APPLICATION OF THE MARKOV THEORY TO QUEUING NETWORKS *

UDC 519.217 : 519.61

Goran Petrović, Nikola Petrović, Zoran Marinković

Mechanical Engineering Faculty, University of Niš A. Medvedeva 14, 18000 Niš, Serbia,
E-mail: pgoran@masfak.ni.ac.yu

Abstract. *The paper presents an application of the Markov theory to the model of networked transport system as another different view of determining the system performances (average number of transport units and average time of the given unit's stay in each subsystem of the given model). The application of the Markov theory as one very important statistical tool is described with reference to the simplest model M/M/1. Since the solution of complex tasks from the field of transport material flow very often requires a combination of particular system elements, the queuing network system is the subject of this paper's research. The overall results of the research represent a developed procedure for determining the system performances which can further on be used in developing conceptions of material flow systems.*

Key words: *Queuing System, Markov theory, Modeling, Queuing Networks*

1. INTRODUCTION

Queuing networks find a wide application in many spheres of life such as manufacturing systems, computer networks, telecommunications, transport, logistics and the like. A series of interconnected stations for serving in which each user, after departing from a particular station, can pass into another one or exit from the whole system is called the networked queuing system. The probability and statistical methods are the most frequently used tools for the system performances determination. The aim of the paper is to represent the basic techniques of probability modeling for analyzing the performances of queuing networks by means of the Markov processes. This field has been quite explored so far [1,2,3], while in the paper special attention is devoted to networked systems, that is, the systems which cannot be analyzed in the classical way. A property of Markov, that is, the property that, if the "past" and "present" are known, then the "future" depends only on

the "present" but not on the "past," enables the application of simple mathematical means to the analysis of the performances of random processes. The queuing networks are most often represented in the form of a direct graph whose nodes represent serving stations and whose branches represent links among particular stations. The pathways by which the users move from one station to another are determined by the probabilities of transition from one state to another. The transition time from one state to another, in the general state, implies that it is equal to zero.

Today, in the world at large, the material flow systems are so conceived as to minimize the waiting time since in this way the stocks are reduced and so are capital costs. Appropriate analyses should be carried out even in the planning phase since, if this is not done, later on, in the real queues systems there emerges the need for creating "keeping" space. Queuing systems depend on the way and characteristics of service processes at "bottlenecks" which, in its way, represents a decisive factor in dimensioning the waiting space which is considered as a temporary or *buffer* storage. Short-time waiting that rarely happens assumes no special action to be taken unless the material flow process is essentially violated. For longer waiting it is necessary to plan, in transport flows, appropriately dimensioned waiting premises (inter-storages, delay routes, accumulation, storages, and the like).

2. QUEUING SYSTEMS

Queuing systems can be described as called upon by particular entities (users) who ask for services. In each serving system we can distinguish arrival process, service process and one or more service stations or servers. The general assumption is that one station cannot at the same time serve two or more arrival entities. If the station is busy, the user has to wait for service. At the very moment when the station becomes free, the entity is taken over from the queuing according to the pre-defined rules – discipline in the queues – and its service is done. During the service, the entities can pass from one or more service states before departing from the system. This notation was defined in the mid-twentieth century by D. G. Kendall [4] and it is still used as such. The scheme of such defined system is given in Fig. 1.
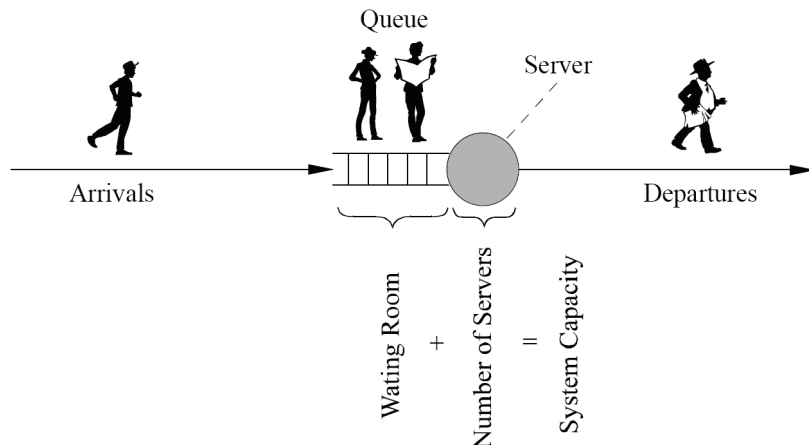
Fig. 1 Scheme of Queuing System

The arrival process is a stochastic process defined by adequate statistical distribution. Very often the arrival process can be described by exponential distribution of interim of the entity's arrival to its service or by Poisson's distribution of the number of arrivals. The parameter which characterizes the arrival process is *arrival rate* $\lambda = 1/E(t_d)$ and it represents an average number of arrivals per unit of time. Likewise, the service process is defined by statistical distribution (very often exponential) while the basic parameter which characterizes this process is *service rate* $\mu = 1/E(t_s)$ and it represents average number of users per unit of time.

Another very important parameter is *loading factor* of the service station or *utilization factor* ρ. Utilization factor can be defined, in general, for any service system, as part of the time in which the service station is busy. As the system cannot physically do more jobs than its capacity allows for, utilization factor is, in the upper respect, limited by unit [5]:

$$\rho = \min\left\{\frac{\lambda}{m \cdot \mu}, 1\right\}, \tag{1}$$

where *m* is the number of service stations. In the case that the arrival rate is greater than the service rate the system is completely busy. An increasing number of entities has to wait in a queue which then tends to increase to infinity. Such systems are called non-stationery or unstable. The system is stable only if the arrival rate is *less* than the service rate ($\rho < 1$).

3. MARKOV MODEL OF THE SIMPLEST QUEUING SYSTEM

The simplest model in the queuing theory is that of M/M/1 model. In this model both the arrival process and the service process are defined as random processes with exponential distribution of interim arrivals, that is service *(denotation M)*. Likewise, there is only one service station $m = 1$ which serves the entity which is the first to reach it (FIFO principle). For the analysis of such systems of practical importance are homogeneous Markov's models with discrete number of states. An example of such models are processes of births and deaths which serve as models for describing various natural, engineering and social systems while the name of these processes is derived from their original application to the description of biological population [6].

Any queuing system described by the process of births and deaths can be represented by a graph of discrete states as given in Fig. 2.
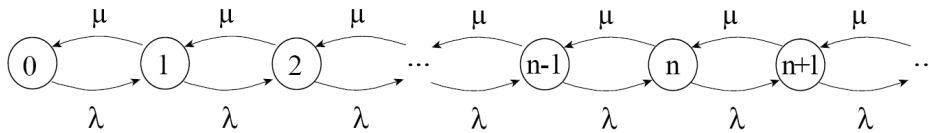

Fig. 2 Graph of State for M/M/1 Model of Queuing System

Each system state has adjoined a certain number denoting the number of units in the system ($0,1,2,...,n$). The arrows point to the direction of possible transitions from the state

system with transition rates λ and μ. In such a defined graph of state there can appear changes as with real systems when an entity joins the system ($n \to n + 1$) or when the entity is served ($n \to n - 1$). Such a defined model directly imposes two questions:

1. With what probability can one entity, for constantly little time Δt, be included into the system?
2. With what probability can one entity, for constantly little time Δt, be excluded from the system?

The answers to these questions are given by the following analysis. The probability of state transition, that is, probability that the system will pass, from state i at moment t, into state j at moment $t + \Delta t$ is equal to the product of some constant λ and time interval Δt. This assumption is very easily proved if we have in mind Poisson's process of units' arrival for serving as well as exponential distribution of service time. Parameter λ represents a number of events per unit of time which, regarding the queuing theory, is the number of arrivals per unit of time, that is, arrival rate. Similarly we can also define the probability of returns to some former state, that is, if the system is, at moment t, in state j, then the probability that the system, at moment $t + \Delta t$, will be in state i is equal to the product of some constant μ and time interval Δt, which, in this case, represents service rate.
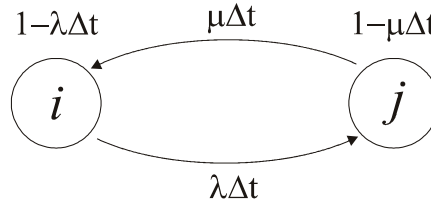

Fig. 3 Diagram of State Transition

The probability that the system is, at moment of time $t + \Delta t$, in state i is equal to the sum of the probability that the system has, even after time interval Δt has expired, remained in state i, that is, that no event has taken place (which can be written in the form $P_i \cdot (1 - \lambda \cdot \Delta t)$) and the probability that system has passed, from state j, in time interval t, into state i in time interval $t + \Delta t$ (that is, $P_j \cdot \mu \cdot \Delta t$). On the other hand, the probability that, at moment of time $t + \Delta t$, the system is in state j is equal to the sum of the probability that the system was in state j even at moment of time t ($P_j \cdot (1 - \mu \Delta t)$) and the probability that the system has passed, from state i, at moment of time t, into state j in time interval $t + \Delta t$ (that is, $P_i \cdot \mu \cdot \Delta t$). This analysis leads to the following relations:

$$P_i(t + \Delta t) = P_i(t) \cdot (1 - \lambda \cdot \Delta t) + P_j(t) \cdot \mu \cdot \Delta t \tag{2}$$
$$P_j(t + \Delta t) = P_i(t) \cdot \lambda \cdot \Delta t + P_j(t) \cdot (1 - \mu \cdot \Delta t)$$

Since time t is a continuous variable, by searching for boundary value ($\Delta t \to 0$) of previously defined expressions (2) and by generalizing for n discrete states, we obtain differential equations in the form:

$$\frac{dP_i(t)}{dt} = \sum_{k=1}^{i-1} P_k(t) \cdot \mu_{i,k} + \sum_{r=i+1}^{n} P_r(t) \cdot \lambda_{i,r} - P_i(t) \cdot \lambda_i \qquad (3)$$

Or in the matrix form:

$$\frac{dP(t)}{dt} = P(t) \cdot A \qquad (4)$$

where $P(t) = [P_i(t)]$ is vector of state probabilities while A is matrix of state transition which is of the following form:

$$A = \begin{bmatrix} -\lambda_1 & \lambda_{1,2} & \lambda_{1,3} & \cdots \\ \mu_{2,1} & -\lambda_2 & \lambda_{2,3} & \cdots \\ \mu_{3,1} & \mu_{3,2} & -\lambda_3 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \text{ where: } \lambda_i = \sum_{k=1}^{i-1} \mu_{i,k} + \sum_{r=i+1}^{n} \lambda_{i,r} \qquad (5)$$

If it is $t \rightarrow \infty$, that is, "if the system works long enough," then probabilities $P_i(t)$ tend to constant values P_i^* which, in this case, represent the probabilities of stationary system state (equilibrium) and do not depend on time. In this case, the first derivative is:

$$\frac{dP(t)}{dt} = 0 \qquad (6)$$

System of differential equations (4) becomes a system of common linear equations in the form:

$$0 = P \cdot A \qquad (7)$$

The solution of the system of linear equations defined in this way is relatively easy under the condition that the sum of all the state probabilities must be equal to one:

$$\sum_{i=1}^{n} P_i = 1 \qquad (8)$$

There is a great number of developed methods for solving the posed problem. If it is necessary to determine the system performances at different moment of time, then the system of differential equations (4) will be solved by some of the numerical methods such as, for instance, Runge Kutt; if, on the other hand, the knowledge of the system equilibrium, that is, its steady-state is enough, then it is necessary to solve a system of common linear equations (7) by means of some of the well-known direct or iterative methods (Gauss-Seidel,....). The application of the modern mathematical software such as MATLAB or MATHEMATICA provides for simple solutions of the set equations as shown in the example presented in the second section of the paper.

4. QUEUING NETWORKS OF MATERIAL FLOWS

The solution of the tasks related to material flows very often requires a combination of particular elements in the form of networked systems. Queuing networks, composed of a great many of service stations, are much more suitable for describing complex structures such as real transport systems. We can distinguish open, closed and mixed queuing networks [7]. An open queuing network is that in which transport units can enter from the surroundings and which they can depart from after being served. The closed networks are those in which transport units cannot enter the network or depart from it, that is, the number of transport units in the system is constant. Likewise, the system in which a new unit enters the very moment when another departs from the system represents a closed network. If the queuing network comprises both open and closed contours it is called a mixed network.

The mathematical means presented in the previous section can be very successfully applied when it comes to queuing networks. The determination of stationary probabilities of all the possible states can be defined as the central problem of the queuing theory [8]. The mean values of all the other system performances can simply be determined if the stationary probabilities of all the possible states are known.

The most important probability characteristics of each queuing network are *marginal probabilities*. Marginal probability $\pi_i(k)$ that is, probability that i-th node comprises exactly $k_i = k$ transport units can be determined in the following way:

$$\pi_i(k) = \sum_{k_i = k} \pi(k_1, ..., k_N) \tag{9}$$

where: N - number of nodes in the system.

Accordingly, $\pi_i(k)$ is the sum of probabilities of all the possible multi-dimensional states of system $(k_1, ... k_N)$ that satisfy the condition that i-th node comprises exactly $k_i = k$ transport units. Surely, it is also necessary that the normalization condition should be valid, that is, that the sum of probabilities of all the possible system states is equal to zero:

$$\sum \pi(k_1, ..., k_N) = 1 \tag{10}$$

Knowing marginal probabilities, it is relatively simple to determine average number of units at each system's node, average value of time of keeping the unit in the system as well as other system performances. Average number of units at i-th node would be:

$$E[N_i] = \sum_{k=0}^{\infty} k \cdot \pi_i(k) \tag{11}$$

By applying Little's theorem [4] time of keeping the unit at i-th node of the system would be:

$$E[T_i] = \frac{E[N_i]}{\lambda_i^*} \tag{12}$$

where λ_i^* - overall arrival rate of units at i-th node.

5. NUMERICAL METHOD

The above-presented methodology will be illustrated by means of an example of the real transport system model which consists of two subsystems, namely PS1 i PS2 (Fig. 4).

Transport units TU enter subsystem PS1 through the transporter whose maximal capacity is limited by transporter length; then they are processed and, in part, they depart from the system while, partly, they come to the second subsystem PS2. At the entrance of the subsystem PS2 the units coming from the outside are included. Likewise, the capacity of the second subsystem PS2 is limited by transporter length. After being processed in subsystem 2, the units depart from the system. The aim of the numerical example is to determine, for the given parameters of both the subsystems, average number of transport units in each subsystem as well as average time of keeping the unit within each subsystem.
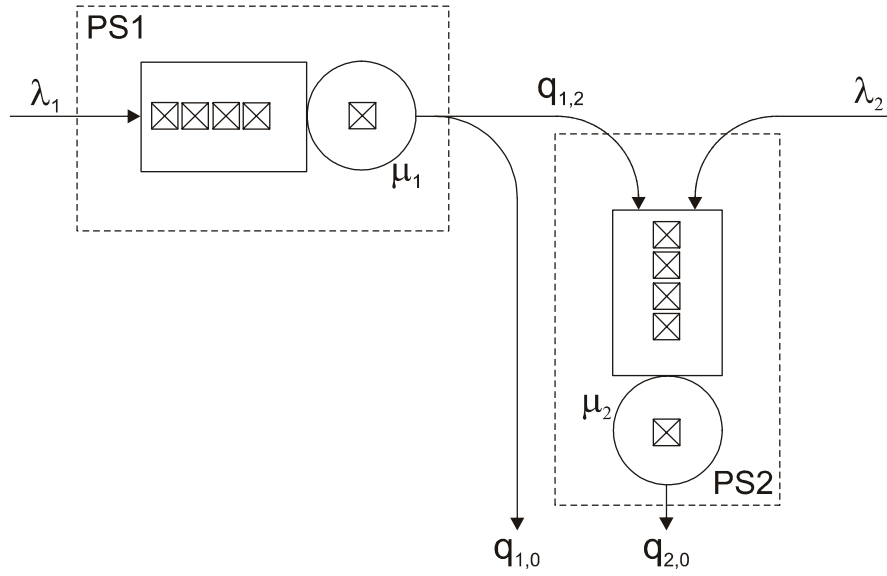

Fig. 4 Model of Real Transport System

Number of transport units entering subsystems PS1 is defined by Poisson's distribution with parameter λ_1, while procession time TU in subsystem PS1 is defined by exponential distribution with parameter μ_1. Number of TUs departing from the system is defined by parameter $q_{1,0}$, while the number of units entering into queue of subsystem PS2 is defined by parameter $q_{1,2}$. The queue of subsystem PS2 also includes the units coming from the surroundings by Poisson's distribution with parameter λ_2. TUs leave subsystem PS2 after the processing which is defined by exponential distribution of service time with parameter μ_2. Thus defined numerical example represents an open network of the queuing system for whose solution we have applied The Markov theory presented in Section 3. The set model represented in the form of the graph of state transition is given in Fig. 5.
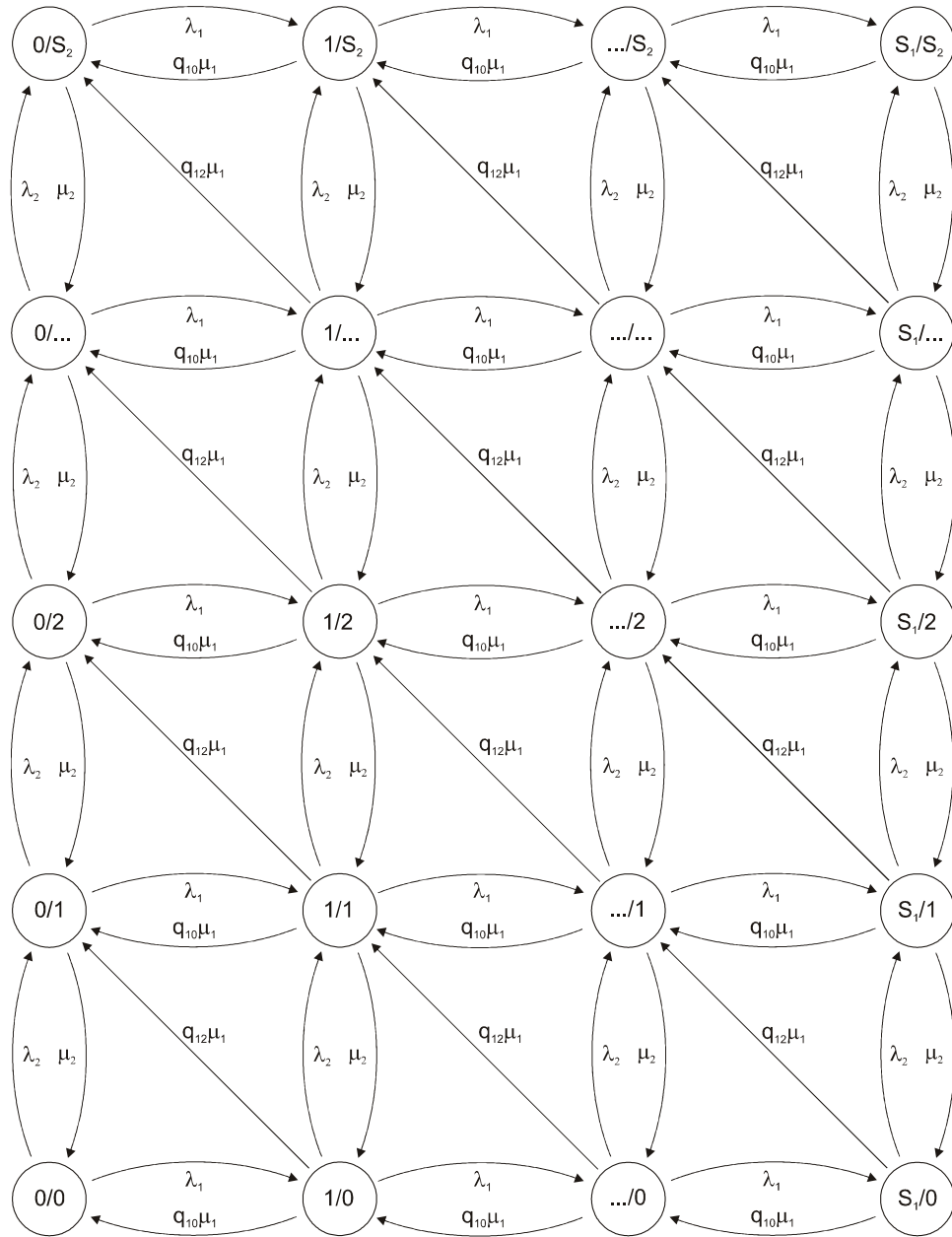
Fig. 5 Graph of Transition State of the Given Model

If the capacity of subsystem PS1 is $S_1 = 5$ and of subsystem PS2 $S_2 = 5$ then the matrix of transition state (5) gets the following form:

Table 1 Matrix of Transition State of the Given Model

	(0,0)	(1,0)	(2,0)	(3,0)	(4,0)	(5,0)	(0,1)	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(0,2)	(1,2)	...	(4,5)	(5,5)
(0,0)	$-\Sigma_1$	λ_1	0	0	0	0	λ_2	0	0	0	0	0	0	0		0	0
(1,0)	$q_{10}\mu_1$	$-\Sigma_2$	λ_1	0	0	0	$q_{12}\mu_1$	λ_2	0	0	0	0	0	0		0	0
(2,0)	0	$q_{10}\mu_1$	$-\Sigma_3$	λ_1	0	0	0	$q_{12}\mu_1$	λ_2	0	0	0	0	0		0	0
(3,0)	0	0	$q_{10}\mu_1$	$-\Sigma_4$	λ_1	0	0	0	$q_{12}\mu_1$	λ_2	0	0	0	0		0	0
(4,0)	0	0	0	$q_{10}\mu_1$	$-\Sigma_5$	λ_1	0	0	0	$q_{12}\mu_1$	λ_2	0	0	0		0	0
(5,0)	0	0	0	0	$q_{10}\mu_1$	$-\Sigma_6$	0	0	0	0	$q_{12}\mu_1$	λ_2	0	0		0	0
(0,1)	μ_2	0	0	0	0	0	$-\Sigma_7$	λ_1	0	0	0	0	λ_2	0	...	0	0
(1,1)	0	μ_2	0	0	0	0	$q_{10}\mu_1$	$-\Sigma_8$	λ_1	0	0	0	$q_{12}\mu_1$	λ_2		0	0
(2,1)	0	0	μ_2	0	0	0	0	$q_{10}\mu_1$	$-\Sigma_9$	λ_1	0	0	0	$q_{12}\mu_1$		0	0
(3,1)	0	0	0	μ_2	0	0	0	0	$q_{10}\mu_1$	$-\Sigma_{10}$	λ_1	0	0	0		0	0
(4,1)	0	0	0	0	μ_2	0	0	0	0	$q_{10}\mu_1$	$-\Sigma_{11}$	λ_1	0	0		0	0
(5,1)	0	0	0	0	0	μ_2	0	0	0	0	$q_{10}\mu_1$	$-\Sigma_{12}$	0	0		0	0
(0,2)	0	0	0	0	0	0	μ_2	0	0	0	0	0	$-\Sigma_{13}$	λ_1		0	0
(1,2)	0	0	0	0	0	0	0	μ_2	0	0	0	0	$q_{10}\mu_1$	$-\Sigma_{14}$		0	0
...										...							
(4,5)	0	0	0	0	0	0	0	0	0	0	0	0	0	0		$-\Sigma_{35}$	λ_1
(5,5)	0	0	0	0	0	0	0	0	0	0	0	0	0	0		$q_{10}\mu_1$	$-\Sigma_{36}$

Where $-\Sigma_i$ represents negative sum of all the elements in i-th row.

To solve equation (7), under the condition defined by expression (8), the MATLAB software is used. For random values $\lambda_1 = 0{,}7$ [1/s], $\lambda_2 = 0{,}2$ [1/s], $\mu_1 = 1$ [1/s], $\mu_2 = 1$ [1/s], $q_{10} = 2\,/\,7$, $q_{12} = 5\,/\,7$, which are characteristic for the class of engineering systems, the final (stationary) probabilities of multi-dimensional system states are given in Tab. 2.

Table 2 Matrix of Final Probabilities of Multi-dimensional System States

(0,0) 0.0934	(0,1) 0.0572	(0,2) 0.0312	(0,3) 0.0185	(0,4) 0.0106	(0,5) 0.0068
(1,0) 0.0771	(1,1) 0.0512	(1,2) 0.0291	(1,3) 0.0171	(1,4) 0.0099	(1,5) 0.0063
(2,0) 0.0677	(2,1) 0.0455	(2,2) 0.0272	(2,3) 0.0158	(2,4) 0.0095	(2,5) 0.0045
(3,0) 0.0612	(3,1) 0.0289	(3,2) 0.0172	(3,3) 0.0100	(3,4) 0.0061	(3,5) 0.0031
(4,0) 0.0620	(4,1) 0.0368	(4,2) 0.0212	(4,3) 0.0125	(4,4) 0.0069	(4,5) 0.0049
(5,0) 0.0619	(5,1) 0.0389	(5,2) 0.0228	(5,3) 0.0134	(5,4) 0.0076	(5,5) 0.0062

The numeration in the Table 2 has the following meaning: $p^*(2,3) = 0.0158$ - final probability that in subsystem PS1 there are exactly two TUs while in subsystem PS2 there are exactly three TUs. Based on determined final probabilities, it is possible, by applying expression (10), to determine marginal probabilities for each subsystem (PS1 and PS2), that is, for possible number of TUs in each subsystem (0,1,...,5), as given in Table 3.

Table 3 Matrix of Marginal Probabilities π for Subsystems PS1 and PS2

π	Number of transport units in respective subsystem					
	TU=0	TU=1	TU=2	TU=3	TU=4	TU=5
PS1	0.2176	0.1907	0.1702	0.1265	0.1443	0.1507
PS2	0.4233	0.2584	0.1486	0.0873	0.0505	0.0318

Finally, on the basis of marginal probabilities, by applying expression (11), we determine average number of units in each subsystem (13), whose values are given in Table 4.

$$E[N_1] = 0 \cdot \pi_1(0) + 1 \cdot \pi_1(1) + 2 \cdot \pi_1(2) + 3 \cdot \pi_1(3) + 4 \cdot \pi_1(4) + 5 \cdot \pi_1(5) \qquad (13)$$

$$E[N_2] = 0 \cdot \pi_2(0) + 1 \cdot \pi_2(1) + 2 \cdot \pi_2(2) + 3 \cdot \pi_2(3) + 4 \cdot \pi_2(4) + 5 \cdot \pi_2(5)$$

Table 4 Average Number of Units in Subsystems PS1 and PS2.

$E[N_1]$	$E[N_2]$
2.2414	1.1789

Likewise, by applying equations (12), we determine average values of keeping time of TUs in each subsystem. The arrival rates into each subsystem will be:

$$\lambda^*_1 = \lambda_1 = 0.7 \ [1/s] \qquad (14)$$

$$\lambda^*_2 = \lambda_2 + q_{12} \cdot \mu_1 = 0.914 \ [1/s]$$

Average time of keeping transport units in each subsystem is given in Table 5:

Table 5 Average Keeping Time of TU in Subsystems PS1 and PS2.

$E[T_1]$	$E[T_2]$
3.202 [s]	1.290 [s]

The represented procedure for determining the system performances is very suitable for analyzing the system itself. By varying input parameters λ_1 and λ_2 it is possible to track the changes of average number of TUs in each of the two subsystems, namely, PS1 and PS2.

The first part of the analysis assumes changes of parameter λ_1 in the range from 0.2 to 0.7 with the rate of 0.1 and constant value of parameter $\lambda_2 = 0,2$ [1/s]. In the second part of the analysis, variable is parameter λ_2 in the same range as in the first part, while parameter λ_1 has remained constant $\lambda_1 = 0,7$ [1/s]. All the other system parameters have remained unchanged: $\mu_1 = 1$ [1/s], $\mu_2 = 1$ [1/s], $q_{10} = 2 / 7$, $q_{12} = 5 / 7$.

The results of this analysis are shown in Table 6, that is they are graphically processed in Fig. 6.

Table 6 Average Number of TUs in Subsystems PS 1 and PS2 for Different Input Parameters

$\lambda_2 = 0.2 = const.$					
λ_1	0.2	0.3	0.4	0.5	0.6	0.7
$E[N_1]$	1.7235	1.8320	1.9784	2.0769	2.1635	2.2414
$E[N_2]$	0.7068	0.8029	0.8969	0.9903	1.0841	1.1789

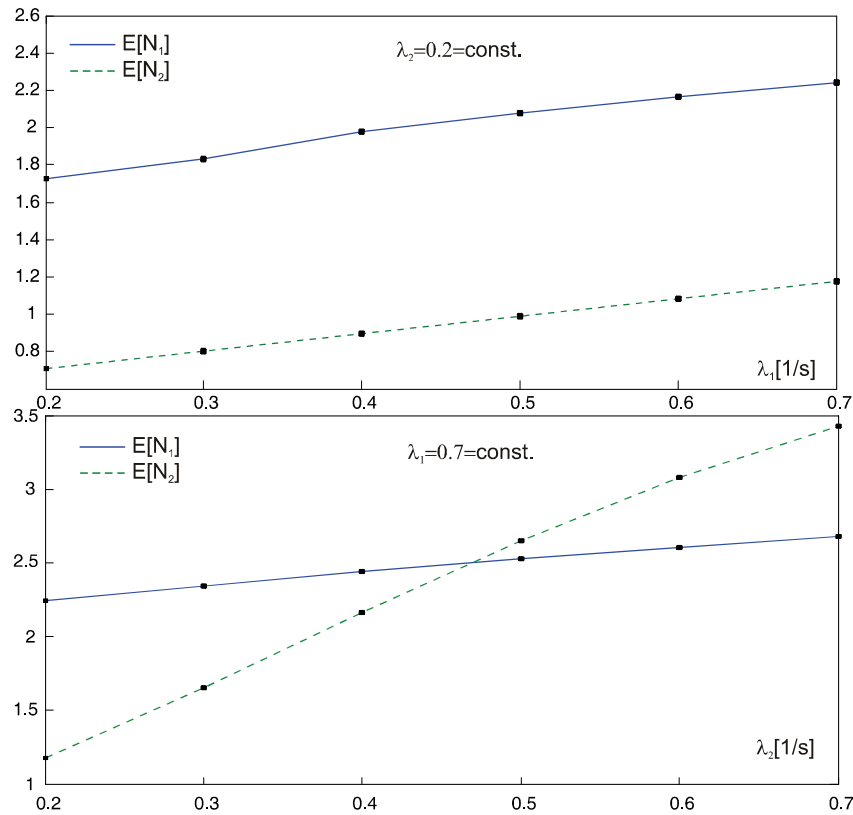| $\lambda_1 = 0.7 = const.$					
λ_2	0.2	0.3	0.4	0.5	0.6	0.7
$E[N_1]$	2.2414	2.3447	2.4389	2.5257	2.6061	2.6805
$E[N_2]$	1.1789	1.6538	2.1635	2.6518	3.0789	3.4303

Fig. 6 Graph of Changes of Number of TUs in Subsystems PS 1 and PS2

6. CONCLUSION

Regarding all that is above-said, the following conclusions can be drawn:

- Queuing systems have a wide application in all the spheres of life and since waiting is an unpleasant category, they deserve special consideration of research studies and analyses,
- Adequate modeling of queuing systems of the type M/M/... is very successfully realized by implementing the Markov theory,
- Above-presented mathematical apparatus is universal, that is, its simple application is possible for all the systems in which the processes of arrival and serving are described by exponential distribution,
- Software solutions, once developed, enable different analyses regardless of the subsystem capacities since complex matrices (matrix of state transition, matrix of final probabilities...) are automatically generated, and.
- Practical tasks from material flows are really described (represented) by means of complex networked queuing systems.

REFERENCES

1. Breuer L., Baum D., 2005, *An introduction to queueing theory and matrix-analytic methods*, Springer.
2. Asmussen S., 2003, *Applied Probabilities and Queues - Second Edition*, Springer.
3. Medhi J., 1991, *Stochastic Models in Queuing Theory*, Academic Press.
4. Cooper R. B., 1981, *Introduction to Queuing Theory - Second Edition*, Elsevier.
5. Skripta auditornih vježbi, 2001, *Teorija prometa*, Zagreb.
6. Stojaković M., 2007, Verovatnoća, statistika i slučajni procesi, Univerzitet u Novom Sadu - Fakultet Tehničkih nauka.
7. Bolch G., Greiner S., Meer H. Trivedi K. S., 1998, *Queuing Networks and Markov Chains*, John Wiley & Sons.
8. Kelly F.P., Zachary S., Ziedins I., 1996, *Stochastic Networks Theory and Applications*, Oxford University Press.
9. Stojiljković M., Rančić M., 1997, *Theory of graphs and some possibilities of finite automates optimization*, The scientific journal FACTA UNIVERSITATIS Series: Mechanical Engineering Vol.1, No 4.

PRIMENA TEORIJE MARKOVA U MREŽNIM SISTEMIMA MASOVNOG OPSLUŽIVANJA

Goran Petrović, Nikola Petrović, Zoran Marinković

U radu je prikazana primena teorije Markova na model umreženog transportnog sistema kao jedno, drugačije, viđenje određivanja performansi sistema (prosečnog broja transportnih jedinica i prosečnih vremena zadržavanja jedinica u svakom podsistemu datog modela). Primena teorije Markova kao jednog veoma bitnog statističkog alata opisana je na najjednostavnijem modelu M/M/1. Kako rešavanje složenih zadataka iz oblasti transportnih tokova materijala često zahteva kombinaciju pojedinačnih elemenata sistema, umreženi sistemi masovnog opsluživanja su predmet istraživanja u ovom radu. Sveukupni rezultat istraživanja predstavlja razvijeni postupak za određivanje performansi sistema koje se dalje mogu upotrebiti u koncipiranju sistema tokova materijala.

Ključne reči: *masovno opsluživanje, teorija Markova, modeliranje, mrežni sistemi*