Journal of the Society of Korea Industrial and Systems Engineering
Vol, 30, No. 4, pp.39－45, December 2007.

통신망 트래픽 제어를 위한 BMAP/M/N/0 대기행렬모형 분석

이석준* · 김제숭**†

*상지대학교 경영학과
**상지대학교 산업공학과

Analysis of BMAP/M/N/0 Queueing System for Telecommunication Network Traffic Control

Seokjun Lee* · Chesoong Kim**†

*School of Business Administration, Sangji University
**Department of Industrial Engineering, Sangji University

The *BMAP/M/N/0* queueing system operating in Markovian random environment is investigated. The stationary distribution of the system is derived. Loss probability and other performance measures of the system also are calculated. Numerical experiments which show the necessity of taking into account the influence of random environment and correlation in input flow are presented.

Keywords : *BMAP/M/N/0* Queueing Model, Stationary State Distribution, Loss Probability

1. Introduction

Various queueing systems have been studied for traffic control to support traffic streams with different traffic characteristics in telecommunication networks. Queueing systems suit for description of a variety of real-life processes, in particular, description of operation of telecommunication networks and they have got a lot of attention in probabilistic literature. Important class of queueing systems assumes that customers arrive into the system in batches. It is usually assumed that, at a batch arrival epoch, all customers of this batch arrive into the system simultaneously.

A. K. Erlang founded queueing theory in the early 1900th. Since that time, well known Erlang's *B*-formula for loss probability in the *M/M/N/0* queueing system, i.e. *N* server queueing system without waiting space, with stationary Poisson arrival process and exponential service time distribution, provided a good mathematical tool for capacity planning and performance evaluation in the classic telephone networks. This is explained by the facts that ① the flows of information in such networks were well described by the stationary Poisson arrival process (stationary ordinary arrival process with no after effect) and ② distribution of the number of busy servers in the *M/G/N/0* system (and loss probability correspondingly) is insensitive with respect to the service time distribution under the fixed mean service time. The latter fact was proven by B. A. Sevastjanov in [10].

† 교신저자 dowoo@sangji.ac.kr
※ This research has been supported by Sangji University Research Fund 2007. This work was supported by grant No. R01-2006-000-10668-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

So, assumption by A. K. Erlang that service time distribution is exponential (what is not true in real life systems) was not rough.

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 arrival process, the modern real life flows are non-stationary, group and correlated. The *BMAP* (Batch Markovian Arrival Process) is one of the appropriate models of such flows. The *BMAP* was introduced as a versatile Markovian point process (*VMPP*) by M. F. Neuts in the 1970th.

Note that although Erlang's *B*-formula is very simple, to facilitate its use by practical engineers, special books were published which contained tables giving the value of loss probability for different values of the traffic intensity and different number of servers. Traffic intensity is defined as ratio of a rate of stationary Poisson arrival process and a service rate. In the case when the arrival process is the *BMAP* ; arrival process is described by a bunch of parameters, so it is not possible to have a set of tables which allows to calculate, in advance, loss probability for all *BMAP*s. Instead, it is required development of computer programs which can allow quickly calculate loss probability for any fixed *BMAP*: Note, that the problem of constructing the *BMAP*; which fits well real arrival process, is not very simple. However, this problem has practical importance and is intensively solving. For relevant references and fitting algorithms see, e.g., [11].

The Erlang loss model for the case of BMAP input was investigated in [5]. The essential extension of results to the case of *PH* (Phase type) service time distribution was considered in [4] where numerically stable algorithms for calculating the stationary distribution of the number of customers in the system are developed. These algorithms are realized as computer software. Numerical results, which are obtained by use of this software and presented in [4], evidently show that loss probability in the case of correlated bursty *BMAP* input is sensitive with respect to the service time distribution and also varies essentially for different correlation and variation of inter-arrival times under the fixed value of the mean arrival rate. This implies that simple Erlang's *B*-formula is inappropriate for capacity planning and performance evaluation in the modern telephone networks and algorithms from [4] should be used for this purpose. However, even the involved *BMAP/PH/N/0* system 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 effect of correlation in 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 these *BMAP* arrival process and *PH* service process are assumed to be stationary within the borders of the model analyzed in [4]. While in many real-life systems the input and service processes are not absolutely stable. 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 modelled by means of the *BMAP/PH/N/0* 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 : ① 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); ② changing the mean arrival rate due breakdowns, overflow or underflow of alternative information transmission channels. Thus, to get the mathematical tool for adequate modelling such information transmission channels, more complicated queues than the *BMAP/PH/N/0* queueing system investigated in [4] 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* as the model of arrival process and *PH* as the model of the service process, must take into account the effect of influence of the random external factors. In some extent, it can be done by means of analyzing the models of queues operating in a random environment.

Speaking about the queue in the *random environment* (*RE*), we assume that there are a queueing system and an external finite state space stochastic process called *RE*. Under the fixed state of the *RE*, queueing system operates as a classic queueing system of the correspondent type. However, when the *RE* jumps into another state, the parameters of the queueing system (inter-arrival times distribution or arrival rate, service times distribution or service rate, number of servers, retrial rate, etc) can immediately change their values. History of investigation of queues operating in a random environment can be traced back to the book by B.V. Gnedenko and I. N. Kovalenko where the *M/G/1* type queueing model with so called partial failures was investigated.

States of the random environment correspond to levels of non-operability of the server. The problem of calculating distribution of the total length of customers in the system was considered. In [12], the problem of calculating the stationary distribution of the queue in the *M/M/*1 system was under study. Essential role in investigation of queues operating in a random environment was played by M. Neuts who investigated the *M/M/*1 type queue operating in a random environment, see [8] and the *M/SM/*1 type queue, see [9], by means of the matrix analytic methods. As the simplest case of queues operating in a random environment, Queues with unreliable servers can be mentioned as the simplest case of queues operating in a random environment. Queues with arrival process modulated by some stochastic process (e.g., the Batch Markov Arrival Process mentioned above, its important partial cases such as Markov Modulated Poisson Process, see [2], Interrupted Poisson Process, Switched Poisson Process, and more general heterogeneous Marked Markov Arrival Process, see [3], and Semi-Markov Arrival Process) as well as queues with service process modulated by some stochastic process (e.g, Semi-Markov Service Process, Markov Service Process, see [1], Phase-type Service Process) can be also considered as important special cases of queues operating in a *RE*.

Importance of investigation of queues operating in a *RE* drastically increased in the last years due to the following reason. Flows of information in the modern communication networks are essentially heterogeneous. Some types of information are very sensitive with respect to delay and jitter but tolerant with respect to losses. Another ones are tolerant with respect to delay but very sensitive with respect to loss of packets. So, different schemes of dynamic bandwidth sharing among these types exist and are developing. They assume that, in case of congestion, transmission of delay tolerant flows is temporarily postponed to provide better conditions for transmission of delay sensitive flows.

Analysis of such schemes requires probabilistic analysis of multi-dimensional processes describing transmission process of different flows. This analysis is often impossible due to mathematical complexity. In such a case, it is reasonable to decompose simultaneous consideration of all flows to separate analysis of processes of transmission of delay sensitive and delay tolerant flows. To this end, we model transmission of delay sensitive flows in terms of queues with controlled service or (and) arrival rate where the service or arrival rate can be changed depending on the queue length or waiting time. Redistribution of the bandwidth to avoid congestion for delay sensitive flows causes variation, at random moments, of available bandwidth for delay tolerant flows. Correspondingly, queues operating in a *RE* naturally arise as the mathematical model for delay tolerant flows transmission. Overwhelming majority of the existing papers is devoted to the investigation of systems operating in Markovian *RE*. Analysis becomes more difficult if the *RE* is defined by semi-Markovian process.

Note, that two kinds of a random environment are considered in literature. More popular is the *RE* which can be called as asynchronous *RE*. Such the *RE* makes its transitions absolutely independent on the state of the queueing system which is governed by this *RE*. Another type, which can be called as the synchronous *RE*, assumes that the *RE* makes its transitions independently on the state of the queueing system, but only at some moments synchronized with moments of customers departure or (and) arrival in the queueing system.

In this paper we extend *BMAP/M/N/0* model assuming that the system has *R* different modes of operations, and the modes are switched by an external random process, so called random environment. The considered queueing model has a wide range of potential applications because in practical systems the input and service processes are not absolutely stable, they are influenced by 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, etc.

The rest of the papers is organized as follows. Mathematical model is formulated in section 2. Section 3 contains analysis of the stationary distribution of the system states and performance measures of the system are given in section 4. Section 5 contains results of numerical experiments and their short analysis.

2. Mathematical Model

We consider an *N*-server queueing system. The behavior of the system 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 $\{1, \cdots R\}$, $R \geq 2$and the infinitesimal generator Q. The input flow into the system is the following modification of the well-known (see, e.g., [7]) *BMAP*. In this input flow, the batch arrivals are directed by the process r_t, $t \geq 0$ (the directing process) with the state space $\{0, 1, \cdots W\}$. Under the fixed state *r* of the random environment, this process behaves as an irreducible continuous time Markov chain. Transitions of the chain r_t, $t \geq 0$, which are not accompanied by arrival, are described by the matrix $D_0^{(r)}$, and transitions, which are accompanied by arrival of k-size batch, are described by the matrix $D_0^{(r)}$, $k \geq 1$, $r=\overline{1,R}$ 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 $\lambda^{(r)}=\theta^{(r)}(D^{(r)}(z))'|_{z=1}e$ and the intensity $\lambda^{(r)}$ of batch arrivals is defined as $\lambda^{(r)}=\theta^{(r)}(-D_0^{(r)})e$ Here $\theta^{(r)}$ is the solution to the equations $\theta^{(r)}D^{(r)}(1)=0$, $\theta^{(r)}e=1$ 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 $(c_{var}^{(r)})^2=\lambda^{(r)}\theta^{(r)}(-D_0^{(r)})^{-1}e-1$ while the correlation coefficient $c_{cor}^{(r)}$ of intervals between successive batch arrivals is calculated as $c_{cor}^{(r)}=(\lambda^{(r)}\theta^{(r)}(-D_0^{(r)})^{-1}(D^{(r)}(1)-D_0^{(r)})(-D_0^{(r)})^{-1}e-1)/(c_{var}^{(r)})^2$The state of the process v_t, $t \geq 0$, is not changed at the epochs of the process r_t, $t \geq 0$, transitions.

The system under consideration has no waiting space. So, if the system has all servers being busy at a batch arrival epoch, the batch leaves the system forever and considered to be lost. If there are free servers at arrival epoch, however the number of these servers is less than the number of customers in the group so called partial admission discipline is used. It means that only a part of the group corresponding to a number of free servers is allowed to enter the service while the rest of the group is lost. It is assumed that all servers are identical and operate independently of each other. Service time of a customer by a server has an exponential distribution of intensity $\mu^{(r)}$ under the state r of the random environment. Our aim is to calculate the stationary state distribution and main performance measures of the described queueing model.

3. Stationary State Distribution

It is easy to see that operation of the considered queueing model is described in terms of the irreducible continuous-time Markov chain $\xi_t=\{i_t, r_t, v_t\}$, $t \geq 0$, where i_t is the number of customers in the system (the number of busy servers), r_t is the state of random environment, $r_t=\overline{1,R}$, and v_t is the state of the *BMAP* directing process at the epoch t, $t \geq 0$. Enumerate the states of the chain ξ_t, $t \geq 0$ in the lexicographic order and form the row vectors P_i, $i=\overline{0,N}$ of probabilities corresponding to the state i of the first component of the process ξ_t, $t \geq 0$. Denote also $p=(p_0, \cdots p_N)$.

The vector p satisfies the system of linear algebraic equations of the form :

$$pA=0, \ pe=1 \tag{1}$$

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

Let i be an identity matrix of size listed as the low index, $I_0=1$ $\otimes$ and $\oplus$ be the symbols of Kronecker's product and sum of matrices ;

$$D_i = diag\{D_i^{(r)},\ r=\overline{1,R}\},\ i=\overline{0,N-1};$$

$$D_{N,l} = diag\left\{\sum_{k=l}^{\infty} D_k^{(r)},\ r=\overline{1,R}\right\},\ i=\overline{0,N};$$

$$\mu = diag\{\mu^{(r)}, r=\overline{1,R}\}.$$

Lemma 1. Infinitesimal generator A of the Markov chain has the following block structure ξ_t, $t \geq 0$:

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

$$=\begin{pmatrix} D_0+Q\otimes I_{\overline{W}} & D_1 & D_2 & \cdots & D_{N,N} \\ \mu I_{\overline{W}} & D_0+Q\otimes I_{\overline{W}}-\mu I_{\overline{W}} & D_1 & \cdots & D_{N,N-1} \\ 0 & 2\mu I_{\overline{W}} & D_0+Q\otimes I_{\overline{W}}-2\mu I_{\overline{W}} & \cdots & D_{N,N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & D_{N,0}+Q\otimes I_{\overline{W}}-N\mu I_{\overline{W}} \end{pmatrix} \tag{2}$$

To solve system (1) we use the effective stable procedure basing on the special structure of the matrix A (it is upper block hessenbergian) and probabilistic meaning of unknown vector **p**. Such a procedure was done in [4]. It is briefly described in the following statement.

Proposition 1. The stationary probability vectors p_i, $i=\overline{0,N}$, are calculated as follows :

$$p_l,\ i=p_0 F_l,\ l=\overline{1,N},$$

where the matrices F_l are calculated recurrently

$$F_l=(\overline{A_{0,l}}+\sum_{i=1}^{l-1}F_i\overline{A_{i,l}})(-\overline{A_{l,l}})^{-1},\ l=\overline{1,N-1},$$

$$F_N=(A_{0,N}+\sum_{i=1}^{N-1}F_iA_{i,N})(-A_{N,N})^{-1},$$

the matrices $\overline{A}_{i,N}$ are calculated from the backward recursion

$$\overline{A_{i,N}} = A_{i,N}, \quad i=\overline{0,N},$$
$$\overline{A_{i,l}} = A_{i,l} + \overline{A_{i,l+1}}G_l, \quad i=\overline{0,l}, \ l=\overline{N-1,0}$$

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

$$G_i = (-A_{i+1,i+1} - \sum_{l=1}^{N-i-1} A_{i+1,i+1+l}G_{i+l}G_{i+l-1} \cdots G_{i+1})^{-1}A_{i+1}, \ i=\overline{N-1,0},$$

the vector p_0 is calculated as the unique solution to the following system of linear algebraic equations:

$$p_0\overline{A}_{0,0} = 0, \ p_0(\sum_{l=1}^{N} F_l e + e) = 1 \qquad (3)$$

The proof of the Proposition 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 explicit form.

4. Performance Measures

Having the vector p be calculated, we are able to calculate the main performance measure of the considered model. It is the probability P_{loss} that an arbitrary customer is lost in the system (loss probability).

Theorem 1. Loss probability P_{loss} is calculated as follows:

$$P_{loss} = 1 - \frac{1}{\lambda}\sum_{i=1}^{N} p_i Q_{i,i-1} e \qquad (4)$$

Here λ is a mean arrival rate into the system and is calculated by $\lambda = xD(z)'|_{z=1}e$, where the row vector x is the unique solution to the system $x(Q \otimes I_{\overline{W}} + D(1)) =, xe = 1$.

Proof. According to a formula of the total probability, the probability P_{loss} is calculated as

$$P_{loss} = 1 - \sum_{i=0}^{N-1}\sum_{k=1}^{\infty} P_k P_i^{(k)} R^{(i,k)} \qquad (5)$$

where P_k is a probability that an arbitrary customer arrivals 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 arrivals; $R^{(i,k)}$is a probability that an arbitrary customer will not be loss conditional it arrivals in a batch consisting of k customers and i servers are busy at the arrival epoch. Taking into account that the row vector x characterizes the stationary state distribution of the continuous time Markov chains with the generator $Q \otimes I_{\overline{W}} + D(1)$, which governs customers arrival into the system, one can see that $xD_k^0 e$ is an intensity of arrival of batch of size k, $kxD_k^0 e$ is an intensity of arrival of customers in batches of size k, and $\lambda = xD(z)|'e$ is a mean arrival rate.

It can be shown that

$$P_i^{(k)} = \frac{p_i D_k^{(i)} e}{D_k^{(0)} e}, \ i=\overline{0,N-1}, \ k \ge 1 \qquad (6)$$

$$P_k = \frac{kxD_k^{(0)}e}{xD(z)|'e} = k\frac{xD_k^{(0)}e}{\lambda}, \ k \ge 1 \qquad (7)$$

$$R^{(i,k)} = \begin{cases} 1, & k \le N-i \\ \frac{N-i}{k} & k > N-i, i=\overline{0,N-1} \end{cases} \qquad (8)$$

By substituting (6)-(8) into (5) after some algebra we get (4).

In trivial way we can also to calculate a number of other stationary performance measures of the considered model.

- The probability to see i busy servers

$$p_i = p_i e, \ i=\overline{0,N}$$

- The mean number of busy servers

$$N_{busy} = \sum_{i=1}^{N} i p_i e.$$

- The joint probability to see i busy servers, the random environment in the state r and the process v_t in the state v

$$p(i,r,v) = p_i \begin{pmatrix} 0_{r-1} \\ 0 \\ 0_{R-r} \end{pmatrix} \otimes e_{\overline{W}}, \ i=\overline{0,N}, \ r=\overline{1,R}, \ v=\overline{0,W}$$

where e_n and 0_n are n-dimensional column vectors consisting of units and zeros respectively.

- The joint probability to see i busy servers and the random environment in the state r

$$p_i(r) = \sum_{v=0}^{\overline{W}} p(i,r,v), \ i=\overline{0,N}, \ r=\overline{0,R}.$$

5. Numerical Examples

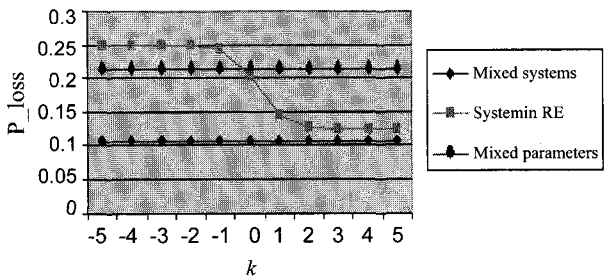
Present the results of two experiments. The goal of the first experiment is confirmation of some intuitively clear reasoning relating the possibility of approximation of the system operating in random environment (system in RE).

The first type approximation is described as follows. We extend *BMAP/M/N/0* model assuming that the system has R different modes of operations. Parameters of the r-th system are defined by the parameters of the r-th operation mode. To approximate some performance characteristic of the system in RE we calculate the same characteristic for each of R systems without RE and then average them according to the stationary distribution of the RE.

The second type approximation is described as follows. The approximated characteristic is calculated as the corresponding characteristic of an averaged *BMAP/M/N/0* system. The parameters of this system are obtained by means of averaging the corresponding parameters of the initial system in RE according to the stationary distribution of the RE. <Figure 1> illustrates the dependence of the value P_{loss} for the system in RE and the same value calculated by the first type approximation ("mixed system") and by the second type approximation ("mixed parameters") on the RE rate. We define the random environment with different rates by their generators having the form $Q^{(k)} = Q^{(0)} \cdot 10^k$, where the generator $Q^{(0)}$ describes the RE whose rates are comparable with the rates of input and service processes.

$$Q^{(0)} = \begin{pmatrix} -5 & 5 \\ 15 & -15 \end{pmatrix}$$

We vary the parameter k from -5 to 5 what corresponds to the variation of the RE rate from "very slow" (comparing the rates of input flow, service and retrial processes) to "very fast".



<Figure 1> Dependence of the loss probability on the RE rate

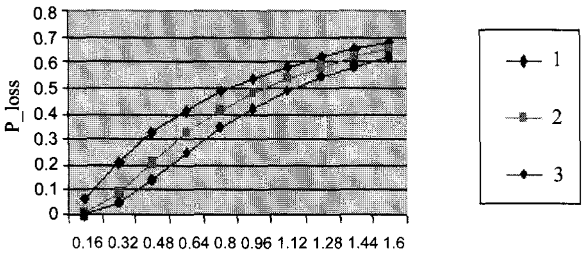
The input flow is described by the $BMAP_3$ and $BMAP_4$ and service process is defined by PH_7 and PH_8. The results are presented on <Figure 1>. Mention that the exact value of the loss probability lies in the interval between approximate values given by "mixed system" and "mixed parameters" approximations on <Figure 1> where the correlation in the $BMAP_3$ and $BMAP_4$ is not very high and differs not very essentially. Figure 1 shows that the first type approximation is good in case of "slow RE" and the second one can be applied in case of "fast RE". And there is an interval for RE rate (approximately the interval (-1, 2)) where we cannot use the estimates for P_{loss} calculated by the considered approximating models. <Figure 1> confirm the importance of investigation implemented in this paper. Simple engineering approximations can lead to unsatisfactory performance evaluation and capacity planning in real life systems.

In the second experiment we compare the main performance measures of the original system in RE and more simple exponential queueing systems, which can be considered as "engineer" approximations of the original system. The first type approximate model is the system *M/M/N/0* in the RE. It differs from the original system by assumption that input flows in its modes are stationary Poisson ones whose intensities are equal to fundamental rate of corresponding *BMAP*s in the original system. The second type approximate model is the Erlang system *M/M/N/0* whose parameters are obtained by means of averaging the corresponding parameters of just described system *M/M/N/0* in the RE according to the stationary distribution of the RE.

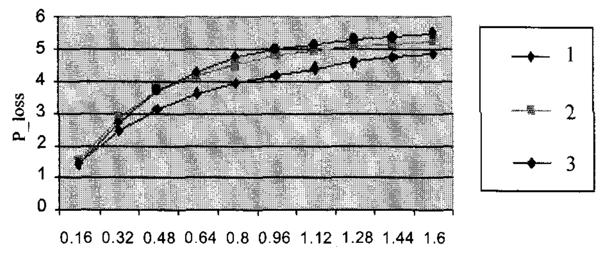
Let us assume that the *RE* has two states. One state corresponds to peak traffic periods, the second one corresponds to normal traffic periods. Service times during these periods are defined by PH_1 and PH_2 distributions. Arrivals during these periods are defined by the stationary Poisson flow with the rates λ_1 and λ_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 arrival rate during peak periods and to increase it suitable during the normal traffic periods) without changing the average arrival rate, the loss probability 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 redistribution.

We assume that the averaged arrival rate λ should be 1.25 and consider four different situations: very big difference $\lambda_1 = 10\lambda_2$ (curve 3), big difference $\lambda_1 = 3\lambda_2$ (curve 2) and equal arrival rates $\lambda_1 = \lambda_2$(curve 1). The generator of the random environment is as follows $Q = \begin{pmatrix} -15 & 15 \\ 5 & -5 \end{pmatrix}$.

It can be seen from <Figure 2> that smoothing the peak rates can cause essential decrease of the loss probability. <Figure 2> shows the dependence of the loss probability on the system load in the original system (curve 1), in the first type approximate model (curve 2) and in the second type approximate model (curve 3)



<Figure 2> Dependence of the loss probability in the system in RE and in the exponential system on the system load



<Figure 3> Dependence of mean number of busy servers in the system in RE and in the exponential system on the system load.

<Figure 3> shows the dependence of the mean number of busy servers on the system load in the original system (curve 1), in the first type approximate model (curve 2) and in the second type approximate model (curve 3).

One conclusion from the figures, which is expectable due to [4], is that the variation of service time has significant impact on the loss probability. One more conclusion, which agrees with numerical results presented in [4], is that the lower variance of service time implies the higher value of the loss probability. More surprising conclusion is that the loss probability is lowest, except the case of low traffic, when periods of high service time variation alternate with periods when the service time has small variation under the constant mean service rate.

References

[1] Bocharov, P. P., D'Apice, C., Pechinkin, A. V., and Salerno, S.; "The stationary characteristics of the *G/MSP/1/r* queueing system," *Automation and Remote Control,* 64 : 288-301, 2003.

[2] Fisher, W. and Meier-Hellstern, K. S.; "The Markov-modulated Poisson Process (MMPP) cookbook," *Performance Evaluation*, 18 : 149-171, 1993.

[3] He, Q. M.; "Queues with marked customers," *Advances in Applied Probability*, 28 : 567-587, 1996.

[4] Kim, C. S., Klimenok, V. I., Orlovsky, D. S., Dudin, and A. N.; "Lack of invariant property of the Erlang loss model in case of MAP input," *Queueing Systems*, 49 : 187-213, 2005.

[5] Klimenok, V. I.; "Characteristics calculation for multi-server queue with losses and bursty traffic," *Automatic Control and Computer Sciences*, 12 : 393-415, 1999.

[6] Klimenok, V. I. and Dudin, A. N.; "Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory," *Queueing Systems*, 5 4 : 245-259, 2006.

[7] Lucantoni, D.; "New results on the single server queue with a batch Markovian arrival process," *Comm. Stat. -Stochastic Models*, 7 : 1-46, 1991.

[8] Neuts, M. F.; *Matrix-geometric solutions in stochastic models*, The Johns Hopkins University Press, Baltimore, 1981.

[9] Neuts, M. F.; *Structured stochastic matrices of M/G/1 type and their applications*. Marcel Dekker, NY, 1983.

[10] Panchenko, A. and Buchholz, P.; "A hybrid algorithm for parameter fitting of Markovian arrival," *14th Int. Conf. Analytical and stochastic modelling technique and applications*, Prague, 2007.

[11] Sevastjanov, B. A.; "Erlang formula in telephone systems under the arbitrary distribution of conversations duration," *Proceedings of 3rd All-Union Mathematical Congress, Academy of Science, Moscow*, 4 : 68-70, 1959.

[12] Yechiali, U. and Naor, P.; "Queueing problems with heterogeneous arrivals and service," *Operations Research*, 19 : 722-734, 1971.