DOI 10.1007/s10958-022-06124-z
*Journal of Mathematical Sciences, Vol. 267, No. 2, October, 2022*

# THE EFFECT OF OPERATION TIME OF THE SERVER ON THE PERFORMANCE OF FINITE-SOURCE RETRIAL QUEUES WITH TWO-WAY COMMUNICATIONS TO THE ORBIT

**J. Sztrik[1], Á. Tóth[1], Á. Pintér[2], and Z. Bács[3]**

In this paper a retrial queuing system is considered with the help of two-way communication where the server is subject to random breakdowns. This is a M/M/1//N type of system so the population of the source is finite. The server becoming idle enables calls the customers in the orbit (outgoing call or secondary customers). The service time of the primary and secondary customers follows exponential distribution with different rates $\mu_1$ and $\mu_2$ respectively. All the random variables included in the model construction are assumed to be totally independent of each other. The novelty of this paper is to show the effect of the different distributions of failure time on the main performance measures such as the mean waiting time of an arbitrary customer or the utilization of the service unit. In order to achieve a valid comparison a fitting process is done; thus, in case of every distribution the mean value and dispersion is the same. Graphical illustrations are given with the help of the self-developed simulation program.

## 1. Introduction

Nowadays, it is a very difficult task to analyze communications systems or create optimal designing patterns of this type of schemes owing to traffic growth and the rapidly increasing number of users. In every area of life an exchange of information is present; therefore, it is essential to develop mathematical and simulation models of telecommunication systems or alter existing ones. Retrial queues are suitable and effective tools for modeling real-life problems arising in systems such as telecommunication systems, networks, mobile networks, call-centres, etc. Many papers and books are devoted to studying many retrial queuing systems with repeated calls, for example, [7–9, 16, 17]. These models are utilized in many application fields improving the efficiency of cellular mobile networks, computer networks, local-area networks with random access protocols, and with multiple access protocols [2, 18].

We examine a special retrial queuing system with two-way communication features. The two-way communication scheme has become quite a popular topic in recent years because the operation of certain real-life systems can be compared with models of two-way communication systems. This is especially appropriate regarding call-centers when the service unit, besides attending to the incoming calls, may perform other activities in the idle state including selling, promoting, and advertising products. In our investigation once the server becomes idle after some random time it can call in a customer from the orbit and these are known as secondary customers. Of course, monitoring the utilization of the service unit plays an important role and has been investigated in many papers, for example, [1, 3, 6, 12, 20].

In some cases researchers assume that the service units are available constantly but failure or sudden acts may happen during its operation resulting in the rejection of an incoming customer. Examining

[1] Doctoral School of Informatics, Faculty of Informatics, University of Debrecen, Debrecen, Hungary, e-mail: {toth.adam,sztrik.janos}@inf.unideb.hu

[2] Institute of Mathematics, Department of Algebra and Number Theory, University of Debrecen, Debrecen, Hungary, e-mail: apinter@science.unideb.hu

[3] the Faculty of Economics and Business, Institute of Accounting and Finance, University of Debrecen, Debrecen, Hungary, e-mail: bacs.zoltan@econ.unideb.hu

Proceedings of the XXXVI International Seminar on Stability Problems for Stochastic Models, Petrozavodsk, Russia, 22–26 June, 2020. Part I.

 

devices utilized in the different fields of industry are subject to breakdowns, meaning that considering reliable operation is quite an optimistic and unrealistic approach. Likewise, in the case of wireless communication numerous elements have influence on the transmission rate; therefore, throughout the packet transmission interruption may easily occur. The characteristics of unreliability of retrial queuing systems significantly reform the operation of the system and the performance measures. Recently, retrial queuing systems with unreliable servers have been extensively studied in many papers, for example, [13–15, 19, 24, 27].

The main aim of this work is to inspect the unreliable operation of this system and to compare different distributions of failure time on performance measures such as the mean waiting time of an arbitrary customer or the total utilization of the service unit. The present paper is the continuation of [25] where the server was reliable, but now the considered system does have a server, which can break down from time to time. In order to gather the desired performance measures we developed a simulation model using SimPack [10], which is a collection of C/C++ libraries and executable programs for computer simulation. In this collection various algorithms are supported connected with simulation, including discrete event simulation, continuous simulation, and combined (multi-model) simulation. In many cases providing exact formulas is problematic or almost impossible; thus, simulation is a very good alternative to approximating the performance measures, applying as many distributions as we desired. The developer has the freedom of how and what performance measures are calculated. The novelty of this paper is to present a sensitivity analysis of failure time on the main measures using various distributions. We demonstrate results by graphically depicting interesting phenomena of sensitivity problems.

## 2. Model description and notations

As shown in Fig. 1 the considered system is a finite source queuing system with retrials that contains a nonreliable server. $N$ customers populate the source, where every customer generates request (primary customers in the system) with rate $\lambda/N$, so that each inter-arrival time is exponentially distributed with parameter $\lambda/N$. It should be noted that our model does not comprise queues; thus, an incoming customer occupies the server when it is available and not busy. The service time of primary customers follows exponential distribution with parameter $\mu_1$. After a successfully executed service the customer returns to the source. Otherwise, when an arriving customer (either from the source or orbit) encounters the server in a busy or failed state then the requests are forwarded to the orbit. In the orbit the customer may initiate attempts to get its service requirement after an exponentially distributed random time with parameter $\sigma/N$. It is supposed that a server breaks down according to gamma, hypo-exponentially, hyper-exponentially, Pareto, and lognormal distribution with different parameters but with the same mean value. The repair process starts instantaneously upon failure of the server. Repair time is also an exponentially distributed random variable with parameter $\gamma_2$. When the busy server gets into a failed state then the customer is transmitted immediately to the orbit. Even though the service unit is not available, all the customers residing in the source are able to generate requests but these find themselves in the orbit. However, an idle server can make an outgoing call to the customers in the orbit, which is performed after some random time. This random variable is exponentially distributed with rate $\tau$. The service time of these secondary customers follows exponential distribution with parameters $\mu_2$. The reason for using rates $\lambda/N$ and $\sigma/N$ is that in [22, 23] similar systems were evaluated by an asymptotic method where $N$ tends to infinity. In these papers it was shown with proof that the number of customers in the systems follows a normal distribution.

It is assumed that all the random variables in model creation are totally independent of each other.

## 3. Simulation results

We used a statistic module class that originates from an adaptation of the statistics package written by Andrea Francini in 1994 [11]. This class is a statistical analysis tool with which it is feasible to perform

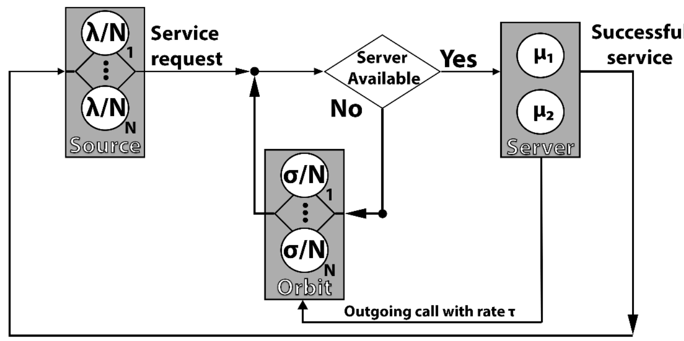


**Fig. 1.** System model.

**Table 1.** Numerical values of model parameters

| N | $\gamma_2$ | $\sigma/N$ | $\mu_2$ | $\nu$ |
|---|---|---|---|---|
| 100 | 1 | 0.01 | 1.1 | 0.02 |

a quantitative estimate of the mean and variance values of the observed variables by applying the method of batch mean. It gathers a sequence of independent samples by aggregating $n$ successive observations of a steady-state simulation. It is one of the common methods for establishing a confidence interval for the steady-state mean of a process. To assure that the sample averages would be approximately independent, sizable batches are required. More about the batch mean method can be found, for example, in [4,5,21]. The simulations are performed with the confidence level 99.9%. The relative half-width of the confidence interval required to stop the simulation run is 0.00001.

The applied values of the input parameters are shown in Table 1. In [25] the server was reliable and now we want to show the effect of the nonreliable operation of the system. The service time of primary customers follows gamma distribution with parameters $\alpha = \beta = 0.04$ to make a comparison with the results of [25]. Figure 2 reveals how the mean waiting time starts to increase and then decreases despite the increment of arrival intensity. But curiously, when the failure rate is low it results in a lower mean waiting time than in the case of reliable operation. When the server breaks down the customer is sent back to the orbit and its service does not continue. After it becomes operational again it can call a customer from the orbit.

### 3.1. Squared coefficient of variation is greater than one

In this section the parameters of failure time in every distribution were selected in a way in which the mean value and the dispersion would be equal. To do that, a fitting process has to be done that can be found in the following paper [26]. Throughout this section we work with four different distributions to ascertain their effect on the performance measures. Using hyper-exponential distribution ensures that the squared coefficient of the variation is more than one. Table 3 demonstrates the input parameters of the various distributions. Similarly, Table 2 displays the applied values of the other parameters.

Figure 3 shows the steady-state distribution using different distributions of failure time. Taking a

**Table 2.** Used numerical values of model parameters

| N | $\lambda/N$ | $\gamma_2$ | $\sigma/N$ | $\mu$ | $\mu_2$ | $\nu$ |
|---|---|---|---|---|---|---|
| 100 | 0.01 | 1 | 0.01 | 1 | 1.2 | 0.02 |

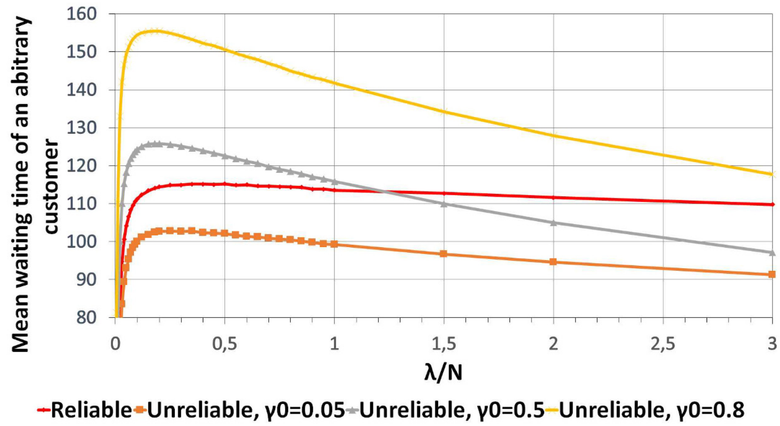


**Fig. 2.** Mean waiting time in function of arrival intensity with different failure rates.

**Table 3.** Parameters of failure time

| Distribution | Gamma | Hyper-exponential | Pareto | Lognormal |
|---|---|---|---|---|
| Parameters | $\alpha = 0.6$<br>$\beta = 0.5$ | $p = 0.25$<br>$\lambda_1 = 0.41667$<br>$\lambda_2 = 1.25$ | $\alpha = 2.2649$<br>$k = 0.67018$ | $m = -0.3081$<br>$\sigma = 0.99037$ |
| Mean | 1.2 | | | |
| Variance | 2.4 | | | |
| Squared coefficient of variation | 1.6666666667 | | | |

closer look on the curves it can be observed that all of them correspond with the normal distribution. In the case of Pareto it seems that more customers are located in the system, but huge differences cannot be observed among the applied distributions.

Figure 4 displays the mean waiting time of the customers in the function of arrival intensity. After analyzing Fig. 3 it is not surprising that we find the highest values of mean waiting time at the case of Pareto but among the other distributions the difference is quite obvious. Lowest values can be found using gamma distribution. It is an interesting phenomenon that, along with the increasing arrival intensity, after a while the mean waiting time starts to decrease. It is a general characteristic of retrial queuing systems with finite source and with suitable parameter settings.

In Fig. 5 we can see how the total utilization of the service unit increases with the increment of arrival intensity. Total utilization of the server includes the service of primary and secondary customers and all the interrupted ones when a busy server suffers a failure. In every case of the applied distributions the obtained values differ from each other in a small compass, even though in Fig. 4 the differences are quite significant. It can be stated that the total utilization of the server is nearly identical respecting the investigated distributions.

### 3.2. Squared coefficient of variation is less than one

After noticing the results of the previous section we were intrigued by how the performance measures modify with the change of the parameters of the failure time. Now the parameters were chosen so that the squared coefficient of variation should be less than one. We exchange hyper-exponential with hypo-exponential distribution because in the case of hypo-exponential distribution the squared coefficient of variation is always less than one. We will go through the same figures as in the previous section but by using the new parameters of failure time, which is shown in Table 4. The other parameters remain the same (see Table 1).

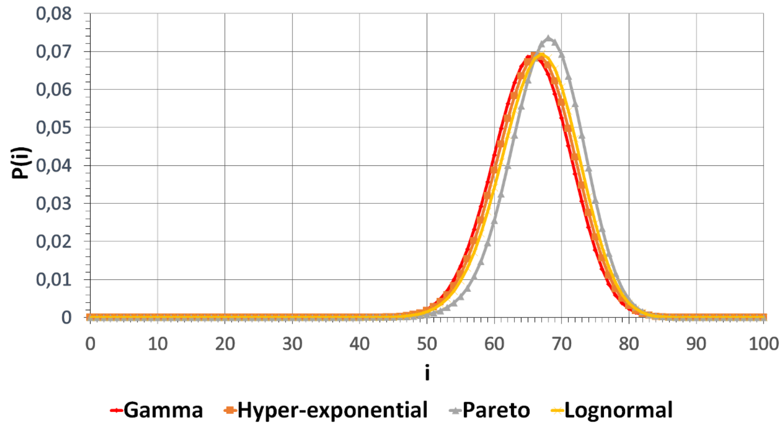


**Fig. 3.** Comparison of steady-state distributions.

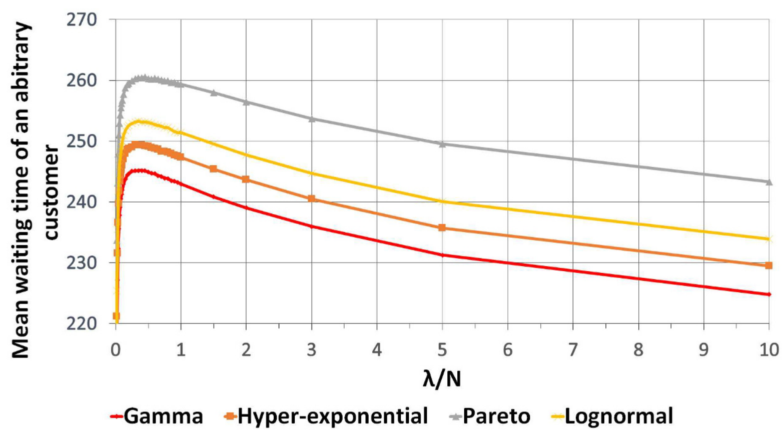


**Fig. 4.** Mean waiting time vs. arrival intensity.

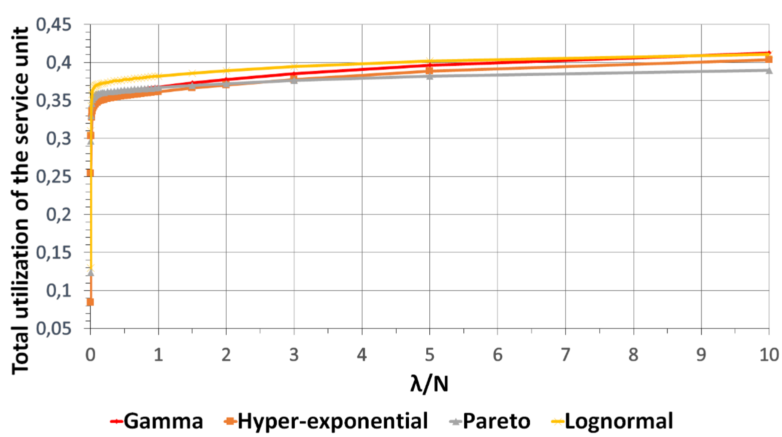


**Fig. 5.** Total utilization of the server vs. arrival intensity.

Figure 6 presents the comparison of the steady-state distributions and now the squared coefficient variation is less than one. The curves almost perfectly cover each other, meaning that on average regardless of the utilized distribution of failure time the customers in the system are the same. Compared

**Table 4.** Parameters of failure time

| Distribution | Gamma | Hypo-exponential | Pareto | Lognormal |
|---|---|---|---|---|
| Parameters | $\alpha = 1.3846$<br>$\beta = 1.1538$ | $\mu_1 = 1$<br>$\mu_2 = 5$ | $\alpha = 2.5442$<br>$k = 0.7283$ | $m = -0.0894$<br>$\sigma = 0.7373$ |
| Mean | 1.2 | | | |
| Variance | 1.04 | | | |
| Squared coefficient of variation | 0.722222 | | | |

with Fig. 3, even the averages are identical, which is around 65–66.

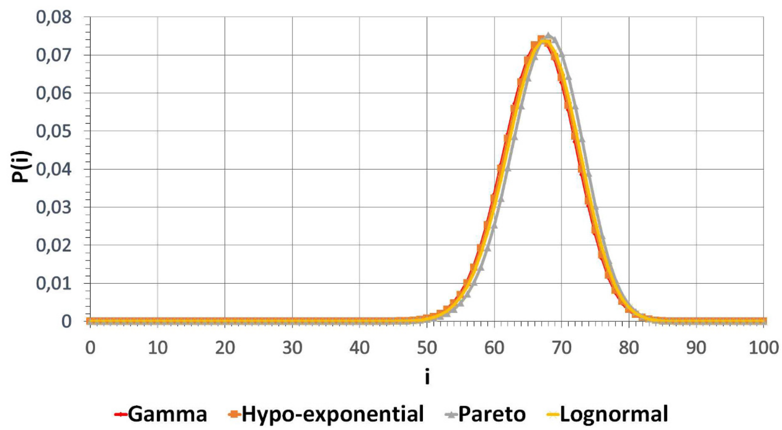


**Fig. 6.** Comparison of steady-state distributions.

In Fig. 7 it can be seen how the mean waiting of a customer forms besides increasing arrival intensity. By examining closer the curves it is observable that they are much closer to each other compared with Fig. 4, however, minor differences are there among the applied distributions. Similar to Fig. 4 the highest values are obtainable in the case of Pareto. When the squared coefficient of variation is less than one in terms of every distribution the values of mean waiting are higher as in the previous section.

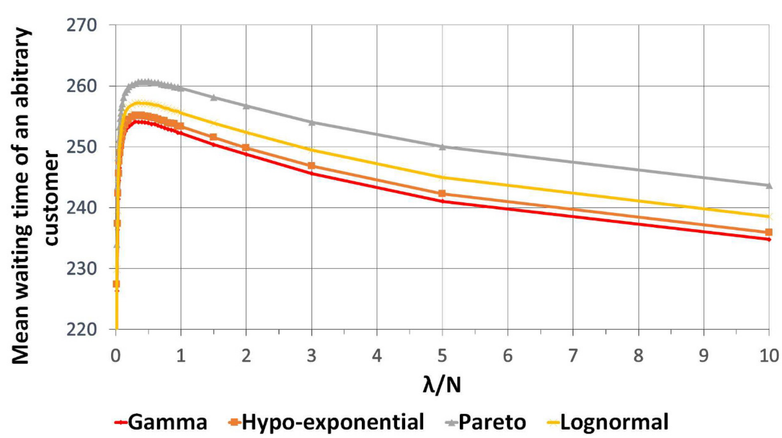


**Fig. 7.** Mean waiting time vs. arrival intensity.

Figure 8 demonstrates the utilization of the service unit in the function of arrival intensity. Among the various distributions it is nearly the same, except fot Pareto, where the utilization is considerably smaller. The same tendency can be detectable compared with previous sections so every distribution possesses approximately the same value of utilization. Of course, the utilization increases as the arrival intensity grows.

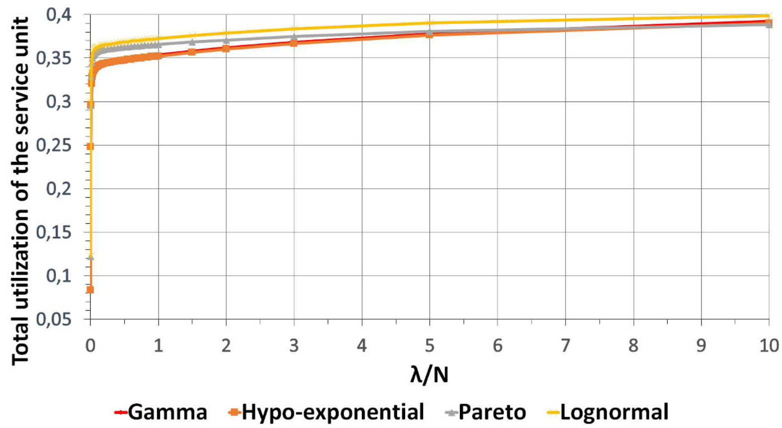


**Fig. 8.** Total utilization of the server vs. arrival intensity.

## 3.3. Conclusions

A finite-source retrial queuing system with the help of two-way communication is presented with an unreliable server, which may summon customers from the orbit after a random idle time. We show the effect of unreliable operation with different failure rates on the mean waiting time of an arbitrary customer. Interestingly, we find lower values in the case of an unreliable server with a small failure rate than with a reliable server because the service of outgoing customers is faster and the number of interruptions is not that high. To carry out a sensitivity analysis we integrated several random number generators into the simulation program to analyze the effect of different distributions on the main performance measures. When the squared coefficient of variation was greater than one having the same mean and variance among the values of performance measures, such as the mean waiting time of a customer disparity appears but the influence was very slight when it was less than one. In the future further modifications are planned to be done such as dealing with more distributions or complementing the system with other features as blocking.

## Acknowledgments

The research was financed by the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary, within the framework of the NKFIH-1150-6/2019 thematic programme of the University of Debrecen.

## REFERENCES

1. Z. Aksin, M. Armony, and V. Mehrotra, “The modern call center: A multi-disciplinary perspective on operations management research,” *Prod. Oper. Manag.*, **16**, No. 6, 665–688 (2007).
2. J. R. Artalejo and A. Gomez-Corral, *Retrial Queueing Systems: A Computational Approach*, Springer, Berlin (2008).
3. L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao, “Statistical analysis of a telephone call center: A queueing-science perspective,” *J. Amer. Statist. Assoc.*, **100**, No. 469, 36–50 (2005).
4. E. Carlstein, “The use of subseries values for estimating the variance of a general statistic from a stationary sequence,” *Ann. Statist.*, **14**, No. 3. 1171–1179 (1986).

5. E. J. Chen and W. D. Kelton, "A procedure for generating batch-means confidence intervals for simulation: Checking independence and normality," *Simulation*, **83**, No. 10, 683–694 (2007).
6. V. Dragieva and T. Phung-Duc, "Two-way communication M/M/1//N retrial queue," in: *International Conference on Analytical and Stochastic Modeling Techniques and Applications*, Springer (2017), pp. 81–94.
7. V. I. Dragieva, "Number of retrials in a finite source retrial queue with unreliable server," *Asia-Pac. J. Oper. Res.*, **31**, No. 2, 23 (2014).
8. G. I. Falin and J. R. Artalejo, "A finite source retrial queue," *Eur. J. Oper. Res.*, **108**, 409–424 (1998).
9. D. Fiems and T. Phung-Duc, "Light-traffic analysis of random access systems without collisions," *Ann. Oper. Res.*, 1–17 (2017).
10. P. A. Fishwick, "Simpack: Getting started with simulation programming in C and C++," in: *Winter Simulation Conference* (1992), pp. 154–162.
11. A. Francini and F. Neri, "A comparison of methodologies for the stationary analysis of data gathered in the simulation of telecommunication networks," in: *Proceedings of MASCOTS 96 — 4th International Workshop on Modeling, Analysis and Simulation of Computer and Telecommunication Systems* (1996), pp. 116–122.
12. N. Gans, G. Koole, and A. Mandelbaum, "Telephone call centers: Tutorial, review, and research prospects," *Manufactur. Serv. Oper. Manag.*, **5**, No. 2, 79–141 (2003).
13. N. Gharbi and C. Dutheillet, "An algorithmic approach for analysis of finite-source retrial systems with unreliable servers," *Comput. Math. Appl.*, **62**, No. 6, 2535–2546 (2011).
14. N. Gharbi and M. Ioualalen, "GSPN analysis of retrial systems with servers breakdowns and repairs," *Appl. Math. Comput.*, **174**, No. 2, 1151–1168 (2006).
15. N. Gharbi, B. Nemmouchi, L. Mokdad, and J. Ben-Othman, "The impact of breakdowns disciplines and repeated attempts on performances of small cell networks," *J. Comp. Sci.*, **5**, No. 4, 633–644 (2014).
16. A. Gomez-Corral and T. Phung-Duc, "Retrial queues and related models," *Ann. Oper. Res.*, **247**, No. 1, 1–2 (2016).
17. W. Jinting, "Reliability analysis M/G/1 queues with general retrial times and server breakdowns," *Progr. Nat. Sci.*, **16**, No. 5, 464–473 (2006).
18. J. Kim and B. Kim, "A survey of retrial queueing systems," *Ann. Oper. Res.*, **247**, No. 1, 3–36 (2016).
19. A. Krishnamoorthy, P. K. Pramod, and S. R. Chakravarthy, "Queues with interruptions: a survey," *TOP*, **22**, No. 1, 90–320 (2014).
20. A. Kuki, J. Sztrik, Á Toth, and T. Bérczes, "A contribution to modeling two-way communication with retrial queueing systems," in: *Information Technologies and Mathematical Modelling. Queueing Theory and Applications*, Springer, Berlin (2018), pp. 236–247.
21. A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill (1991).
22. A. Nazarov, J. Sztrik, and A. Kvach, "A survey of recent results in finite-source retrial queues with collisions," in: *Information Technologies and Mathematical Modelling. Queueing Theory and Applications*, Springer (2018), pp. 1–15.
23. A. Nazarov, J. Sztrik, A. Kvach, and T. Berczes, "Asymptotic analysis of finite-source M/M/1 retrial queueing system with collisions and server subject to breakdowns and repairs," *Ann. Oper. Res.*, **277**, No. 2, 213–229 (2019).

24. J. Roszik, "Homogeneous finite-source retrial queues with server and sources subject to breakdowns and repairs," *Ann. Univ. Sci. Budap. Rolando Eötvös, Sect. Comput.*, **23**, 213–227 (2004).

25. J. Sztrik, Á. Tóth, Á. Pintér, B. Zoltán, "Simulation of finite-source retrial queues with two-way communications to the orbit," in: *Information Technologies and Mathematical Modelling. Queueing Theory and Applications*, A. Dudin, A. Nazarov, and A. Moiseev (eds.), Springer, Cham (2019), pp. 270–284.

26. A. Toth, J. Sztrik, A. Kuki, T. Berczes, and D. Effosinin, "Reliability analysis of finite-source retrial queues with outgoing calls using simulation," in: *International Conference on Information and Digital Technologies (IDT)* (2019), pp. 504–511.

27. J. Wang, "Analysis of the finite source retrial queues with server breakdowns and repairs," *J. Indust. Manag. Optim.*, **7**, No. 3, 655–676 (2011).