networks. In opposite to the stationary Poisson input (stationary ordinary input with no aftereffect), the modern real life flows are non-stationary, group and correlated. The *BMAP* (*Batch Markovian Arrival Process*) arrival process was introduced as a versatile Markovian point process (*VMPP*) by M.F. Neuts in the 70th. The original development of *VMPP* contained extensive notations; however these notations were simplified greatly in [7] and ever since this process bears the name *BMAP*. The class of *BMAP*s includes many input flows considered previously, such as stationary Poisson ( $M$ ), Erlangian ( $E_k$ ), Hyper-Markovian ( $HM$ ), Phase-Type ( $PH$ ), Interrupted Poisson Process ( $IPP$ ), Markov Modulated Poisson Process ( $MMPP$ ). Generally speaking, the *BMAP* is correlated, so it is ideal to model correlated and (or) bursty traffic in modern telecommunication networks.

As it was mentioned above, the question why the inter-arrival times in the classical networks have the exponential distribution was answered in literature. However, Erlang's assumption that the service time distribution has the exponential distribution is not supported by the real networks measurements. In the case of the  $M/M/N/0$  system, good fitting of performance measures of this system with the respective measures of real world systems is easily explained by Sevastjanov's result. But in the case of the  $M/M/N$  system, it was necessary to generalize results by Erlang to the cases of another, than exponential, service time distributions. This work was started by Erlang who offered so called Erlang's distribution. He introduced Erlangian of order  $k$  distribution as a distribution of a sum of  $k$  independent identically exponentially distributed random variables (phases). Further, so called phase type ( $PH$ ) distribution was introduced into consideration as the straightforward generalization of Erlangian distribution, see, e.g., [8].  $PH$  distribution includes as the special cases the exponential, Erlangian, Hyper-exponential, Coxian distributions. In our chapter we assume that service times at the fixed operation mode of the system have  $PH$  distribution.

It follows from discussion above that it is interesting to extend investigation of Erlang's models to the case of the *BMAP* input and  $PH$  type service process. This work was started by M. Combe in [1] and V. Klimenok in [5] where the *BMAP/M/N* and *BMAP/M/N/0* models, respectively, were investigated.

In paper [2], we investigated the *BMAP/PH/N/0* model having no buffer. It was shown there that the stationary distribution of the system states essentially depends on the shape of

the service time distribution and so Sevastjanov's invariant property does not hold true in the case of the general *BMAP* arrival process. Here we analyze the *BMAP/PH/N/L* system with a finite buffer and the *BMAP/PH/N* system with infinite buffer. Simultaneously, we make one more essential generalization of the model under study. Motivation of this generalization is as follows.

Even if one will use such general models of the arrival and service process as the *BMAP* and *PH*, he may fail in application to practical systems. The reason is the following. Assumption that the input flow is described by the *BMAP* allows to take into consideration a burstiness, an effect of correlation in the arrival process and variation of inter-arrival times. Assumption that the service process is described by the *PH* distribution allows to take into consideration variation of service times. But the *BMAP* arrival process and *PH* service process are assumed to be stationary and independent of each other within the borders of the models of *BMAP/PH/N/L* type,  $0 \leq L \leq \infty$ . While in many real world systems the input and service processes are not absolutely stable and may be mutually dependent. They may be influenced by some external factors, e.g., the different level of the noise in the transmission channel, hardware degradation and recovering, change of the distance by a mobile user from the base station, parallel transmission of high priority information, etc. Information transmission channel modeled by means of the *BMAP/PH/N/L* queueing system can be a part of complex communication network. The rest of the network may essentially vary characteristics of the arrival and service process in this system by means of: (i) changing the bandwidth of the channel (due to reliability factors or the needs to provide good quality of service in another parts of the network when congestion occurs); (ii) changing the mean arrival rate due breakdowns, overflow or underflow of alternative information transmission channels. Thus, to get the mathematical tool for adequate modeling such information transmission channels, more complicated queues than the *BMAP/PH/N/L* queueing system should be analyzed. These queues, in addition to the account of complicated internal structure of the arrival and service processes by means of considering the *BMAP* and *PH*, must take into account the influence of random external factors. In some extent, it can be done by means of analyzing the models of queues operating in a random environment. Such an analysis is the topic of this chapter.

Importance of investigation of the queues operating in a random environment (*RE*) drastically increased in the last years due to the following reason. The flows of information in the modern communication networks are essentially heterogeneous. Some types of information are very sensitive with respect to a delay and an jitter but tolerant with respect to losses. Another ones are tolerant with respect to the delay but very sensitive with respect to the loss of the packets. So, different schemes of the dynamic bandwidth sharing among these types exist and are developing. They assume that, in the case of congestion, transmission of the delay tolerant flows is temporarily postponed to provide better conditions for transmission of the delay sensitive flows. Analysis of such schemes requires the probabilistic analysis of the multi-dimensional processes describing transmission process of the different flows. This analysis is often impossible due to the mathematical complexity. In such a case, it is reasonable to decompose a simultaneous consideration of all flows to separate analysis of the processes of transmission of the delay sensitive and the delay tolerant flows. To this end, we model transmission of the delay sensitive flows in terms of the queues with the controlled service or (and) arrival rate where the service or the

arrival rate can be changed depending on the queue length or the waiting time. Redistribution of a bandwidth to avoid congestion for the delay sensitive flows causes a variation, at random moments, of an available bandwidth for the delay tolerant flows. Correspondingly, the queues operating in a random environment naturally arise as the mathematical model for the delay tolerant flows transmission.

Mention that the  $BMAP/PH/N/0$  model operating in the  $RE$  was recently investigated in [4]. Short overview of the recent research of queues operating in the  $RE$  can be found there. In this chapter, we consider the models  $BMAP/PH/N/L$  and  $BMAP/PH/N$  operating in the  $RE$ .

## 2. The Mathematical Model

We consider the queueing system having  $N$  identical servers. The system behavior depends on the state of the stochastic process (random environment)  $r_t, t \geq 0$ , which is assumed to be an irreducible continuous time Markov chain with the state space  $\{1, \dots, R\}$ ,  $R \geq 2$ , and the infinitesimal generator  $Q$ .

The input flow into the system is the following modification of the  $BMAP$ . In this input flow, the arrival of batches is directed by the process  $\nu_t, t \geq 0$ , (the underlying process) with the state space  $\{0, 1, \dots, W\}$ . Under the fixed state  $r$  of the  $RE$ , this process behaves as an irreducible continuous time Markov chain. Intensities of transitions of the chain  $\nu_t, t \geq 0$ , which are accompanied by arrival of  $k$ -size batch, are described by the matrices  $D_k^{(r)}, k \geq 0$ ,

$r = \overline{1, R}$ , with the generating function  $D^{(r)}(z) = \sum_{k=0}^{\infty} D_k^{(r)} z^k, |z| \leq 1$ . The matrix  $D^{(r)}(1)$  is an

irreducible generator for all  $r = \overline{1, R}$ . Under the fixed state  $r$  of the random environment, the average intensity  $\lambda^{(r)}$  (fundamental rate) of the  $BMAP$  is defined as  $\lambda^{(r)} = \theta^{(r)} (D^{(r)}(z))' \big|_{z=1} \mathbf{e}$ , and the intensity  $\lambda_b^{(r)}$  of batch arrivals is defined as  $\lambda_b^{(r)} = \theta^{(r)} (-D_0^{(r)}) \mathbf{e}$ . Here the row vector  $\theta^{(r)}$  is the solution to the equations  $\theta^{(r)} D^{(r)}(1) = \mathbf{0}, \theta^{(r)} \mathbf{e} = 1$ ,  $\mathbf{e}$  is a column vector of appropriate size consisting of 1's. The variation coefficient  $c_{var}^{(r)}$  of intervals between batch arrivals is given by

$$\left(c_{var}^{(r)}\right)^2 = 2\lambda_b^{(r)} \theta^{(r)} (-D_0^{(r)})^{-1} \mathbf{e} - 1,$$

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

$$c_{cor}^{(r)} = \left(\lambda_b^{(r)} \theta^{(r)} (-D_0^{(r)})^{-1} (D^{(r)}(1) - D_0^{(r)}) (-D_0^{(r)})^{-1} \mathbf{e} - 1\right) / \left(c_{var}^{(r)}\right)^2.$$

At the epochs of the process  $r_t, t \geq 0$ , transitions, the state of the process  $\nu_t, t \geq 0$ , is not changed, but the intensities of its transitions are immediately changed.



The service process is defined by the modification of the *PH*-type service time distribution. Service time is interpreted as the time until the irreducible continuous time Markov chain  $m_t, t \geq 0$ , with the state space  $\{0, 1, \dots, M+1\}$  reaches the absorbing state  $M+1$ . Under the fixed value  $r$  of the random environment, transitions of the chain  $m_t, t \geq 0$ , within the state space  $\{1, \dots, M\}$  are defined by an irreducible sub-generator  $S^{(r)}$  while the intensities of transition into the absorbing state are defined by the vector  $S_0^{(r)} = -S^{(r)}\mathbf{e}$ . At the service beginning epoch, the state of the process  $m_t, t \geq 0$ , is chosen according to the probabilistic row vector  $\beta^{(r)}, r = \overline{1, R}$ . It is assumed that the state of the process  $m_t, t \geq 0$ , is not changed at the epoch of the process  $r_t, t \geq 0$ , transitions. Just the exponentially distributed sojourn time of the process  $m_t, t \geq 0$ , in the current state is re-started with a new intensity defined by the sub-generator corresponding to the new state of the random environment  $r_t, t \geq 0$ .

The system under consideration has  $L, 0 \leq L \leq \infty$ , waiting positions. In the case of an infinite buffer ( $L = \infty$ ) all customers are always admitted to the system. In the case of a finite  $L$ , the system behaves as follows. If the system has all servers being busy at a batch arrival epoch, the batch looks for the available waiting position, and occupies it in case of success. If the system has all servers and all waiting positions being busy, the batch leaves the system forever and is considered to be lost. Due to a possibility of the batch arrivals, it can occur that there are free servers or waiting positions in the system at an arrival epoch, however the number of these positions is less than the number of the customers in an arriving batch. In such situation the acceptance of the customers to the system is realized according to the partial admission (*PA*) discipline (only a part of the batch corresponding to the number of free servers is allowed to enter the system while the rest of the batch is lost), the complete rejection (*CR*) discipline (a whole batch leaves the system if the number of free servers is less than the number of customers in the batch), complete admission (*CA*) discipline (a part of the batch corresponding to a number of free servers starts the service immediately while the rest of the batch waits for a service in the system in some special waiting space). All these disciplines are popular in the real life systems and got a lot of attention in the literature. Here, we consider all these disciplines.

Our aim is to calculate the stationary state distribution and main performance measures of the described queueing model.

For the use in the sequel, let us introduce the following notation:

- $\mathbf{e}_n$  ( $\mathbf{0}_n$ ) is a column (row) vector of size  $n$ , consisting of 1's (0's). Suffix may be omitted if the dimension of the vector is clear from context;
- $I(O)$  is an identity (zero) matrix of appropriate dimension (when needed the dimension of this matrix is identified with a suffix);
- $\text{diag}\{a_k, k = \overline{1, K}\}$  is a diagonal matrix with diagonal entries or blocks  $a_k$ ;
- $\otimes$  and  $\oplus$  are symbols of the Kronecker product and sum of matrices;
- $\Omega^{\otimes l} = \underbrace{\Omega \otimes \dots \otimes \Omega}_l, l \geq 1, \Omega^{\otimes 0} = 1, \Omega^{\oplus l} = \sum_{m=0}^{l-1} I_{n^m} \otimes \Omega \otimes I_{n^{l-m-1}}, l \geq 1$ , where  $n$  is the dimension of square matrix  $\Omega$ ;

$$\begin{aligned}
& \bullet D(z) = \sum_{k=0}^{\infty} \text{diag}\{D_k^{(r)}, r = \overline{1, R}\} z^k; \\
& \bullet \mathcal{D}_k^{(n)} = \text{diag}\{D_k^{(r)} \otimes I_{M^n}, r = \overline{1, R}\}, n = \overline{0, N}, k \geq 0; \\
& \bullet \mathcal{D}^{(n)}(z) = \sum_{k=0}^{\infty} \mathcal{D}_k^{(n)} z^k, n = \overline{0, N}; \\
& \bullet \mathcal{B}_l^{(n)} = \text{diag}\{I_{\overline{W}} \otimes I_{M^n} \otimes (\boldsymbol{\beta}^{(r)})^{\otimes l}, r = \overline{1, R}\}, n = \overline{1, N}, \overline{W} = W + 1; \\
& \bullet \mathcal{S}^{(n)} = \text{diag}\{I_{\overline{W}} \otimes (\mathbf{S}^{(r)})^{\oplus n}, r = \overline{1, R}\}, n = \overline{1, N}; \\
& \bullet \mathcal{S} = \text{diag}\{\mathbf{S}^{(r)}, r = \overline{1, R}\}; \\
& \bullet \mathcal{S}_0^{(n)} = \text{diag}\{I_{\overline{W}} \otimes (\mathbf{S}_0^{(r)})^{\oplus n}, r = \overline{1, R}\}, n = \overline{1, N}; \\
& \bullet \overline{\mathcal{S}}_0^{(N)} = \text{diag}\{I_{\overline{W}} \otimes (\mathbf{S}_0^{(r)} \boldsymbol{\beta}^{(r)})^{\oplus N}, r = \overline{1, R}\}; \\
& \bullet \mathcal{C}^{(n)} = Q \otimes I_{\overline{W}} \otimes I_{M^n} + \mathcal{D}_0^{(n)} + \mathcal{S}^{(n)}, n = \overline{0, N}; \\
& \bullet \overline{\mathcal{C}}^{(n)} = Q \otimes I_{\overline{W}} \otimes I_{M^{\min\{n, N\}}} + \mathcal{D}_0^{(\min\{n, N\})} + \sum_{k=N+L-n+1}^{\infty} \mathcal{D}_k^{(\min\{n, N\})} + \mathcal{S}^{(\min\{n, N\})}, n = \overline{0, N}; \\
& \bullet \overline{\mathcal{C}}^{(N+L)} = Q \otimes I_{\overline{W}} \otimes I_{M^N} + \sum_{k=0}^{\infty} \mathcal{D}_k^{(N)} + \mathcal{S}^{(N)}.
\end{aligned}$$

### 3. Process of the System States

It is easy to see that operation of the considered queueing model is described in terms of the regular irreducible continuous-time Markov chain

$$\xi_t = \{n_t, r_t, \nu_t, m_t^{(1)}, \dots, m_t^{(\min\{n_t, N\})}\}, t \geq 0,$$

where

•  $n_t$  is the number of customers in the system, where  $n_t = \overline{0, N+L}$  in case of *PA* and *CR* disciplines and  $n_t \geq 0$  in case of *CA* discipline;

•  $r_t$  is the state of the random environment,  $r_t = \overline{1, R}$ ;

•  $\nu_t$  is the state of the *BMAP* underlying process,  $\nu_t = \overline{0, W}$ ;

•  $m_t^{(n)}$  is the phase of *PH* service process in the  $n$ th busy server,  $m_t^{(n)} = \overline{1, M}$ ,  $n_t = \overline{1, N}$ , (we assume here that the busy servers are numerated in order of their occupying, i.e. the server, which begins the service, is appointed the maximal number among all busy servers; when some server finishes the service, the servers are correspondingly enumerated) at epoch  $t, t \geq 0$ .

Let us enumerate the states of the chain  $\xi_t, t \geq 0$ , in the lexicographic order and form the row vectors  $\mathbf{p}_n$  of probabilities corresponding to the state  $n$  of the first component of the process  $\xi_t, t \geq 0$ . Denote also  $\mathbf{p} = (\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \dots)$ .

It is well known that the vector  $\mathbf{p}$  satisfies the system of the linear algebraic equations (so called equilibrium equations or Chapman-Kolmogorov equations) of the form:

$$\mathbf{p}A = \mathbf{0}, \quad \mathbf{p}\mathbf{e} = 1, \quad (1)$$

where  $A$  is the infinitesimal generator of the Markov chain  $\xi_t, t \geq 0$ .

Structure of this generator and methods of system (1) solution vary depending on the admission discipline.

### 3.1. The Case of Partial Admission Discipline

**Lemma 1.** Infinitesimal generator  $A$  of the Markov chain  $\xi_t, t \geq 0$ , in the case of partial admission discipline has the following block structure:

$$A = (A_{n,n'})_{n,n'=0,\overline{N+L}} =$$

$$= \begin{pmatrix} \mathcal{C}^{(0)} & \Psi_{1,1}^{(0)} & \dots & \Psi_{N-1,N-1}^{(0)} & \Psi_{N,N}^{(0)} & \Psi_{N+1,N}^{(0)} & \dots & \Psi_{N+L-1,N}^{(0)} & \hat{\Psi}_{N+L,N}^{(0)} \\ \mathcal{S}_0^{(1)} & \mathcal{C}^{(1)} & \dots & \Psi_{N-2,N-2}^{(1)} & \Psi_{N-1,N-1}^{(1)} & \Psi_{N,N-1}^{(1)} & \dots & \Psi_{N+L-2,N-1}^{(1)} & \hat{\Psi}_{N+L-1,N-1}^{(1)} \\ O & \mathcal{S}_0^{(2)} & \dots & \Psi_{N-3,N-3}^{(2)} & \Psi_{N-2,N-2}^{(2)} & \Psi_{N-1,N-2}^{(2)} & \dots & \Psi_{N+L-3,N-2}^{(2)} & \hat{\Psi}_{N+L-2,N-2}^{(2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \dots & \mathcal{C}^{(N-1)} & \Psi_{1,1}^{(N-1)} & \Psi_{2,1}^{(N-1)} & \dots & \Psi_{L,1}^{(N-1)} & \hat{\Psi}_{L+1,1}^{(N-1)} \\ O & O & \dots & \mathcal{S}_0^{(N)} & \mathcal{C}^{(N)} & \mathcal{D}_1^{(N)} & \dots & \mathcal{D}_{L-1}^{(N)} & \hat{\mathcal{D}}_L^{(N)} \\ O & O & \dots & O & \overline{\mathcal{S}}_0^{(N)} & \mathcal{C}^{(N)} & \dots & \mathcal{D}_{L-2}^{(N)} & \hat{\mathcal{D}}_{L-1}^{(N)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \dots & O & O & O & \dots & \overline{\mathcal{S}}_0^{(N)} & \overline{\mathcal{C}}^{(N+L)} \end{pmatrix}$$

where

$$\Psi_{m,m'}^{(k)} = \mathcal{D}_m^{(k)} \mathcal{B}_{m'}^{(k)}, \quad k = \overline{0, N-1}, \quad m \geq 1, \quad m' = \overline{1, N},$$

$$\hat{\Psi}_{m_1,m'}^{(k)} = \sum_{m=m_1}^{\infty} \Psi_{m,m'}^{(k)}, \quad \hat{\mathcal{D}}_{m_1}^{(k)} = \sum_{m=m_1}^{\infty} \mathcal{D}_m^{(k)}.$$

Proof of the Lemma follows from analysis of Markov chain  $\xi_t, t \geq 0$ , transitions during an infinitesimal interval. Block entries of the generator have the following meaning. The non-diagonal entries of the matrix  $\mathcal{C}^{(k)}$  define intensity of transition of the components  $\{r_t, v_t, m_t^{(1)}, \dots, m_t^{(n_t)}\}$  of the Markov chain  $\xi_t, t \geq 0$ , which do not lead to the change of the number  $k$  of busy servers. The diagonal entries of the matrix  $\mathcal{C}^{(k)}$  are negative and define, up to the sign, intensity of leaving the corresponding states of the Markov chain  $\xi_t, t \geq 0$ . The entries of the matrix  $\Psi_{m,m'}^{(k)} = \mathcal{D}_m^{(k)} \mathcal{B}_{m'}^{(k)}$  define intensity of transitions of the components  $\{r_t, v_t, m_t^{(1)}, \dots, m_t^{(n_t)}\}$  of the Markov chain  $\xi_t, t \geq 0$ , which are accompanied by arrival of  $m$



customers and occupying  $m'$  servers conditional the number of busy servers is  $k$ . The entries of the matrix  $S_0^{(k)}$  define intensity of transitions, which are accompanied by a departure of a customer, conditional the number of busy servers is  $k$ .

To solve system (1) with the matrix  $A$  defined by Lemma 1, we use the effective numerically stable procedure developed in [2] that exploits the special structure of the matrix  $A$  (it is upper block Hessenberg) and probabilistic meaning of the unknown vector  $\mathbf{p}$ . This procedure is given by the following statement.

**Theorem 1.** In case of partial admission, the stationary probability vectors  $\mathbf{p}_i, i = \overline{0, N+L}$ , are computed as follows:

$$\mathbf{p}_l = \mathbf{p}_0 F_l, l = \overline{1, N+L},$$

where the matrices  $F_l$  are calculated recurrently:

$$F_l = \left( \bar{A}_{0,l} + \sum_{i=1}^{l-1} F_i \bar{A}_{i,l} \right) \left( -\bar{A}_{l,l} \right)^{-1}, l = \overline{1, N+L-1},$$

$$F_{N+L} = \left( A_{0,N+L} + \sum_{i=1}^{N+L-1} F_i A_{i,N+L} \right) \left( -A_{N+L,N+L} \right)^{-1},$$

the matrices  $\bar{A}_{l,N+L}$  are calculated from the backward recursions:

$$\bar{A}_{l,N+L} = A_{l,N+L}, i = \overline{0, N+L},$$

$$\bar{A}_{i,l} = A_{i,l} + \bar{A}_{i,l+1} G_l, i = \overline{0, l}, l = N+L-1, N+L-2, \dots, 0,$$

the matrices  $G_i, i = \overline{0, N+L-1}$ , are calculated from the backward recursion:

$$G_i = \left( -A_{i+1,i+1} - \sum_{l=1}^{N+L-i-1} A_{i+1,i+1+l} G_{i+l} G_{i+l-1} \dots G_{i+1} \right)^{-1} A_{i+1,i},$$

$$i = N+L-1, N+L-2, \dots, 0,$$

the vector  $\mathbf{p}_0$  is calculated as the unique solution to the following system of linear algebraic equations:

$$\mathbf{p}_0 \bar{A}_{0,0} = 0, \quad \mathbf{p}_0 \left( \sum_{l=1}^{N+L} F_l \mathbf{e} + \mathbf{e} \right) = 1.$$

### 3.2. The Case of Complete Rejection Discipline

**Lemma 2.** Infinitesimal generator  $A$  of the Markov chain  $\xi_t, t \geq 0$ , in the case of complete rejection discipline has the following block structure:

$$A = (A_{n,n'})_{n,n'=\overline{0, N+L}} =$$

$$= \begin{pmatrix} \bar{\mathcal{C}}^{(0)} & \Psi_{1,1}^{(0)} & \dots & \Psi_{N-1,N-1}^{(0)} & \Psi_{N,N}^{(0)} & \Psi_{N+1,N}^{(0)} & \dots & \Psi_{N+L-1,N}^{(0)} & \Psi_{N+L,N}^{(0)} \\ \mathcal{S}_0^{(1)} & \bar{\mathcal{C}}^{(1)} & \dots & \Psi_{N-2,N-2}^{(1)} & \Psi_{N-1,N-1}^{(1)} & \Psi_{N,N-1}^{(1)} & \dots & \Psi_{N+L-2,N-1}^{(1)} & \Psi_{N+L-1,N-1}^{(1)} \\ O & \mathcal{S}_0^{(2)} & \dots & \Psi_{N-3,N-3}^{(2)} & \Psi_{N-2,N-2}^{(2)} & \Psi_{N-1,N-2}^{(2)} & \dots & \Psi_{N+L-3,N-2}^{(2)} & \Psi_{N+L-2,N-2}^{(2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \dots & \bar{\mathcal{C}}^{(N-1)} & \Psi_{1,1}^{(N-1)} & \Psi_{2,1}^{(N-1)} & \dots & \Psi_{L,1}^{(N-1)} & \Psi_{L+1,1}^{(N-1)} \\ O & O & \dots & \mathcal{S}_0^{(N)} & \bar{\mathcal{C}}^{(N)} & \mathcal{D}_1^{(N)} & \dots & \mathcal{D}_{L-1}^{(N)} & \mathcal{D}_L^{(N)} \\ O & O & \dots & O & \bar{\mathcal{S}}_0^{(N)} & \bar{\mathcal{C}}^{(N)} & \dots & \mathcal{D}_{L-2}^{(N)} & \mathcal{D}_{L-1}^{(N)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \dots & O & O & O & \dots & \bar{\mathcal{S}}_0^{(N)} & \bar{\mathcal{C}}^{(N+L)} \end{pmatrix}.$$

The proof of the Lemma is analogous to the proof of the previous Lemma and takes into account the fact that the number of customers in the system does not change when the number of customers in an arriving batch exceeds the number of free servers. To solve system (1) with the matrix  $A$  defined by Lemma 2, we also use the procedure described by Theorem 1.

3.3. The Case of Complete Admission Discipline

**Lemma 3.** Infinitesimal generator  $A$  of the Markov chain  $\xi_t, t \geq 0$ , in the case of complete admission discipline has the following block structure:

$$A = (A_{n,n'})_{n,n' \geq 0} = \begin{pmatrix} \mathcal{C}^{(0)} & \Psi_{1,1}^{(0)} & \dots & \Psi_{N,N}^{(0)} & \Psi_{N+1,N}^{(0)} & \dots & \Psi_{N+L-1,N}^{(0)} & \Psi_{N+L,N}^{(0)} & \Psi_{N+L+1,N}^{(0)} & \dots \\ \mathcal{S}_0^{(1)} & \bar{\mathcal{C}}^{(1)} & \dots & \Psi_{N-1,N-1}^{(1)} & \Psi_{N,N-1}^{(1)} & \dots & \Psi_{N+L-2,N-1}^{(1)} & \Psi_{N+L-1,N-1}^{(1)} & \Psi_{N+L,N-1}^{(1)} & \dots \\ O & \mathcal{S}_0^{(2)} & \dots & \Psi_{N-2,N-2}^{(2)} & \Psi_{N-1,N-2}^{(2)} & \dots & \Psi_{N+L-3,N-2}^{(2)} & \Psi_{N+L-2,N-2}^{(2)} & \Psi_{N+L-1,N-2}^{(2)} & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \\ O & O & \dots & \Psi_{1,1}^{(N-1)} & \Psi_{2,1}^{(N-1)} & \dots & \Psi_{L,1}^{(N-1)} & \Psi_{L+1,1}^{(N-1)} & \Psi_{L+2,1}^{(N-1)} & \dots \\ O & O & \dots & \bar{\mathcal{C}}^{(N)} & \mathcal{D}_1^{(N)} & \dots & \mathcal{D}_{L-1}^{(N)} & \mathcal{D}_L^{(N)} & \mathcal{D}_{L+1}^{(N)} & \dots \\ O & O & \dots & \bar{\mathcal{S}}_0^{(N)} & \bar{\mathcal{C}}^{(N)} & \dots & \mathcal{D}_{L-2}^{(N)} & \mathcal{D}_{L-1}^{(N)} & \mathcal{D}_L^{(N)} & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \\ O & O & \dots & O & O & \dots & \bar{\mathcal{C}}^{(N)} & \mathcal{D}_1^{(N)} & \mathcal{D}_2^{(N)} & \dots \\ O & O & \dots & O & O & \dots & \bar{\mathcal{S}}_0^{(N)} & \bar{\mathcal{C}}^{(N+L)} & O & \dots \\ O & O & \dots & O & O & \dots & O & \bar{\mathcal{S}}_0^{(N)} & \bar{\mathcal{C}}^{(N+L)} & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Essential difference of complete admission discipline is that the state space of the Markov chain  $\xi_t, t \geq 0$ , is infinite and this makes its analysis more complicated. However, the block

rows, except the first  $N + L$  boundary block rows, have only two non-zero blocks and this Markov chain behaves as Quasi-Death process when the state of the first component  $n_t$  of the Markov chain  $\xi_t, t \geq 0$ , is greater than  $N + L$ . It allows to construct effective stable algorithm for calculation of the stationary distribution of this Markov chain. Note, that although the state space of the Markov chain  $\xi_t, t \geq 0$ , is infinite, this Markov chain is ergodic under the standard assumptions about the parameters of the *BMAP* input, the *PH* type service and the random environment. The algorithm for calculation of the stationary distribution is given in the following statement.

**Theorem 2.** In case of complete admission discipline, the stationary probability vectors  $\mathbf{p}_l, l \geq 0$ , are calculated as follows:

$$\mathbf{p}_l = \mathbf{p}_0 F_l, l \geq 1,$$

where the matrices  $F_l$  are calculated recurrently

$$F_l = \left( \bar{A}_{0,l} + \sum_{i=1}^{l-1} F_i \bar{A}_{i,l} \right) \left( -\bar{A}_{l,l} \right)^{-1}, l \geq 1,$$

the matrices  $\bar{A}_{i,l}$  are calculated as:

$$\bar{A}_{i,l} = A_{i,l} + \sum_{k=1}^{\infty} A_{i,l+k} G^{\max\{0, l+k-N-L\}} G_{\min\{N+L, l+k\}-1} G_{\min\{N+L, l+k\}-2} \dots G_l, \quad i = 0, \dots, l, \quad l \geq 1,$$

the matrix  $G$  has a form

$$G = -\left( \bar{C}^{(N+L)} \right)^{-1} \bar{S}_0^{(N)},$$

the matrices  $G_i, i = \overline{0, N+L-1}$ , are calculated from the backward recursion

$$G_i = -\left( A_{i+1,i+1} + \sum_{l=i+2}^{\infty} A_{i+1,l} G^{\max\{0, l-N-L\}} \cdot G_{\min\{N+L, l\}-1} \cdot G_{\min\{N+L, l\}-2} \dots G_{i+1} \right)^{-1} A_{i+1,i},$$

the vector  $\mathbf{p}_0$  is the unique solution of the system:

$$\mathbf{p}_0 \bar{A}_{0,0} = 0, \quad \mathbf{p}_0 \left( \sum_{l=1}^{\infty} F_l \mathbf{e} + \mathbf{e} \right) = 1.$$

The proof of the Theorem follows from the theory of multi-dimensional Markov chains with continuous time, see [6]. It is worth to note that Neuts' matrix  $G$ , which is usually found numerically as solution to matrix equation, see [9], here is obtained in the explicit form.

### 3.4. The Case of an Infinite Size of a Buffer ( $L=\infty$ )

The system under consideration in this section has an infinite waiting space. If an arriving batch of customers sees idle servers, a part of the batch corresponding to the number of free servers occupy these servers while the rest of the batch joins the queue. If the system has all servers being busy at a batch arrival epoch, all customer of the batch go to the queue.

**Lemma 3.** Infinitesimal generator  $A$  of the Markov chain  $\xi_t, t \geq 0$ , has the following block structure:

$$A = (A_{n,n'})_{n,n' \geq 0} = \begin{pmatrix} \mathcal{C}^{(0)} & \Psi_{1,1}^{(0)} & \Psi_{2,2}^{(0)} & \cdots & \Psi_{N-1,N-1}^{(0)} & \Psi_{N,N}^{(0)} & \Psi_{N+1,N}^{(0)} & \Psi_{N+2,N}^{(0)} & \cdots \\ \mathcal{S}_0^{(1)} & \mathcal{C}^{(1)} & \Psi_{1,1}^{(1)} & \cdots & \Psi_{N-2,N-2}^{(1)} & \Psi_{N-1,N-1}^{(1)} & \Psi_{N,N-1}^{(1)} & \Psi_{N+1,N-1}^{(1)} & \cdots \\ O & \mathcal{S}_0^{(2)} & \mathcal{C}^{(2)} & \cdots & \Psi_{N-3,N-3}^{(2)} & \Psi_{N-2,N-2}^{(2)} & \Psi_{N-1,N-2}^{(2)} & \Psi_{N,N-2}^{(2)} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots \\ O & O & O & \cdots & \mathcal{C}^{(N-1)} & \Psi_{1,1}^{(N-1)} & \Psi_{2,1}^{(N-1)} & \Psi_{3,1}^{(N-1)} & \cdots \\ O & O & O & \cdots & \mathcal{S}_0^{(N)} & \mathcal{C}^{(N)} & \mathcal{D}_1^{(N)} & \mathcal{D}_2^{(N)} & \cdots \\ O & O & O & \cdots & O & \overline{\mathcal{S}}_0^{(N)} & \mathcal{C}^{(N)} & \mathcal{D}_1^{(N)} & \cdots \\ O & O & O & \cdots & O & O & \overline{\mathcal{S}}_0^{(N)} & \mathcal{C}^{(N)} & \cdots \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (2)$$

In what follows we perform the steady state analysis of the Markov chain having generator of form (2). To this end, we use the results for continuous time multi-dimensional Markov chain (QTMC) presented in [6].

**Theorem 3.** The necessary and sufficient condition for existence of the Markov chain  $\xi_t, t \geq 0$ , stationary distribution is the fulfillment of the inequality

$$\rho = \lambda / \bar{\mu} < 1, \quad (3)$$

where

$$\begin{aligned} \lambda &= \mathbf{x}_1 D'(z) \Big|_{z=1} \mathbf{e}, \\ \bar{\mu} &= \mathbf{x}_2 \text{diag} \left\{ \left( \mathbf{S}_0^{(r)} \right)^{\oplus N}, r = \overline{1, R} \right\} \mathbf{e}, \end{aligned} \quad (4)$$

the vectors  $\mathbf{x}_n, n = 1, 2$ , are the unique solutions to the following systems of linear algebraic equations:

$$\mathbf{x}_1 (Q \otimes I_{\overline{W}} + D(1)) = \mathbf{0}, \quad \mathbf{x}_1 \mathbf{e} = 1, \quad (5)$$

$$\mathbf{x}_2 \left( Q \otimes I_{M^N} + \text{diag} \left\{ \left( \mathbf{S}^{(r)} + \mathbf{S}_0^{(r)} \beta^{(r)} \right)^{\oplus N}, r = \overline{1, R} \right\} \right) = \mathbf{0}, \quad \mathbf{x}_2 \mathbf{e} = 1. \quad (6)$$

*Proof.* Using the results of [6], we directly obtain the desired condition in the form of inequality

$$\mathbf{x} \left[ \mathcal{C}^{(N)} z + z \mathcal{D}^{(N)}(z) \right]_{z=1}' \mathbf{e} < 0, \quad (7)$$

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

$$\mathbf{x} \left( Q \otimes I_{\overline{WM}^N} + \mathcal{D}^{(N)}(1) + \mathcal{S}^{(N)} + \overline{\mathcal{S}}_0^{(N)} \right) = \mathbf{0}, \quad \mathbf{x} \mathbf{e} = 1. \quad (8)$$

It is easy to show that inequality (7) is reduced to the following inequality:

$$\mathbf{x}_1 \mathcal{D}'(z) \Big|_{z=1} \mathbf{e} < \mathbf{x}_2 \text{diag} \left\{ \left( \mathbf{S}_0^{(r)} \right)^{\oplus N}, r = \overline{1, R} \right\} \mathbf{e}, \quad (9)$$

where  $\mathbf{x}_1 = \mathbf{x} (I_{R\overline{W}} \otimes \mathbf{e}_{M^N})$ ,  $\mathbf{x}_2 = \mathbf{x} (I_R \otimes \mathbf{e}_{\overline{W}} \otimes I_{M^N})$ .

To get the equations for the row vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , we multiply equation (8) by the matrices  $I_{R\overline{W}} \otimes \mathbf{e}_{M^N}$  and  $I_R \otimes \mathbf{e}_{\overline{W}} \otimes I_{M^N}$  respectively. After multiplication and some algebra we obtain equations (5), (6) for the vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$ . So, inequality (9) is equivalent to inequality (3) and the theorem is proved.

The value  $\rho$  has a meaning of the system load. In what follows we assume inequality (3) be fulfilled.

To solve system (1) with the matrix  $A$  defined by (2), we use the effective numerically stable procedure [6] based on the account special structure of the matrix  $A$ , notion of the censored Markov chain and probabilistic meaning of the unknown vector  $\mathbf{p}$ . For more detail see [6]. This procedure is given by the following statement.

**Theorem 4.** The stationary probability vectors  $\mathbf{p}_l, l \geq 0$ , are calculated as follows:

$$\mathbf{p}_l = \mathbf{p}_0 F_l, l \geq 1,$$

where the matrices  $F_l$  are calculated recurrently:

$$F_l = \left( \overline{A}_{0,l} + \sum_{i=1}^{l-1} F_i \overline{A}_{i,l} \right) \left( -\overline{A}_{l,l} \right)^{-1}, l \geq 1,$$

the matrices  $\overline{A}_{i,l}$  are calculated as:

$$\begin{aligned} \overline{A}_{i,l} &= A_{i,l} + \overline{A}_{i,l+1} G_l, \quad 0 \leq i \leq l < N, \\ \overline{A}_{i,l} &= A_{i,l} + \overline{A}_{i,l+1} G, \quad l \geq \max\{i, N\}, \quad i \geq 0, \end{aligned}$$

the matrix  $G$  is calculated from the equation

$$G = \left( -\sum_{l=1}^{\infty} A_{N+1,N+l} G^{l-1} \right)^{-1} A_{N+1,N},$$

the matrices  $G_i, i = \overline{0, N-1}$ , are calculated from the backward recursion:

$$G_i = - \left( A_{i+1, i+1} + \sum_{l=i+2}^{\infty} A_{i+1, l} G^{\max\{0, l-N\}} \cdot G_{\min\{N, l\}-1} \cdot G_{\min\{N, l\}-2} \cdot \dots \cdot G_{i+1} \right)^{-1} A_{i+1, i},$$

the vector  $\mathbf{p}_0$  is calculated as the unique solution to the following system of linear algebraic equations:

$$\mathbf{p}_0 \bar{A}_{0,0} = \mathbf{0}, \quad \mathbf{p}_0 \left( \sum_{l=1}^{\infty} F_l \mathbf{e} + \mathbf{e} \right) = 1.$$

#### 4. Performance Measures

Having the probability vector  $\mathbf{p}$  been computed, we are able to calculate performance measures of the considered model. The main performance measure in the case of a finite buffer is the probability  $P_{loss}$  that an arbitrary customer will be lost (the loss probability).

**Theorem 5.** The loss probability  $P_{loss}$  is calculated as follows

(i) in the case of PA discipline

$$P_{loss} = 1 - \frac{1}{\lambda} \sum_{i=0}^{N+L-1} \mathbf{p}_i \sum_{k=0}^{N+L-i} (k+i-N-L) \mathcal{D}_k^{(i)} \mathbf{e}, \quad (10)$$

(ii) in the case of CR discipline

$$P_{loss} = 1 - \frac{1}{\lambda} \sum_{i=0}^{N+L-1} \mathbf{p}_i \sum_{k=0}^{N+L-i} k \mathcal{D}_k^{(i)} \mathbf{e}, \quad (11)$$

(iii) in the case of CA discipline

$$P_{loss} = 1 - \frac{1}{\lambda} \sum_{i=0}^{N+L-1} \mathbf{p}_i \sum_{k=0}^{\infty} k \mathcal{D}_k^{(i)} \mathbf{e}. \quad (12)$$

Proofs of formulae (10) - (12) are analogous. So, we will prove only formula (10). According to a formula of the total probability, the probability  $P_{loss}$  is calculated as

$$P_{loss} = 1 - \sum_{i=0}^{N+L-1} \sum_{k=1}^{\infty} P_k P_i^{(k)} R^{(i,k)} \quad (13)$$

where  $P_k$  is a probability that an arbitrary customer arrives in a batch consisting of  $k$  customers;  $P_i^{(k)}$  is a probability to see  $i$  servers being busy at the epoch of the  $k$  – size batch arrival;  $R^{(i,k)}$  is a probability that an arbitrary customer will not be lost conditional it arrives in a batch consisting of  $k$  customers and  $i$  servers are busy at the arrival epoch.



It can be shown that

$$P_i^{(k)} = \frac{\mathbf{p}_i \mathcal{D}_k^{(i)} \mathbf{e}}{\mathbf{x}_1 \mathcal{D}_k^{(0)} \mathbf{e}}, \quad i = \overline{0, N+L-1}, k \geq 1, \quad (14)$$

$$P_k = \frac{k \mathbf{x}_1 \mathcal{D}_k^{(0)} \mathbf{e}}{\mathbf{x}_1 D(z) \Big|_{z=1} \mathbf{e}} = k \frac{\mathbf{x}_1 \mathcal{D}_k^{(0)} \mathbf{e}}{\lambda}, \quad k \geq 1, \quad (15)$$

$$R^{(i,k)} = \begin{cases} 1, & k \leq N+L-i, \\ \frac{N+L-i}{k}, & k > N+L-i, \quad i = \overline{0, N+L-1}. \end{cases} \quad (16)$$

By substituting (14)-(16) into (13) after some algebra we get (10).  $\square$

Some performance measures for the case  $L = \infty$  are presented below.

- The probability to see  $i$  customers in the system

$$p_i = \mathbf{p}_i \mathbf{e}, \quad i \geq 0;$$

- The mean number of customers in the system

$$L_{queue} = \sum_{i=0}^{\infty} i \mathbf{p}_i \mathbf{e};$$

- The probability to see  $n$  busy servers

$$p_n = \mathbf{p}_n \mathbf{e}, n = \overline{0, N-1}, \quad p_N = \sum_{n=N}^{\infty} \mathbf{p}_n \mathbf{e};$$

- The mean number of busy servers

$$N_{busy} = \sum_{n=1}^N n \mathbf{p}_n \mathbf{e} + N \sum_{n=N}^{\infty} \mathbf{p}_n \mathbf{e};$$

- The mean number  $N_{idle}$  of idle servers

$$N_{idle} = N - N_{busy};$$

- The vector  $\hat{\mathbf{p}}_n$  whose  $(\overline{W}(r-1) + \nu + 1)$ -th entry is the joint probability to see  $n$  busy servers, the random environment in the state  $r$  and the process  $\nu_i$  in the state  $\nu$

$$\hat{\mathbf{p}}_n = \mathbf{p}_n (I_{\overline{RW}} \otimes \mathbf{e}_{M^n}), n = \overline{0, N-1}, \quad \hat{\mathbf{p}}_N = \sum_{n=N}^{\infty} \mathbf{p}_n (I_{\overline{RW}} \otimes \mathbf{e}_{M^n});$$

• The vector of conditional means of the number of busy servers under the fixed states of the random environment

$$\mathbf{n} = \sum_{n=1}^{\infty} \min\{n, N\} \hat{\mathbf{p}}_n (I_R \otimes \mathbf{e}_{\overline{W}}) \text{diag}\{q_r^{-1}, r = \overline{1, R}\};$$

• The vector  $\mathbf{p}^{(a)}(n)$  whose  $(\overline{W}(r-1) + \nu + 1)$ -th entry is the joint probability that an arbitrary arriving call sees  $n$  busy servers and the random environment in the state  $r$  and the state of the process  $\nu_t$  becomes  $\nu$  after the arrival epoch

$$\mathbf{p}^{(a)}(n) = \lambda^{-1} \hat{\mathbf{p}}_n D'(z) \Big|_{z=1}, n = \overline{0, N};$$

• The probability  $p^{(a)}(n)$  that an arbitrary arriving call sees  $n$  busy servers

$$p^{(a)}(n) = \mathbf{p}^{(a)}(n) \mathbf{e}, n = \overline{0, N};$$

• The vector  $\mathbf{p}_b^{(a)}(n)$  whose  $(\overline{W}(r-1) + \nu + 1)$ -th entry is the joint probability that an arbitrary arriving batch of size  $k$  sees  $n$  busy servers and the random environment in the state  $r$  and the state of the process  $\nu_t$  becomes  $\nu$  after the arrival epoch

$$\mathbf{p}_b^{(a)}(k, n) = \lambda_b^{-1} \hat{\mathbf{p}}_n \text{diag}\{D_k^{(r)}, r = \overline{1, R}\}, n = \overline{0, N}, k \geq 1,$$

where  $\lambda_b = \mathbf{x}(D(1) - D(0))\mathbf{e}$ ;

• The probability  $p_b^{(a)}(k, n)$  that an arbitrary arriving batch of size  $k$  sees  $n$  busy servers

$$p_b^{(a)}(k, n) = \mathbf{p}_b^{(a)}(k, n) \mathbf{e}, n = \overline{0, N}, k \geq 1;$$

• The vector  $\mathbf{p}_b^{(a)}(n)$  whose  $(\overline{W}(r-1) + \nu + 1)$ -th entry is the joint probability that an arbitrary arriving batch sees  $n$  busy servers and the random environment in the state  $r$  and the state of the process  $\nu_t$  becomes  $\nu$  after the arrival epoch

$$\mathbf{p}_b^{(a)}(n) = \lambda_b^{-1} \hat{\mathbf{p}}_n (D(1) - D(0)), n = \overline{0, N};$$

• The probability  $p_b^{(a)}(n)$  that an arbitrary arriving batch sees  $n$  busy servers

$$p_b^{(a)}(n) = \mathbf{p}_b^{(a)}(n) \mathbf{e}, n = \overline{0, N};$$

• The probability  $P_{imm}$  that an arbitrary customer will enter the service immediately upon arrival (without visiting a buffer)

$$P_{imm} = \lambda^{-1} \sum_{i=0}^{N-1} \mathbf{p}_i \sum_{k=0}^{N-i} (k+i-N) \mathcal{D}_k^{(i)} \mathbf{e}.$$

## 5. Actual Sojourn Time

Let  $\nu_a(s)$ ,  $\operatorname{Re} s > 0$ , be the Laplace-Stieltjes transform (LST) of the sojourn time distribution and  $\bar{\nu}_a$  be the mean sojourn time of the arbitrary customer in the system.

**Theorem 6.** The Laplace-Stieltjes transform  $\nu_a(s)$  is calculated as follows

$$\nu_a(s) = \frac{1}{\lambda} \left\{ \sum_{i=0}^{N-1} \mathbf{p}_i \sum_{k=1}^{\infty} \min\{k, N-i\} \mathcal{D}_k^{(i)} (I_R \otimes \mathbf{e}_{\overline{Wm}^i}) + \right. \quad (17)$$

$$\left. + \sum_{i=0}^{\infty} \mathbf{p}_i \sum_{k=\max\{1, N-i+1\}}^{\infty} \mathcal{D}_k^{(\min\{i, N\})} \mathcal{B}^{(\max\{0, N-i\})} \sum_{l=\max\{1, N-i+1\}}^k (\mathcal{F}(s))^{i-N+l} (I_R \otimes \mathbf{e}_{M^N}) \right\} \mathcal{H}(s) \mathbf{e}_R,$$

where

$$\begin{aligned} \mathcal{H}(s) &= -\operatorname{diag}\{\boldsymbol{\beta}^{(r)}, r = \overline{1, R}\} (sI - (Q \otimes I_M + \mathcal{S}))^{-1} \operatorname{diag}\{\mathbf{S}_0^{(r)}, r = \overline{1, R}\}, \\ \mathcal{F}(s) &= -\left( sI - \left( Q \otimes I_{M^N} + \operatorname{diag}\left\{ (S^{(r)})^{\oplus N}, r = \overline{1, R} \right\} \right) \right)^{-1} \operatorname{diag}\left\{ (\mathbf{S}_0^{(r)} \boldsymbol{\beta}^{(r)})^{\oplus N}, r = \overline{1, R} \right\}, \\ \mathcal{B}^{(n)} &= \operatorname{diag}\left\{ \mathbf{e}_{\overline{W}} \otimes I_{M^{N-n}} \otimes (\boldsymbol{\beta}^{(r)})^{\oplus n}, r = \overline{1, R} \right\}, \quad n = \overline{0, N}, \\ \mathcal{S} &= \operatorname{diag}\{S^{(r)}, r = \overline{1, R}\}. \end{aligned}$$

**Proof.** We derive the expression for the LST  $\nu_a(s)$  by means of the method of collective marks (method of additional event, method of catastrophes) for references see, e.g. [3], [11]. To this end, we interpret the variable  $s$  as the intensity of some virtual stationary Poisson flow of catastrophes. So,  $\nu_a(s)$  has the meaning of probability that no one catastrophe arrives during the sojourn time of an arbitrary customer. Then, the proof of the theorem follows from the formula of total probability if we analyze the states of the system at an arbitrary customer arrival epoch and take into account the probabilistic meaning of the involved matrices. The matrix  $\mathcal{H}(s)$  is the matrix LST of an arbitrary customer service time distribution. It is the  $R$ -size square matrix whose  $(r, r')$  entry is a probability that during the service time of a customer a catastrophe does not arrive and the process  $r_t, t \geq 0$ , transits from the state  $r$  to the state  $r', r, r' = \overline{1, R}$ . It is defined by the formula:

$$\mathcal{H}(s) = \operatorname{diag}\{\boldsymbol{\beta}^{(r)}, r = \overline{1, R}\} \int_0^{\infty} e^{-st} e^{(Q \otimes I_M + \operatorname{diag}\{S^{(r)}, r = \overline{1, R}\})t} dt \operatorname{diag}\{\mathbf{S}_0^{(r)}, r = \overline{1, R}\}.$$

Analogously, the entries of the matrix LST  $\mathcal{F}(s)$  are the probabilities of no catastrophe arrival and corresponding transitions of the process  $\{r_t, m_t^{(1)}, \dots, m_t^{(N)}\}, t \geq 0$ , during the time

interval from an arbitrary moment when all  $N$  servers are busy till the first epoch when one of these servers finishes the service of a customer. This matrix is defined by the formula:

$$\mathcal{F}(s) = \int_0^\infty e^{-st} e^{(Q \otimes I_{M^N} + \text{diag}\{(S^{(r)})^{\oplus N}, r = \overline{1, R}\})t} dt \text{diag}\{(S_0^{(r)} \beta^{(r)})^{\oplus N}, r = \overline{1, R}\}.$$

**Theorem 7.** The mean sojourn time  $\bar{v}_a$  of an arbitrary customer in the system is calculated by

$$\begin{aligned} \bar{v}_a = & -\frac{1}{\lambda} \left\{ \sum_{i=0}^\infty \mathbf{p}_i \sum_{k=\max\{1, N-i+1\}}^\infty \mathcal{D}_k^{(\min\{i, N\})} \mathcal{B}^{(\max\{0, N-i\})} \sum_{l=\max\{1, N-i+1\}}^k \sum_{m=0}^{i+l-N-1} (\mathcal{F}(0))^m \mathcal{F}'(0) (I_R \otimes \mathbf{e}_{M^N}) + \right. \\ & \left. + \sum_{i=0}^{N-1} \mathbf{p}_i \sum_{k=1}^\infty \min\{k, N-i\} \mathcal{D}_k^{(i)} (I_R \otimes \mathbf{e}_{\overline{WM}^i}) \mathcal{H}'(0) + \right. \\ & \left. + \sum_{i=0}^\infty \mathbf{p}_i \sum_{k=\max\{1, N-i+1\}}^\infty \mathcal{D}_k^{(\min\{i, N\})} \mathcal{B}^{(\max\{0, N-i\})} \sum_{l=\max\{1, N-i+1\}}^k (\mathcal{F}(0))^{i-N+l} (I_R \otimes \mathbf{e}_{M^N}) \mathcal{H}'(0) \right\} \mathbf{e}_R \end{aligned} \quad (18)$$

where

$$\begin{aligned} \mathcal{H}'(0) &= -\text{diag}\{\beta^{(r)}, r = \overline{1, R}\} [Q \otimes I_M + \mathcal{S}]^{-2} \text{diag}\{S_0^{(r)}, r = \overline{1, R}\}, \\ \mathcal{F}'(0) &= -\left[ Q \otimes I_{M^N} + \text{diag}\{(S^{(r)})^{\oplus N}, r = \overline{1, R}\} \right]^{-2} \text{diag}\{(S_0^{(r)} \beta^{(r)})^{\oplus N}, r = \overline{1, R}\}. \end{aligned}$$

Proof. To get expression (18) for  $\bar{v}_a$  we differentiate (17) at the point  $s=0$  and use the formula  $\bar{v}_a = -\bar{v}'_a(0)$ .

## 6. Numerical Examples

The goal of the numerical experiments is to demonstrate the feasibility of the proposed algorithms for computing the stationary distributions of the number of customers and the sojourn time in the system and to give some insight into behavior of the considered queueing systems. In particular, the following issues are addressed:

- Comparison of the mean sojourn time of an arbitrary customer and the probability of immediate access to the servers in the systems with varying traffic intensities and different coefficients of correlation in the *BMAP*s (experiment #1);
- Comparison of the mean sojourn time of an arbitrary customer and the probability of immediate access to the servers in the original system in a *RE* and in more simple queueing systems for different system loads (experiments #2);
- Demonstration of possible positive effect of redistribution of traffic between the peak traffic periods and normal traffic periods (experiment #3);
- Comparison of the exact value of performance measures of the system in a *RE* and their simple engineering approximations in cases of slowly and quickly varying *RE* (experiment #4);
- Investigation of the rate of convergence of the mean sojourn time and the probability of immediate access in the system with the finite buffer to corresponding

performance measures of the system with an infinite buffer when the buffer size increases (experiment #5);

- Demonstration of the possibility to apply the presented results for optimization of the number of servers in the system (experiment #6).

In numerical examples, we consider the systems operating in the *RE* which has two states ( $R = 2$ ). The generator of the random environment is  $Q = \begin{pmatrix} -5 & 5 \\ 15 & -15 \end{pmatrix}$ . The stationary distribution of the *RE* states is defined by the vector  $\mathbf{q} = (0.75, 0.25)$ . The number of servers is  $N = 3$ .

In the presented examples, we will use several different *MAP* s and *BMAP* s for description of the arrival process and two *PH* type distributions for description of the service processes under the fixed value of the *RE*. For the use in the sequel, let us define these processes.

We consider four arrival processes *MAP*<sub>*r*</sub>,  $r = \overline{1, 4}$ . *MAP*<sub>*r*</sub> is defined by the matrices  $D_0^{(r)}$ ,  $D_1^{(r)}$ ,  $r = \overline{1, 4}$ , where

$$\begin{aligned} D_0^{(1)} &= \begin{pmatrix} -3.9 & 0.15 & 0.15 \\ 0.13 & -0.6 & 0.1 \\ 0.15 & 0.14 & -0.5 \end{pmatrix}, & D_1^{(1)} &= \begin{pmatrix} 3.5 & 0.08 & 0.02 \\ 0.03 & 0.3 & 0.04 \\ 0.02 & 0.06 & 0.13 \end{pmatrix}; \\ D_0^{(2)} &= \begin{pmatrix} -6.4 & 0.1 & 0.1 \\ 0.04 & -0.6 & 0.1 \\ 0.07 & 0.1 & -0.44 \end{pmatrix}, & D_1^{(2)} &= \begin{pmatrix} 6.06 & 0.12 & 0.02 \\ 0.01 & 0.4 & 0.05 \\ 0.01 & 0.06 & 0.2 \end{pmatrix}; \\ D_0^{(3)} &= \begin{pmatrix} -2.9 & 0.73 & 0.77 \\ 0.87 & -3.06 & 0.53 \\ 0.85 & 0.5 & -2 \end{pmatrix}, & D_1^{(3)} &= \begin{pmatrix} 0.68 & 0.45 & 0.27 \\ 0.48 & 1.08 & 0.1 \\ 0.35 & 0.05 & 0.25 \end{pmatrix}; \\ D_0^{(4)} &= \begin{pmatrix} -1.3 & 0.21 & 0.17 \\ 0.16 & -2.04 & 0.21 \\ 0.01 & 0.16 & -1.3 \end{pmatrix}, & D_1^{(4)} &= \begin{pmatrix} 0.46 & 0.32 & 0.14 \\ 0.13 & 1.34 & 0.2 \\ 0.02 & 0.01 & 1.1 \end{pmatrix}. \end{aligned}$$

All these *MAP* s have fundamental rate  $\lambda^{(r)} = 1.25$ . The *MAP*<sub>1</sub> has the squared variation coefficient  $(c_{var}^{(1)})^2 = 4$  and the coefficient of correlation of the lengths of successive inter-arrival times  $c_{cor}^{(1)} = 0.2$ . For the rest of the *MAP* s, the corresponding parameters are:  $(c_{var}^{(2)})^2 = 4$ ,  $c_{cor}^{(2)} = 0.3$ ;  $(c_{var}^{(3)})^2 = 1.097$ ,  $c_{cor}^{(3)} = 0.0052$ ;  $(c_{var}^{(4)})^2 = 1.037$ ,  $c_{cor}^{(4)} = 0.0065$ .

Based on these *MAP* s, we construct batch flows *BMAP* s as follows. If the *MAP* is defined by the matrices  $D_0^{(r)}$ , and  $D_1^{(r)}$ ,  $r = \overline{1, 4}$ , then the *BMAP* having the maximal size of a batch equal to  $K$  is defined by the matrices  $D_0^{(r)}$ ,  $D_k^{(r)} = D_1^{(r)} q^{k-1} (1 - q) / (1 - q^K)$ ,  $k = \overline{1, K}$ ,  $r = \overline{1, R}$ , where  $q = 0.9$ .

Following this way, we construct the  $BMAP_1$ ,  $BMAP_2$ ,  $BMAP_3$ ,  $BMAP_4$  flows based on the  $MAP_1$ ,  $MAP_2$ ,  $MAP_3$ ,  $MAP_4$  correspondingly, with  $K = 5$ . Note that the coefficients of variation and correlation of all  $BMAP$ s are the same as these coefficients for the corresponding  $MAP$ s. Fundamental rate  $\lambda^{(r)}$  and the mean batch size  $\bar{k}^{(r)}$  of the  $BMAP$ s are the following:  $\lambda^{(1)} = \lambda^{(2)} = \lambda^{(3)} = \lambda^{(4)} = 3.488$ ,  $\bar{k}^{(1)} = \bar{k}^{(2)} = \bar{k}^{(3)} = \bar{k}^{(4)} = 1.989$ .

The  $PH_r$ ,  $r = \overline{1, 2}$ , service processes are defined by the vectors  $\beta^{(1)} = (1, 0)$ ,  $\beta^{(2)} = (0.2, 0.8)$  and the matrices

$$S^{(1)} = \begin{pmatrix} -4 & 4 \\ 0 & -4 \end{pmatrix}, \quad S^{(2)} = \begin{pmatrix} -10 & 2 \\ 2 & -20 \end{pmatrix}.$$

The mean rates of service are  $\mu^{(1)} = 2$ ,  $\mu^{(2)} = 14$ . The coefficients of variation of the service time distribution are defined by  $(c_{var}^{(1)})^2 = 0.5$ ,  $(c_{var}^{(2)})^2 = 1.24$ .

In the **first experiment**, we compare the dependence of  $\bar{v}_a$  and  $P_{imm}$  on the system load  $\rho$  for the  $BMAP$ s with different correlations.

In the experiment we use service processes defined by  $PH_1$  and  $PH_2$  and four different input flows which are described by  $BMAP_1$ ,  $BMAP_2$ ,  $BMAP_3$  and  $BMAP_4$  having the same mean fundamental rate equal to 3.488 but different correlation coefficients.

We consider three queueing systems which have different combinations of the  $BMAP$ s under the first and second states of the  $RE$ .

The input flow in the first system is defined by  $BMAP_1$  and  $BMAP_2$ . These  $BMAP$ s have large coefficients of correlation  $c_{cor}^{(1)} = 0.2$  and  $c_{cor}^{(2)} = 0.3$ .

The input flow in the second system is defined by  $BMAP_3$  and  $BMAP_4$ . These  $BMAP$ s have small coefficients of correlation  $c_{cor}^{(3)} = 0.0052$  and  $c_{cor}^{(4)} = 0.0065$ .

In the third system the input is defined by  $BMAP_1$  and  $BMAP_4$ . The correlation coefficients of these  $BMAP$ s differ significantly.

Figures 1 and 2 show the dependence of the mean sojourn time  $\bar{v}_a$  and the probability  $P_{imm}$  on the system load  $\rho$  for all these systems. Variation of the value of  $\rho$  in all experiments is performed by means of multiplying the entries of the matrices, which define the corresponding  $BMAP$ , by some varying factor  $\gamma$ . This implies the increase of the fundamental rate of all the  $BMAP$  by a factor  $\gamma$ . Service time distributions are not modified. It is clear from Figure 1 that correlation in  $BMAP$  has a great impact on the sojourn time in the system. An increase of correlation at least in one of the  $BMAP$ s describing input in the system implies an increase of the sojourn time in the system in all range of the system load.



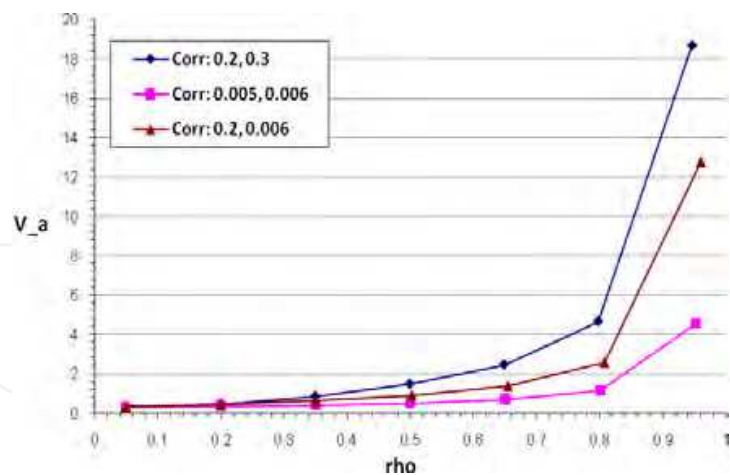


Fig. 1. Mean sojourn time in the system as a function of the system load for different correlations in the *BMAP* s

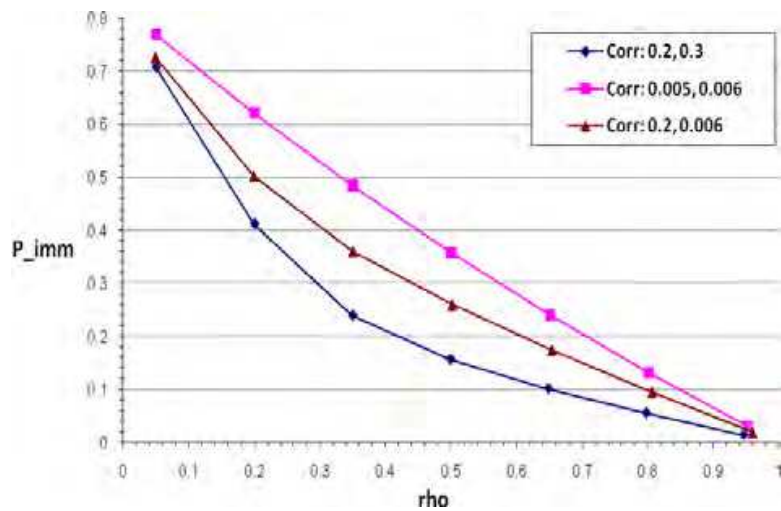


Fig. 2. Probability of immediate access to the servers in the system as a function of the system load for different correlations in the *BMAP* s

In the **second experiment** we compare the values  $\bar{v}_a$  and  $P_{imm}$  in the *BMAP*/*PH*/*N* system operating in the *RE* and in more simple queueing systems which can be considered as its simplified analogs. The first type analog is the  $M^x$ /*PH*/*N* system in the *RE* where, under the fixed value of the *RE*, the input flow is a group stationary Poisson with the same batch size distribution and intensity equal to fundamental rate of the corresponding *BMAP* in the original system. The second type analog is the system  $M^x$ /*M*/*N* with parameters of arrival and service processes which are obtained by means of averaging, according to stationary distribution of the *RE*, parameters of the original system.

Input flow is described by *BMAP*<sub>1</sub> and *BMAP*<sub>2</sub>. Service processes are *PH*<sub>1</sub> and *PH*<sub>2</sub>.

Figures 3 and 4 show the dependence of the the mean sojourn time  $\bar{v}_a$  and the probability  $P_{imm}$  on the value of  $\rho$ .

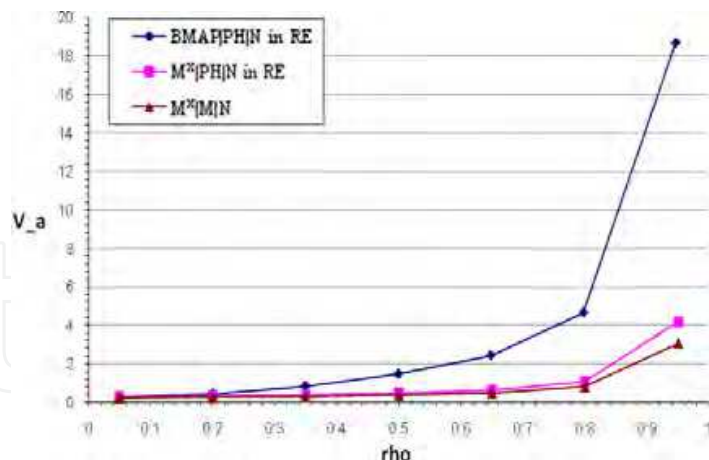


Fig. 3. Mean sojourn time in original system and more simple queueing systems

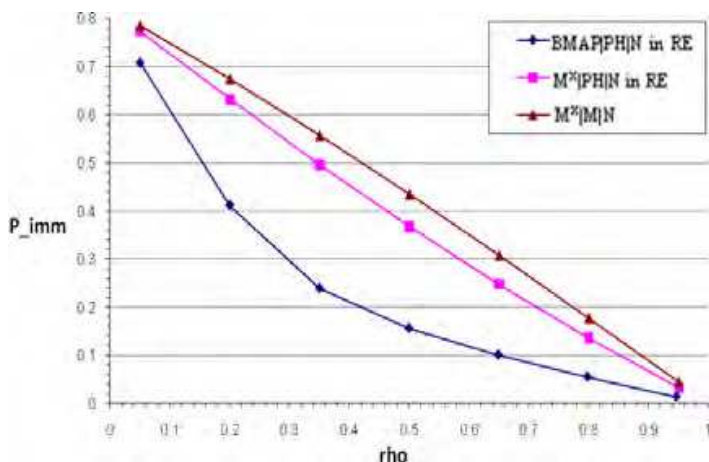


Fig. 4. Probability of immediate access to the servers in original system and more simple queueing systems

It can be seen from *Figures 3 and 4* that an approximation of the mean sojourn time and the probability that an arbitrary call reaches the server immediately by means of their values in some specially constructed more simple queueing system can be rather bad.

The idea of the **third experiment** is the following. Let us assume that the *RE* has two states. One state corresponds to the peak traffic periods, the second one corresponds to the normal traffic periods. Service times during these periods are defined by  $PH_1$  and  $PH_2$  distributions. Arrivals during these periods are defined by the stationary Poisson flow with the rates  $\lambda_1$  and  $\lambda_2$  correspondingly and initially we assume that  $\lambda_1 \gg \lambda_2$ . It is intuitively clear that if it is possible to redistribute the arrival processes (i.e., to reduce the arrival rate during the peak periods and to increase it correspondingly during the normal traffic periods) without changing the total average arrival rate, the mean sojourn time in the system can be reduced. In real life system such a redistribution is sometimes possible, e.g., by means of controlling tariffs during the peak traffic periods. The goal of this experiment is to show that this intuitive consideration is correct and to illustrate the effect of the redistribution.

We assume that the averaged arrival rate  $\lambda$  should be 12.5 and consider four different situations: a huge difference of arrival rates  $\lambda_1 = 50\lambda_2$ , a very big difference  $\lambda_1 = 10\lambda_2$ , a big difference  $\lambda_1 = 3\lambda_2$  and equal arrival rates  $\lambda_1 = \lambda_2$ . The generator of the random environment is  $Q = \begin{pmatrix} -15 & 15 \\ 5 & -5 \end{pmatrix}$ .

It can be seen from *Figures 5 and 6* that the smoothing of the peak rates can cause essential decrease of the mean sojourn time and the increase of the probability that an arbitrary call reaches the server immediately upon arrival in the system.

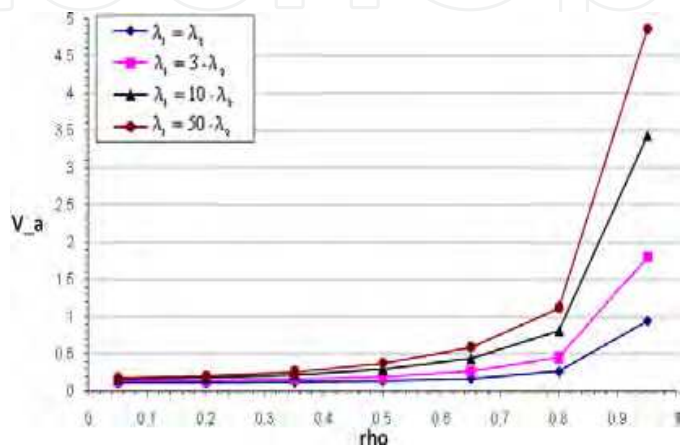


Fig. 5. Mean sojourn time as a function of system load for different relations of arrival rates

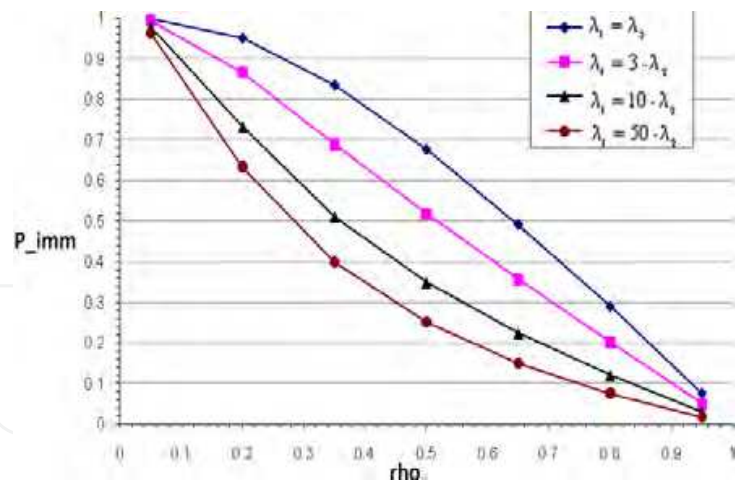


Fig. 6. Probability of immediate access to the servers as a function of system load for different relations of arrival rates

In the second experiment, we have seen that an approximation of the system performance measures by means of their values in more simple queueing system can be bad. However, it is intuitively clear the following. If the random environment is "very slow" (the rate of the RE is much less then the rates of the input flow and the service processes), an approximation called below as "mixed system" can be applied successfully. This approximation consists of calculation of the system characteristics under the fixed states of

the  $RE$  and their averaging by the  $RE$  distribution. If the random environment is "very fast", approximation called below as "mixed parameters" can be successfully applied. This approximation consists of averaging parameters of the arrival and service processes by the distribution of the  $RE$  and calculation of performance measures in  $BMAP/PH/N$  system with the averaged arrival and service rates.

In the **fourth experiment**, we show numerically that sometimes the described approximations make sense. However, in situations when environment is neither "very slow" nor "very fast", these approximations can be very poor. We consider the  $RE$ s with

different rate which are characterized by the generators of the form  $Q^{(k)} = \begin{pmatrix} -5 & 5 \\ 15 & -15 \end{pmatrix} \cdot 10^k$ .

We vary the parameter  $k$  from -7 to 4 what corresponds to the variation of the  $RE$  rate from "very slow" to "very fast". In this and further experiments, the input flow is described by the  $BMAP_1$  and  $BMAP_2$  and the service process is defined by the  $PH_1$  and  $PH_2$ . The results are presented in *Figures 7, 8 and 9*. In application of "mixed system" approximation, the averaged arrival rates under both states of the  $RE$  are equal to 3.488. The averaged service rate is equal to 2 at the first state of the  $RE$  and is equal to 14 at the second state. The mean sojourn times of an arbitrary customer at these states are equal to 3.5297 and 0.0998, respectively; the probabilities of immediate access to the servers are equal to 0.2021 and 0.81399; the mean numbers of customers in the system are equal to 12.311 and 0.3482. The averaged, according to the stationary distribution of the  $RE$ , mean sojourn time of an arbitrary customer is equal to 2.6722 and the probability that an arbitrary arriving customer sees an idle server is equal to 0.355. In application of "mixed parameters" approximation, the averaged, according to the stationary distribution of the  $RE$ , arrival rate is equal to 3.778 while the averaged service rate is equal to 4.0625. The value of the mean sojourn time of an arbitrary customer in the system with averaged arrival and service rates is equal to 0.7541, the probability of immediate access to the servers is equal to 0.4168, the mean number of customers in the system is equal to 2.849.

*Figures 7, 8 and 9* confirm the hypothesis that the first type approximation ("mixed system") is good in case of "very slow"  $RE$  and the second one ("mixed parameters") can be applied to case of "very fast"  $RE$ . But sometimes the second type approximation is not very good (see *Figures 8*) because it is not quite clear how to make averaging of service intensity. Simple averaging of service rates under the different states of the  $RE$  may be not correct when the load of the system is not high because there are time intervals when the system is empty and no service is provided. It is worth to note also that there is an interval for  $RE$  rate (interval  $k \in [-3, 0]$ ) where one should not use the values of the system performance measures calculated based on the considered approximating models. The use of these values can lead to the large relative error. Thus, *Figures 7, 8 and 9* confirm the importance of investigation implemented in this chapter. Simple engineering approximations can lead to unsatisfactory performance evaluation and capacity planning in real world systems.

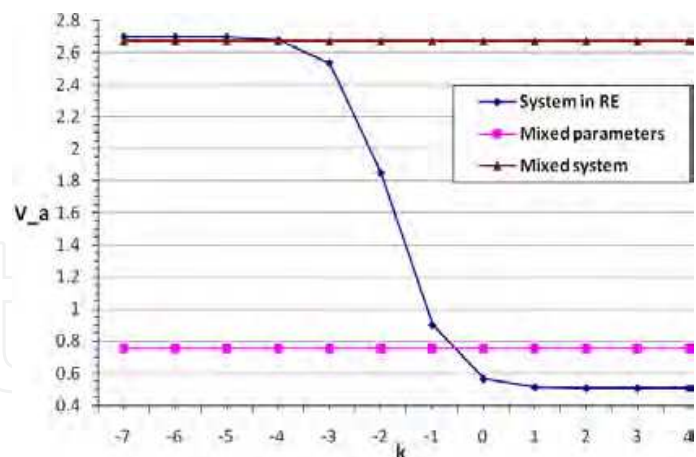


Fig. 7. Mean sojourn time of an arbitrary customer as a function of the *RE* rate

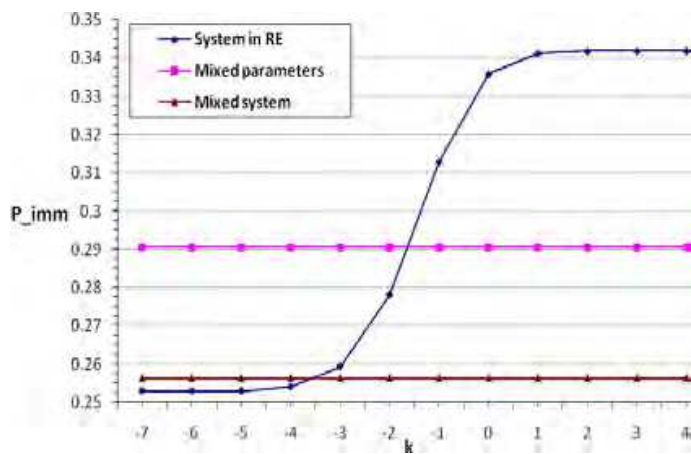


Fig. 8. Probability of immediate access to the servers as a function of the *RE* rate

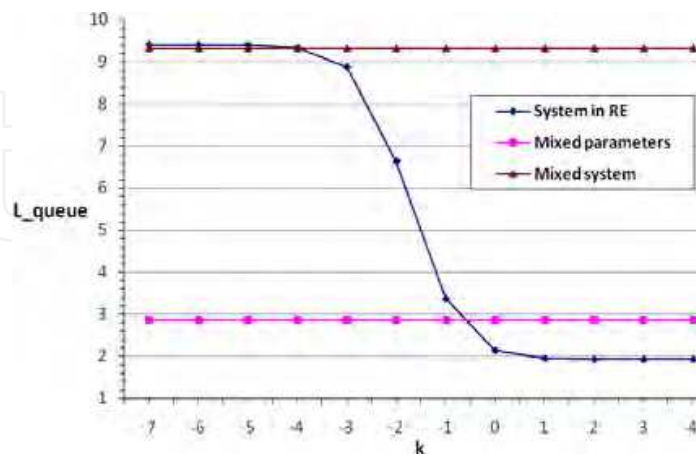


Fig. 9. The mean number of customers in the system  $L_{queue}$  as a function of the *RE* rate

In the **fifth experiment** we compare the mean number of customers, probability of immediate access to the servers and loss probability in the *BMAP/PH/N* and



$BMAP/PH/N/L$  systems operating in the  $RE$  for different values  $L$  of the buffer capacity and different customers admission discipline.

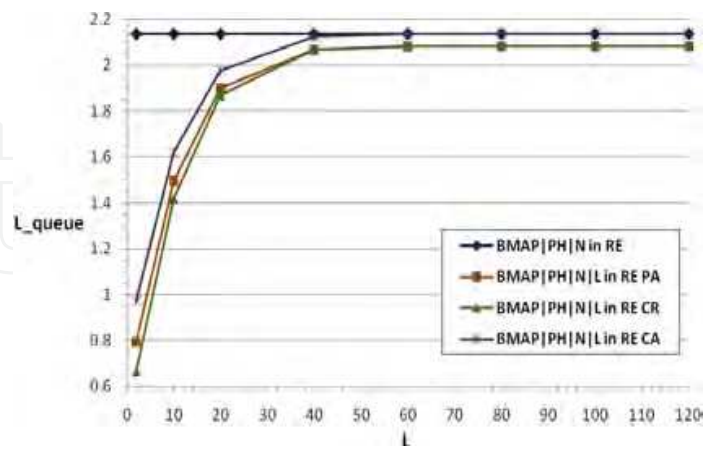


Fig. 10. Mean number of customers in the system as a function of the buffer capacity  $L$

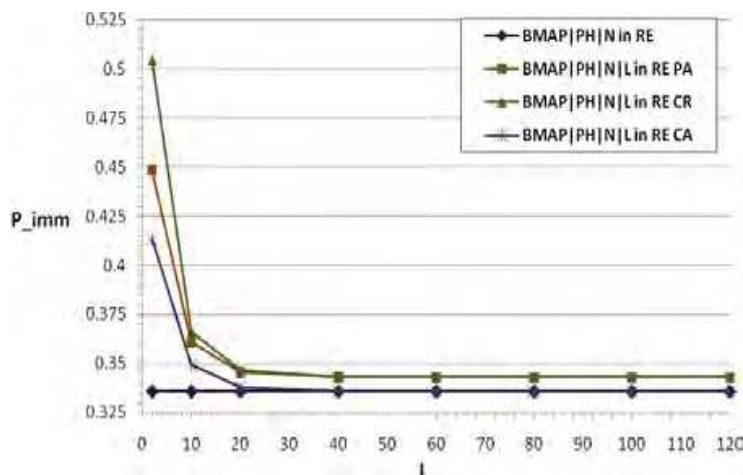


Fig. 11. Probability of immediate access to the servers as a function of the buffer capacity  $L$

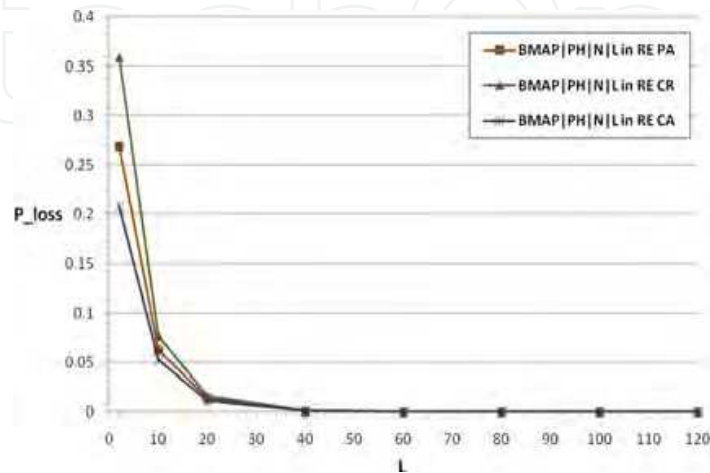


Fig. 12. Loss probability as a function of the buffer capacity  $L$



Looking at *Figures 10-12*, it should be noted that the rate of convergence of the curves corresponding to the disciplines *PA* and *CR* to their limits defined by the system with an infinite buffer is not very high. When we further increase the value  $L$ , we discover that even for the buffer capacity  $L$  about 5000, the difference is not negligible. So, estimation of performance measures of the system with an infinite buffer by the respective measures of the system with a finite buffer can be not very good. This explains why we made the separate analysis of the system with an infinite buffer.

Finally, in the **sixth experiment** we consider the next optimization problem:

$$J(N) = c_1 \bar{v}_a + c_2 N \rightarrow \max_N, \quad (19)$$

where

- $\bar{v}_a$  is the mean sojourn time in the system,
- $N$  is the number of servers,
- $c_1$  is the charge for an unit of customer sojourn time in the system,
- $c_2$  is the cost of a server maintenance per unit of time.

It is clear that this problem is not trivial. When the number of servers is small, the cost of servers maintenance is also small, but the mean sojourn time is large. If we increase the number of servers, the mean sojourn time decreases while the cost of servers maintenance increases.

Let us assume that the cost coefficients be fixed as  $c_1 = 5$  and  $c_2 = 3$ . Service time distribution at both states of the *RE* is exponential with intensities  $\mu^{(1)} = 1$ ,  $\mu^{(2)} = 7$ .

On *Figure 13*, dependence of the cost criterion  $J(N)$  on  $N$  is presented along with the dependences of the summands  $c_1 \bar{v}_a$  and  $c_2 N$ .

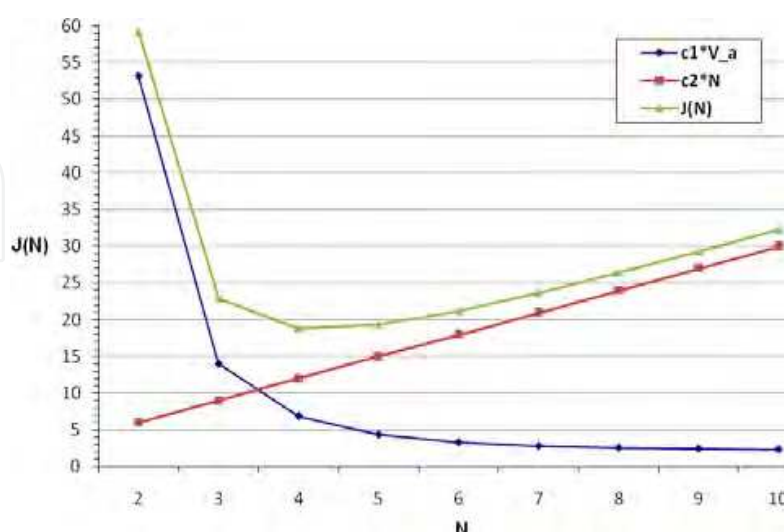


Fig. 13. Criterion  $J(N)$  as a function of number of servers in the system

Based on *Figure 13*, one can conclude that our analysis allows effectively solve the problems of the system design and that the optimal value of the cost criterion (in this example it is provided by  $N = 4$ ) can be significantly smaller than the values of the cost criterion for other values on  $N$ .

## 7. Conclusion

The  $BMAP/PH/N/L$  system operating in a finite state space Markovian random environment is investigated for the finite and infinite buffer capacity. The joint stationary distribution of the number of the customers in the system, the state of the random environment, and the states of the underlying processes of arrival and service processes is calculated. The analytic formulas for performance measures of the system are derived. The Laplace-Stieltjes transform of sojourn time distribution is derived and the mean sojourn time is calculated. Selected results of numerical study are presented. They show an impact of the correlation in arrival process, illustrate the poor quality of the system characteristics approximation by means of more simple models, confirm the positive effect of the traffic redistribution between the peak and normal operation periods. The results can be used for the optimal design, capacity planning, and performance evaluation of real world systems in which operation of the system can be changed depending on some external factors.

## Acknowledgement

This work was supported by the Korea Research Foundation Grant Funded by the Korean Government (MOEHRD)(KRF-2008-313-D01211).

## 8. References

- Combe, M. (1994) Queueing models with dependence structure. *Amsterdam: CWI*, p. 165.
- Kim, C.S., Klimenok, V.I., Orlovsky, D.S., Dudin, A.N. (2005) Lack of invariant property of the Erlang loss model in case of MAP input. *Queueing Systems*, Vol. 49, pp. 187-213.
- Kasten, H., Runnenburg, J.Th. (1956) Priority in waiting line problems. *Mathematisch Centrum, Amsterdam, Holland*. Dec. 1956.
- Kim, C.S., Dudin, A.N, Klimenok, V.I., Khramova, V.V. (2009) Erlang loss queueing system with batch arrivals operating in a random environment. *Computers and Operations Research*, Vol. 36, pp. 674-967.
- Klimenok, V.I. (1999) Characteristics calculation for multi-server queue with losses and bursty traffic. *Automatic Control and Computer Sciences*, Vol. 12, pp. 393-415.
- Klimenok, V.I., Dudin, A.N. (2006) Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory. *Queueing Systems*, Vol. 54 pp. 245-259.
- Lucantoni, D. (1991) New results on the single server queue with a batch Markovian arrival process. *Comm. Stat. Stochastic Models*, Vol. 7: pp. 1-46.
- Nuets, M.F. (1981) Matrix-Geometric Solutions in Stochastic Models. *John Hopkins Univ. Press, Baltimore*, 1981.
- Neuts, M.F. (1989) Structured Stochastic Matrices of  $M/G/1$  Type and Their Applications. *Marcel Dekker, New York*, 1989.

- Sevastjanov, B.A. (1959) Erlang formula in telephone systems under the arbitrary distribution of conversations duration. *In: Proc. Of 3rd All-Union Mathematical Congress (Academy of Science, Moscow), 1959, Vol. 4, pp. 68-70.*
- Van Danzig, D. (1955) Chaines de Markof dans les ensembles abstraits applications aux processus avec regions absorbantes et au problem des boucles. *Ann. De l'Inst. H. Poincare 14 (fasc. 3), pp. 145-199.*

IntechOpen

IntechOpen



## **Trends in Telecommunications Technologies**

Edited by Christos J Bouras

ISBN 978-953-307-072-8

Hard cover, 768 pages

**Publisher** InTech

**Published online** 01, March, 2010

**Published in print edition** March, 2010

The main focus of the book is the advances in telecommunications modeling, policy, and technology. In particular, several chapters of the book deal with low-level network layers and present issues in optical communication technology and optical networks, including the deployment of optical hardware devices and the design of optical network architecture. Wireless networking is also covered, with a focus on WiFi and WiMAX technologies. The book also contains chapters that deal with transport issues, and namely protocols and policies for efficient and guaranteed transmission characteristics while transferring demanding data applications such as video. Finally, the book includes chapters that focus on the delivery of applications through common telecommunication channels such as the earth atmosphere. This book is useful for researchers working in the telecommunications field, in order to read a compact gathering of some of the latest efforts in related areas. It is also useful for educators that wish to get an up-to-date glimpse of telecommunications research and present it in an easily understandable and concise way. It is finally suitable for the engineers and other interested people that would benefit from an overview of ideas, experiments, algorithms and techniques that are presented throughout the book.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Che Soong Kim, Alexander Dudin, Valentina Klimenok, and Valentina Khramova (2010). Performance Analysis of Multi-Server Queueing System Operating under Control of a Random Environment, Trends in Telecommunications Technologies, Christos J Bouras (Ed.), ISBN: 978-953-307-072-8, InTech, Available from: <http://www.intechopen.com/books/trends-in-telecommunications-technologies/Performance-analysis-of-multi-server-queueing-system-operating-under-control-of-a-random-environment>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen