Number IV, Volume V, December 2010

MODELLING AND SIMULATION OF UNRELIABLE M/M/*n*/*n* QUEUEING SYSTEM

MODELOVÁNÍ A SIMULACE NESPOLEHLIVÉHO M/M/*n*/*n* SYSTÉMU HROMADNÉ OBSLUHY

Michal Dorda[1]

Summary: This paper is devoted to modelling and simulation of a Markov multi-server queueing system subject to breakdowns and with an ample repair capacity, the system do not form the queue of waiting customers. The paper introduces a mathematical model of the studied system and a simulation model created by using software CPN Tools. At the end of the paper the outcomes which were reached by both approaches will be shown and statistically evaluated.

Key words: M/M/n/n, Queueing system, Breakdown, Petri net.

Anotace: Příspěvek je věnován modelování a simulaci Markovského vícelinkového systému hromadné obsluhy podléhajícího poruchám a s dostačující opravárenskou kapacitou, systém netvoří frontu čekajících požadavků. Příspěvek představuje matematický model studovaného systému a simulační model vytvořený v software CPN Tools. Na konci příspěvku jsou statisticky vyhodnoceny výsledky, které byly dosaženy oběma přístupy.

Klíčová slova: M/M/n/n, Systém hromadné obsluhy, Porucha, Petriho síť.

INTRODUCTION

In the queueing theory servers generally work without failures. But in many practical cases this assumption is not precise; servers are often technical devices and every technical device can break down. Many authors studied a behavior of single-server queueing systems subject to breakdowns from a different point of view. Modelling of unreliable multi-server queueing systems is not so often done due to mathematical complexity of their analysis.

The paper [1] is devoted to investigation of a multi-server Markov queueing system with an infinite queue capacity, server breakdowns and an ample repair capacity. A multi-server Markov queueing system is considered with an infinite queue capacity and server breakdowns as well, but with a limited repair capacity in papers [2] and [3]. Some experimental outcomes related to a behavior of the unreliable multi-server Markov queueing system are presented in paper [4]. Queueing systems with two unreliable servers are described in papers [5] and [6]. In paper [5] the authors assume homogeneous servers, in paper [6]

[1] Ing. Michal Dorda, Ph.D., VŠB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport , 17. listopadu 15/2172, 708 33 Ostrava – Poruba, Tel.: +420 597 325 754, E-mail: michal.dorda@vsb.cz

heterogeneous servers are considered. Paper [7] is focused on a finite Markov queueing system with balking, reneging and server breakdowns.

1. MATHEMATICAL MODEL

Let us consider a Markov multi-server queuing system formed by n homogenous parallel placed servers subject to breakdowns. The system does not make up a queue of waiting customers, therefore an incoming customer which does not find an idle server is rejected.

Customers come to the system according to the Poisson process with the parameter λ, hence the customers inter-arrival times are exponentially distributed with the mean value equal to $\frac{1}{\lambda}$. The customers service times are described by the exponential distribution with the parameter μ, thus the mean service time is $\frac{1}{\mu}$.

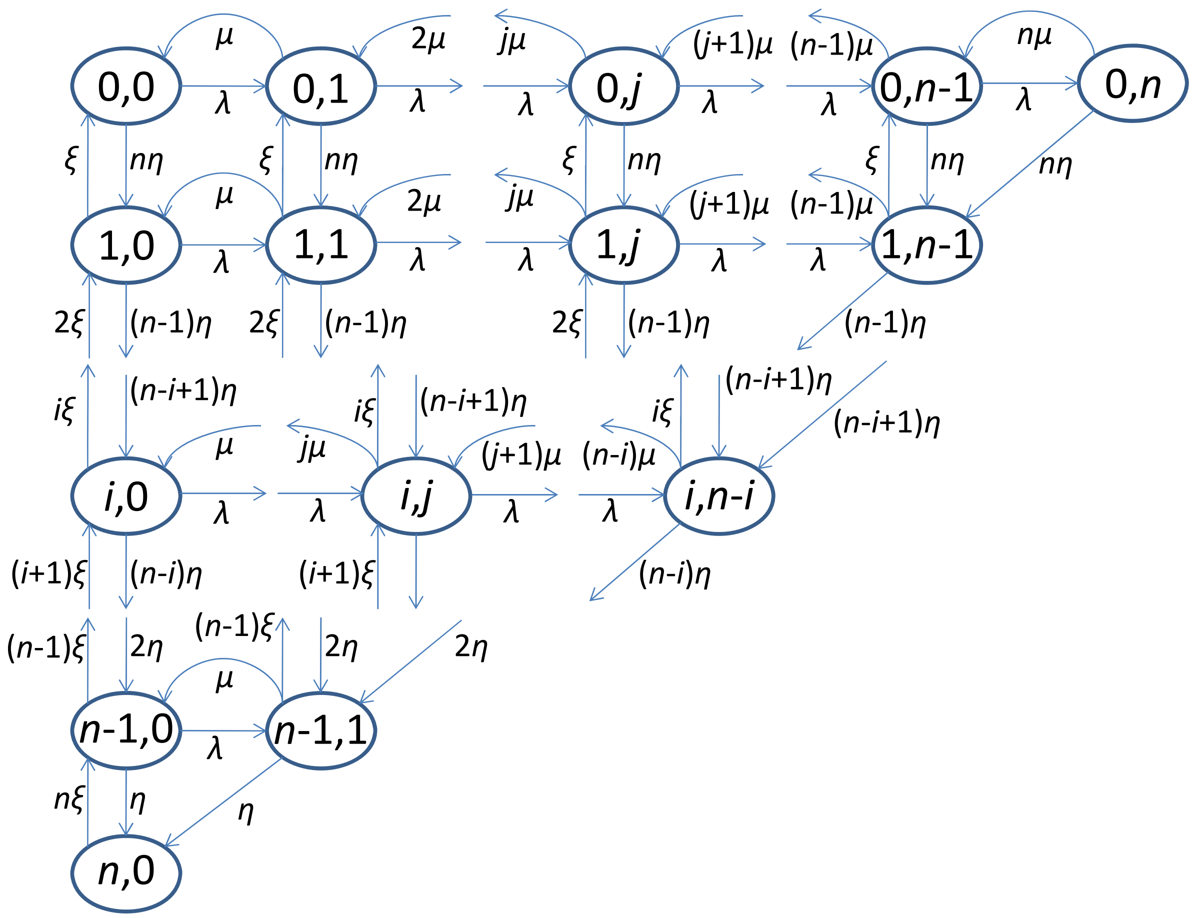
Each of the servers is successively failure-free and broken; let us assume that failures of individual servers are mutually independent. Furthermore let us consider that a server breakdown can happen at any time. That means the server can fault if it is busy or idle. The time of failure-free state is an exponential random variable with the parameter η, thus the mean value is equal to $\frac{1}{\eta}$. The server repairing time is an exponential random variable as well, but with the parameter ξ; the mean repair time is therefore $\frac{1}{\xi}$. Let us assume an ample repair capacity – repairing of each server can start immediately after the occurrence of the breakdown.

Let us assume so called preemptive-repeat discipline that means after the occurrence of the server breakdown (if the server is servicing a customer at the moment) the customer servicing is interrupted and its servicing then starts from the beginning. Two different events can generally occur:

- If there is an idle server in the system, the customer will go over to the idle server.
- If there is not an idle server in the system, the customer will be rejected.

The queueing model can be modelled as a two-dimensional Markov chain. Let us denote states of the chain by pairs (i,j), where $i \in \{0,1,...,n\}$ expresses the number of broken servers and $j \in \{0,...,n-i\}$ represents the number of the customer in the service (or in the system as well).

Let us illustrate the queueing model graphically as a state transition diagram (see in fig. 1). The vertices represent the states of the system and oriented edges indicate the possible transitions with corresponding rate.



Source: Author

Fig. 1 – The state transition diagram

With reference to the state transition diagram we can obtain the finite system of the differential equations for probabilities $P_{(i,j)}(t)$ of the particular states depending on the time *t*. For $t \to \infty$ we get the system of the linear equations for steady state probabilities $P_{(i,j)}$ that are not dependent on the time *t*. The steady state balance equations are:

$$(\lambda + n\eta)P_{(0,0)} = \mu P_{(0,1)} + \xi P_{(1,0)}, \tag{1}$$

$$(\lambda + j\mu + n\eta)P_{(0,j)} = \lambda P_{(0,j-1)} + (j+1)\mu P_{(0,j+1)} + \xi P_{(1,j)} \text{ for } j = 1,\dots,n-1, \tag{2}$$

$$(n\mu + n\eta)P_{(0,n)} = \lambda P_{(0,n-1)}, \tag{3}$$

$$[\lambda + (n-i)\eta + i\xi]P_{(i,0)} = \mu P_{(i,1)} + (n-i+1)\eta P_{(i-1,0)} + (i+1)\xi P_{(i+1,0)} \text{ for } i = 1,\dots,n-1, \tag{4}$$

$$[\lambda + j\mu + (n-i)\eta + i\xi]P_{(i,j)} = \lambda P_{(i,j-1)} + (j+1)\mu P_{(i,j+1)} + (n-i+1)\eta P_{(i-1,j)} + \\ + (i+1)\xi P_{(i+1,j)} \text{ for } i = 1,\dots,n-2;\ j = 1,\dots,n-i-1, \tag{5}$$

$$[(n-i)\mu + (n-i)\eta + i\xi]P_{(i,n-i)} = \lambda P_{(i,n-i-1)} + (n-i+1)\eta P_{(i-1,n-i)} + (n-i+1)\eta P_{(i-1,n-i+1)} \\ \text{for } i = 1,\dots,n-1, \tag{6}$$

$$n\xi P_{(n,0)} = \eta P_{(n-1,0)} + \eta P_{(n-1,1)}. \tag{7}$$

Clearly, the probabilities $P_{(i,j)}$ must satisfy the normalization equation:

$$\sum_{i=0}^{n}\sum_{j=0}^{n-i}P_{(i,j)}=1. \tag{8}$$

Thanks to the solution of the linear equations system, obtained from equations (1) – (6) and (8), we get steady state probabilities of the individual states of the system that are needed for the calculation of the performance measures. Please remember that the equation (7) is a linear combination of the equations (1) – (6), therefore it can be dropped and is replaced by the normalization equation (8).

On the basis of steady state probabilities knowledge the following performance measures can be calculated. The mean number of the costumers in the service denoted as *ES* can be found according to the formula:

$$ES=\sum_{i=0}^{n-1}\sum_{j=1}^{n-i}jP_{(i,j)}. \tag{9}$$

The utilization of the servers denoted as χ_s is expressed by the formula:

$$\chi_s=\frac{ES}{n}=\frac{\sum_{i=0}^{n-1}\sum_{j=1}^{n-i}jP_{(i,j)}}{n}. \tag{10}$$

Because the system does not accept the queue of waiting customers, it must hold on the mean number of the customers finding in the system *EK*:

$$EK=ES=\sum_{i=0}^{n-1}\sum_{j=1}^{n-i}jP_{(i,j)}. \tag{11}$$

For the mean number of broken servers *EP* it can be written:

$$EP=\sum_{i=1}^{n}i\sum_{j=0}^{n-i}P_{(i,j)}. \tag{12}$$

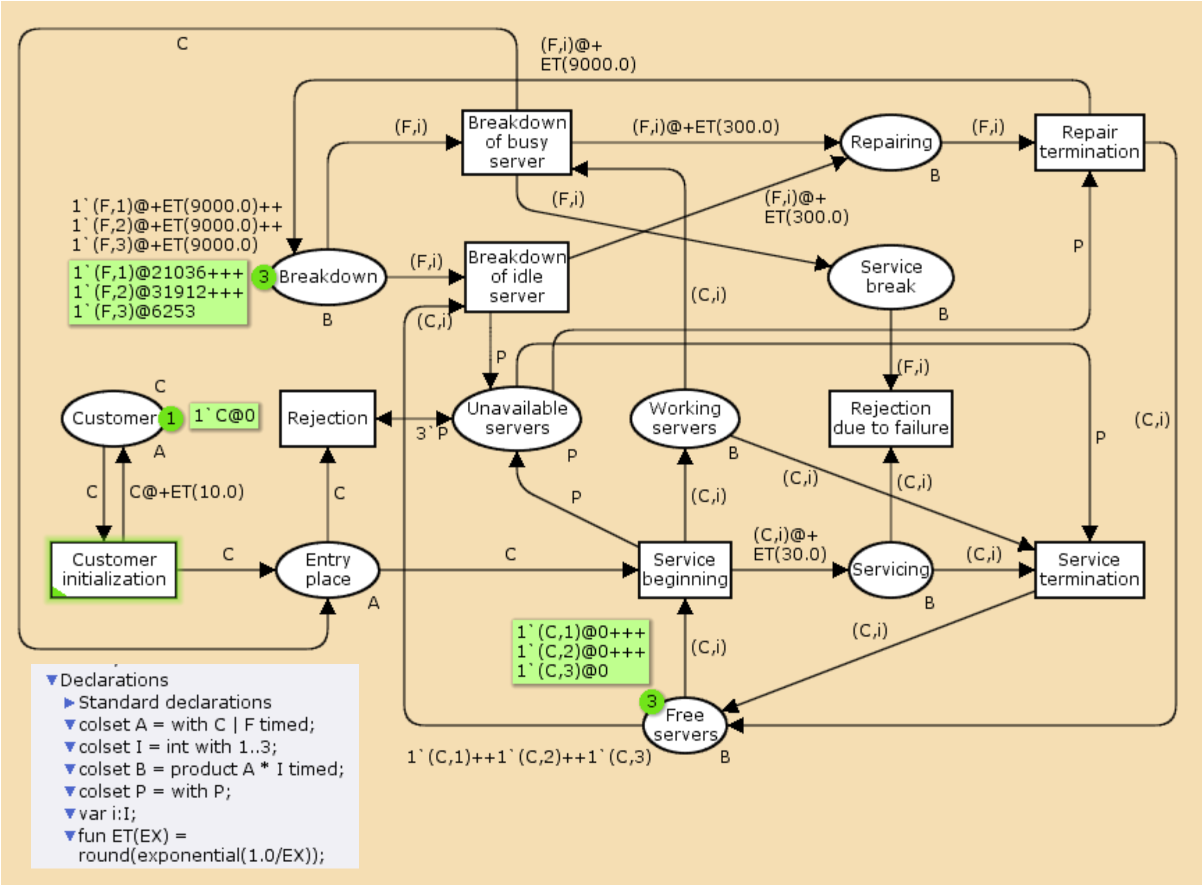
And finally the utilization of the repair capacity χ_r can be expressed by the formula:

$$\chi_r=\frac{EP}{n}=\frac{\sum_{i=1}^{n}i\sum_{j=0}^{n-i}P_{(i,j)}}{n}. \tag{13}$$

2. SIMULATION MODEL

To validate the outcomes, which were reached by the solution of the above-mentioned mathematical model, Petri net model of the studied queueing system was created by using CPN Tools – Version 2.2.0. The software CPN Tools is designed for editing, simulating and analyzing coloured Petri nets. The created simulation model is shown in fig. 2. The model is created by 9 places and 8 transitions. The Petri net presented in fig. 2 models the unreliable M/M/3/3 (thus n=3) queueing system fulfilling the conditions mentioned in Section 1, where λ=6 h^{-1}, μ=2 h^{-1}, η=150^{-1} h^{-1} and ξ=0,2 h^{-1}.

The concrete values of the random variables are generated during a simulation through the defined function fun ET(EX) = round(exponential(1.0/EX)). We define a minute as the applied unit of time and the parameter of the defined function is equal to the mean value of the exponential distribution. So the mean value of each random variable must be expressed in [min]. We substitute $\frac{1}{\lambda} = 10$ min, $\frac{1}{\mu} = 30$ min, $\frac{1}{\eta} = 9000$ min and $\frac{1}{\xi} = 300$ min in this case.



Source: Author

Fig. 2 – Created Petri net modelling the unreliable M/M/3/3 queueing system

The created Petri net works with the following tokens:

- Tokens C represent incoming and waiting customers.
- Tokens (C,*i*), where $i \in \{1,2,3\}$, are customers as well, but in the phase of service process (each customer in the service is labeled by a number of the server which is working on it); these tokens are used for modelling of available and working servers as well.
- Tokens (F,*i*), where $i \in \{1,2,3\}$, model failures of the servers.
- Auxiliary tokens P are used for modelling of free queue places and unavailable servers.

The place “Customer” with the initial marking C and the transition “Customer initialization” model the Poisson arrival process of customers; first customer comes to the system at the time 0. If there is a customer in the place “Entry place” which models customers approaching to the system, it is possible that one of the two following events will happen:

- The service of the customer will start (through the transition "Service beginning" firing), if there is an available server modelled by tokens finding in the place "Free servers" with the initial marking 1`(C,1)++1`(C,2)++1`(C,3).
- If there is not any available server that means there are 3 tokens P in the auxiliary place "Unavailable servers", the customer will be rejected through the transition "Rejection" firing.

In the place "Servicing" there are located tokens modelling servicing of customers, the appropriate service time is ensured by the time stamp update through the transition "Service beginning" firing. The transition "Service termination" will be enabled if there is a customer in the service and corresponding server is working.

Tokens which model failures of the servers are in the place "Breakdowns" with the initial marking 1`(F,1)@+ET(9000.0)++1`(F,2)@+ET(9000.0)++1`(F,3)@+ET(9000.0). When the actual simular time is equal to the value of the time stamp of the token modelling breakdown of *i*-th server, its failure will happen:

- If *i*-th server is idle at this moment (there is the token (C,*i*) in the place "Available servers"), breakdown of idle server will occur.
- If *i*-th server is busy at this moment (there is the corresponding token in the place "Working servers"), breakdown of busy server will happen – this failure will cause the rejection of corresponding customer through the transition "Rejection due to failure" firing and the input of a new costumer to the place "Entry place".

Tokens (F,*i*) which are in the place "Repairing" model repairing of the servers, exponential repairing time is ensured by the time stamp update. The repairing of *i*-th broken server is ended by the transition "Repair termination" firing and corresponding token (F,*i*) is sent with updated time stamp back to the place "Breakdowns".

The two marking size monitoring functions were applied for computing of selected performance measures during the simulation and they are:

- The monitoring function which is bonded with the place "Working servers" enables the estimation of the mean number of the customers in the service *ES*.
- The monitoring function which is associated with the place "Repairing" enables the estimation of the mean number of the servers in failure *EP*.

3. EXECUTED EXPERIMENTS AND THEIR EVALUATION

Let us consider the analysed queueing system with 3 parallel servers (n=3). Let us apply the values of the random variables parameters which are λ=6 h^{-1}, μ=2 h^{-1}, $\eta \in \{150^{-1}, 125^{-1}, 100^{-1}, 75^{-1}, 50^{-1}\}$ h^{-1} and ξ=0,2 h^{-1}, consequently we consider 5 system configurations differing in the parameter η.

Let us focus on the performance measures *ES* and *EP*, which we got according to the formulas (9) and (12). Please notice that the corresponding system of linear equations was solved numerically by using Matlab.

The experimental estimation of the performance measures we obtain by simulation of the coloured Petri net presented in Chapter 2. Thirty independent experiments were executed for each system configuration; each experiment was terminated after a half million steps (a step corresponds to a transition firing).

First of all the normality of the simulation results was tested by using the χ^2 goodness-of-fit test. As for all cases p-value is greater than 0,05 we do not reject individual hypotheses about the normal distribution. We can compute 95% confidence intervals for selected performance measures, because the normality was not rejected. The reached outcomes are summarized in tab. 1 below. Please notice that all statistical computations were executed by using software Statgraphics plus 5.0.

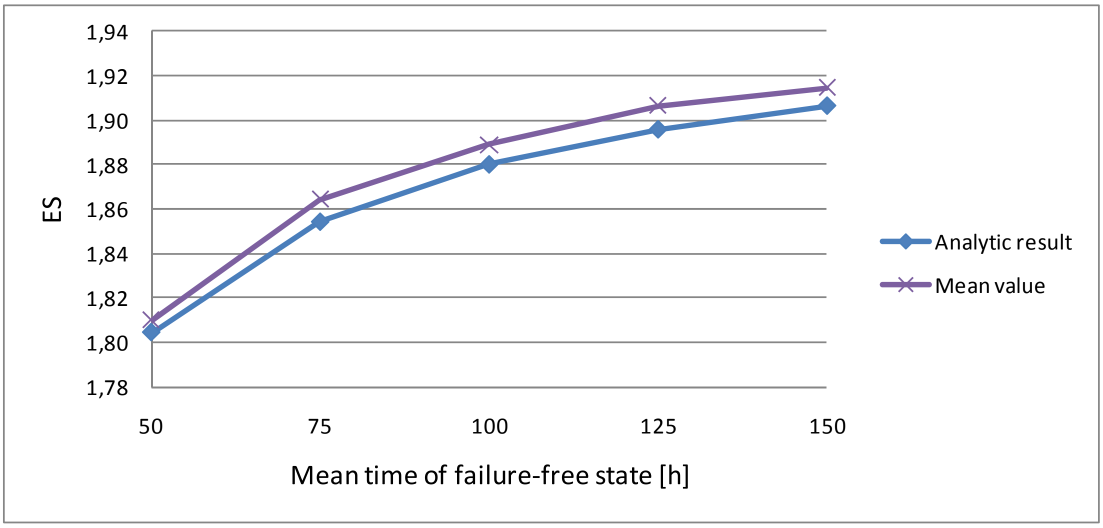
Tab. 1: Summary of reached outcomes

η [h^{-1}]	Performance measure	Normality testing			Analytic result	95% confidence interval	Difference [%]
		Mean value μ	Dispersion σ^2	P-value			
150^{-1}	*ES*	1,91470	2,09E-05	0,19124	1,90637	(1,91299; 1,91641)	0,44
	EP	0,09613	2,58E-05	0,11569	0,09677	(0,09423; 0,09803)	-0,66
125^{-1}	*ES*	1,90641	2,61E-05	0,70293	1,89568	(1,90451; 1,90832)	0,57
	EP	0,11394	4,02E-05	0,70293	0,11538	(0,11157; 0,11631)	-1,25
100^{-1}	*ES*	1,88877	2,35E-05	0,85761	1,87985	(1,88696; 1,89058)	0,47
	EP	0,14255	4,88E-05	0,70293	0,14286	(0,13994; 0,14516)	-0,21
75^{-1}	*ES*	1,86424	2,77E-05	0,78513	1,85401	(1,86228; 1,86621)	0,55
	EP	0,18555	6,00E-05	0,36904	0,18750	(0,18266; 0,18844)	-1,04
50^{-1}	*ES*	1,81009	3,82E-05	0,70293	1,80426	(1,80778; 1,81240)	0,32
	EP	0,27627	7,01E-05	0,11569	0,27273	(0,27314; 0,27939)	1,30

Source: Author

On the basis of the reached outcomes it is possible to say that there are statistical significant differences between the analytic and the simulation outcomes for all cases, because the analytic results do not fall into the corresponding confidence intervals. However these differences are not considerably in practice as we can see in the last column of table 1. In this column the differences expressed in [%] are shown between the analytic results and the estimations of the mean values of the performance measures obtained by the simulation experiments.

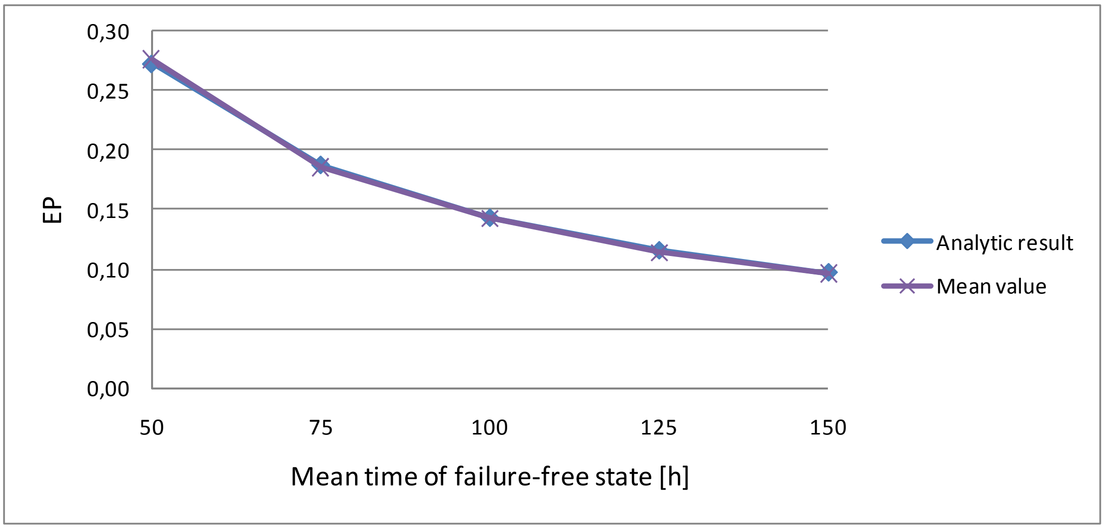
Let us illustrate the reached relations graphically – see figure 3 and 4. In both graphs two lines are drawn, the first line is denoted as “Analytic result” and it corresponds to the value obtained by the mathematical model solution and the second line “Mean value” represents the estimation of the mean value (sample average) got thanks to the simulation experiments.



Source: Author

Fig. 3 – The relationship between *ES* and $\frac{1}{\eta}$.

The figure 3 shows the relationship between the mean number of the customers in the service *ES* and the mean time of failure-free state $\frac{1}{\eta}$. We can observe that the increasing value $\frac{1}{\eta}$ causes the increasing of *ES*, this relationship could be logically expected.



Source: Author

Fig. 4 – The relationship between *EP* and $\frac{1}{\eta}$.

The dependence of the mean number of broken servers *EP* on the mean time of failure-free state $\frac{1}{\eta}$ is drawn in figure 4; this relationship is logically descending.

4. CONCLUSION

The mathematical and the simulation model of the M/M/*n*/*n* queueing system with servers subject to breakdowns and an ample repair capacity was presented in this paper. The simulation model was created by using software CPN Tools. The major part of the paper was focused on the mathematical model of the studied system and on the description of the created simulation model. At the end of the paper reached outcomes were evaluated for the unreliable M/M/3/3 queueing system with concrete parameters of random variables.

Please notice that more accurate simulation outcomes we would probably obtain by using lower unit of time, for example a second. As regards to following research on the studied queueing system we would like to find the formula for the customer rejection probability above all. Consequently the model will be upgraded to the model with an insufficient repair capacity $r<n$ that means only r servers can be repaired simultaneously.

REFERENCES

[1] MITRANI, I. L., AVI-ITZHAK, B. A many-server queue with service interruptions. *Operations Research*, 1968, Vol. 16, No. 3, pp. 628 – 638. ISSN 0030–364X.

[2] NEUTS, M. F., LUCANTONI, D. M. A Markovian queue with N servers subject to breakdowns and repairs. *Management Science*, 1979, Vol. 25, No. 9, pp. 849 – 861. ISSN 1526–5501.

[3] WARTENHORST, P. N parallel queueing systems with server breakdown and repair. *European Journal of Operational Research*, 1995, Vol. 82, No. 2, pp. 302 – 322. ISSN 0377–2217.

[4] MITRANI, I., KING, J. B. Multiserver systems subject to breakdowns: an empirical study. *IEEE Transactions on Computers*, 1983, Vol. 32, No. 1, pp. 96 – 98. ISSN 0018-9340.

[5] MADAN, K. C., ABU-DEYYEH, W., GHARAIBEH, M. On two parallel servers with random breakdowns. *Soochow Journal of Mathematics*, 2003, Vol. 29, No. 4, pp. 413 – 423. ISSN 0250–3255.

[6] YUE, D., YUE, W., YU, J., TIAN, R. A heterogeneous two-server queueing system with balking and server breakdowns. In *The Eighth International Symposium on Operations Research and Its Applications*. Zhangjiajie, China, 2009, pp. 230 – 244. ISBN 978–7–5100–0548–0/O763.

[7] WANG, K., CHANG, Y. Cost analysis of a finite M/M/R queueing system with balking, reneging, and server breakdowns. *Mathematical Methods of Operations Research*, 2002, Vol. 56, No. 2, pp. 169 – 180. ISSN 1432–5217.