Reliability Analysis of a Two-Way Communication System with Searching for Customers

Attila Kuki, Tamas Berczes, Janos Sztrik, Adam Toth
Faculty of Informatics
University of Debrecen
Debrecen, Hungary
{kuki.attila, berczes.tamas, janos.sztrik, toth.adam}@inf.unideb.hu

***Abstract*—In this paper a special two-way communication system is modelled by the help of finite and infinite sources queueing system with retrial. An unreliable system is considered. The server may subject to random breakdowns. Customers from the finite source are the first order or regular customers, while the customers from the infinite source are the second order or the invited customers. The novelty of this paper is to investigate and model of a non-reliable two-way communication system with this special case of searching for the customers. The first order customers reach the server according to an exponentially distributed request generation time. In case of a busy server, they are able to retry their requests. In case of an idle server, the second order customers are called for service. Their inter-arrival times also subject to the exponential distribution. The effect of the breakdown and repair intensities are also investigated. From the system balance equations the steady state probabilities can be obtained. The MOSEL-2 tool is used for these calculations. By the help of these probabilities the common performance measures are calculated and displayed.**

I. Introduction

General queueing theory has been investigated since decades for modeling of various fields of computer science, manufacturing processes, telecommunication systems, etc. As the complexity of the considered system has been increased rapidly, developing new approaches of queueing models were necessary. One of the well-known and widely spread models were the retrial queueing systems. The main demand for these models arose from the telephone switching and call centers . In case of busy lines or operators, the incoming call is not lost, but it is sent to a virtual waiting facility, to the orbit, and can retry the call again. These models were investigated by Falin, Templeton, Artalejo and more authors [1], [2], [3], [4], [5] [6]. Large part of the models assumes, that the customers generate their calls or request from a finite number of population. These assumption leads to study the finite source models [7], [2]. In addition, the considered real-life systems are unfortunately non-reliable, that is the server or other parts of the systems can lose their efficiency or may breakdown. These type of a non-reliable systems were investigated e.g. in [8], [9], [10], [11]. Beside the mentioned retrial queueing system, a new general model was developed for not to lose the customers, who are not able or not disposed to wait the service (in the queue or in the orbit). They can register for the service and later the idle state system can call for these customers. This and similar real-life demands were the motivation for developing the two-way communication systems. The key assumption is, that the idle server makes an outgoing call for the customers. One of the first paper on the retrial queueing system with two-way communication was presented by Falin [12]. Later several authors provided their contributions to this type os models [13], [14], [15], [16], [17], [18], [19]. For example, Artalejo at al. introduced models with exponential service times with different parameters for incoming and outgoing calls [20]. The M/G/1 model with the assumption of distinct arbitrary service distribution for both incoming and outgoing calls was introduced also by Artalejo and Phung-Duc [21]. In business and economic application fields (e.g. trade and IT companies) where the agent can promote their new services, products, discounts, etc. it is very important to increase the performance and the utilization of the core facility (server) of the system [22], [23], [24], [5], [25], [26]. This paper deals with a special case of searching for the customers, and in the background an unreliable server with breakdowns and repairs is supposed. Two types of sources are considered. The organization has a finite number of goodwill customers. They are the first order customers, making primary calls towards the organization (server). These clients are served according to the common retrial queueing discipline. The idle periods of the server is utilized for making outgoing calls towards the customers in the second, infinite source. The clients in this infinite source (second order customers) will contact the organization with some special interest. In case of a busy server (meanwhile another regular customer arrived), this special second order customer is treated as a non-preemptive priority client. In this model there is no distinction made between the service times for the two types of calls. The server is non-reliable, it is subject to random breakdowns. Different cases for busy time breakdown and repair is considered. The remaining parts of this work contain the followings. In Section II the model definition, the underlying Markovian process with 2 dimensions and the applied parameters are described. In Section III the steady-state probabilities are considered, and some performance measures (utilization, response times, etc.) are provided by the help of MOSEL-2 tool. At the end of the paper the results are summarized in a Conclusion.

 978-1-7281-1401-9/19/$31.00

II. Description of the model

The considered system is modelled by a finite and infinite source retrial queueing system with a single server. The functionality of the model is displayed on Figure 1.

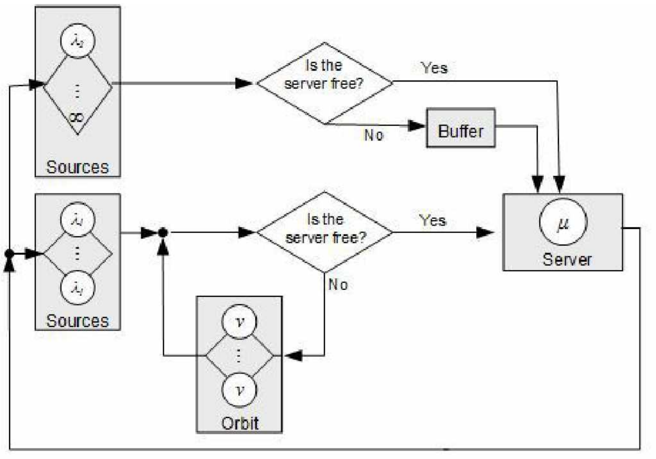

Fig. 1. The system model

The model has two sources. The first one is finite, the number of customers is N. They are the first order customers. These customers generate a job towards the server with an exponentially distributed inter-request time. The generation rate is λ_1. The job arrives to the server. If the server is idle, the service starts immediately. After the service, the job goes back to the source. The service time is again exponentially distributed with parameter μ. When the server is busy at the time of the arrival of the request, the job is transferred to the orbit. The orbit is a virtual waiting room for customers. The maximum size of the orbit is N. From the orbit the jobs after a random time interval keep retrying their request to the server until they are served. The inter-request times from the orbit are also exponentially distributed. The retrial rate is ν. When a customer generates a request, this customer is not able generate an other request during the response time of the job generated for the first. When after servicing this job is transferred back to the source the customer can generate a new request. The system has an infinite number of sources, as well. They are the second order customers. These customers generate triggered requests only. The idle server makes a call towards this infinite source, and the jobs in the source generate a request to be served. The distribution of the inter-generation times are exponential, with parameter λ_2. During the time interval between the call of the server and the request generation time, a new customer may arrive to the server from the finite source or from the orbit. In case, when the server is idle at the time of the arrival of a second order customer, the service starts immediately. The service times are exponentially distributed with parameter of μ, the same parameter as for the first order customers. When a second order customer finds the server busy, several working modes can be considered.

- The second order job is transferred back to the infinite source,
- The second order job takes place in a priority buffer. When the server becomes idle, the service of this job will start.

In this model the single server is an unreliable server, it may subject to breakdown. It is supposed, that there is no distinction between idle or busy time breakdowns. When the server is up, it will breakdown after a random time with exponentially distribution. The breakdown intensity is γ_0. In case of a breakdown, a repair process starts immediately. The repair time is exponentially distributed with parameter γ_1. When a first order customer finds the server down, it will be transferred to the orbit. A second order customer also may arrive. The idle server makes a call for the customers, and during the request generation time (with parameter λ_2) a breakdown might occur. In this situation different cases can be investigated.

- The second order job is transferred back to the infinite source,
- The second order job takes place in a priority buffer. When the server becomes up, the service of this job will start.

The server may breakdown in a busy state, as well. A first order or a second order customer is under service at the time of breakdown. For the first order customers:

- The first order job is transferred to the orbit,
- The first order job is transferred back to the source,
- The first order job remains at the server. The service will start again when the repair is finished. (because of the exponentially distributed service times, the restarted or the continued services have the same characteristics),

In this case the first order customer is sent to the orbit. this customer can retry its request from the orbit by the method described above. For the second order customers there are also some cases to be investigated.

- The second order job remains at the server. The service will start again when the repair is finished. (the equality of continuing os restarting is mentioned above),
- The second order job is sent back to the infinite source.

Let us denote $O(t)$ and $S(t)$ the number of requests in the orbit and the state of the server at a given time point of t.

Let us define the state of the server by $S(t)$, that is

$$S(t) = \begin{cases} 0, & \text{when the server is idle} \\ 1, & \text{when the server is busy} \\ & \text{with a first order customer} \\ 2, & \text{when the server is busy} \\ & \text{with a second order customer} \\ 3, & \text{when the server is down} \end{cases}$$

.

It is easy to see, that the maximum size of the orbit is N. (For example,in case of an extremely long service of a second order request may cause all of the first order customers sent to the orbit.) From here, the state space representation of the Markovian-process $(S(t), O(t))$ can be described as a set of $\{0, 1, 2, 3\} \times \{0, 1, 2, ..., N\}$ elements. Although, the system has an infinite source, the maximum number of the customers in the system is $(N+1)$ (N in the orbit and one second order customer under service), there is no stability problems regarding the system. The state space is finite.

All of the times, time intervals considered in the model, are exponentially distributed and totally independent from each other.

Let us consider the non-buffered model, when a second order customer under service is sent back to the source in case of breakdown. For this case the system balance equations for the steady-state system probabilities can be formulated as follows:

$$p_{i,j} = \lim_{t\to\infty} P(S(t)=i, O(t)=j),$$
$$i = 0,1,2,3 \text{ and } j = 0,1,..N$$

$$[(N-j)\lambda_1 + \lambda_2 + j\nu + \gamma_0]\, p_{0,j} = \mu p_{1,j} + \mu p_{2,j} + \gamma_1 p_{3,j}$$

$$[(N-j-1)\lambda_1 + \mu + \gamma_0]\, p_{1,j} =$$
$$= (N-j)\lambda_1 p_{0,j} + (j+1)\nu p_{0,j+1}$$

$$[(N-j)\lambda_1 + \mu + \gamma_0]\, p_{2,j} = \lambda_2 p_{0,j}$$

$$[(N-j)\lambda_1]\, p_{3,j} = \gamma_0 p_{0,j} + \gamma_0 p_{1,j-1} + \gamma_0 p_{2,j}$$

with $p_{1,-1} = p_{0,N+1} = 0$.

Similarly, consider the case in the non-buffered model, when a first order customer under service remains at the server in case of breakdown. Because the exponentially distributed service time, the restarted or the continued services have the same characteristics. For this case the system balance equations for the steady-state probabilities can be formulated as follows:

$$p_{i,j} = \lim_{t\to\infty} P(S(t)=i, O(t)=j),$$
$$i = 0,1,2,3 \text{ and } j = 0,1,..N$$

$$[(N-j)\lambda_1 + \lambda_2 + j\nu + \gamma_0]\, p_{0,j} = \mu p_{1,j} + \mu p_{2,j} + \gamma_1 p_{3,j}$$

$$[(N-j-1)\lambda_1 + \mu + \gamma_0]\, p_{1,j} =$$
$$= (N-j)\lambda_1 p_{0,j} + (j+1)\nu p_{0,j+1}$$

$$[(N-j)\lambda_1 + \mu + \gamma_0]\, p_{2,j} = \lambda_2 p_{0,j}$$

$$[(N-j)\lambda_1]\, p_{3,j} = \gamma_0 p_{0,j} + \gamma_0 p_{1,j} + \gamma_0 p_{2,j}$$

with $p_{0,N+1} = 0$.

The system balance equations for the steady-state system probabilities in the other cases can be obtained by similar way.

Solving manually these balance equations is rather difficult. There exist several effective tools performing the background calculations. In this paper the MOSEL-2 tool was used. When the steady-state probabilities are calculated, this tool provides the well known performance characteristics. These measures are obtained using the following formulas.

- *Utilization 1*

$$U_1 = \sum_{o=0}^{N} P(1,o)$$

- *Utilization 2*

$$U_2 = \sum_{o=0}^{N} P(2,o)$$

- *Average number of jobs in the orbit*

$$\overline{O} = \sum_{s=0}^{3}\sum_{o=0}^{N} oP(s,o)$$

- *Average number of active primary users*

$$\overline{M} = N - \overline{O} - U_1$$

- *Average generation rate of primary users*

$$\overline{\lambda_1} = \lambda_1 \overline{M}$$

- *Mean time spent in orbit by using Little-formula*

$$\overline{W} = \frac{\overline{O}}{\overline{\lambda_1}}$$

III. Numerical results

The most important goal of these types of stochastic systems is to obtain the performance measures and system characteristics. Usually the throughput, utilization, response times, waiting times, queue length are considered. Here the utilization and waiting time in the orbit are focused.

TABLE I
NUMERICAL VALUES OF MODEL PARAMETERS

Case studies							
No.	N	λ_1	λ_2	μ	ν	γ_0	γ_1
Fig. 2	100	$x-axes$	2	3	0.05	0.01	1
Fig. 3	100	$x-axes$	2	3	0.05	0.1	1
Fig. 4	100	$x-axes$	2	3	0.05	0.01, 0.1	1
Fig. 5	100	$x-axes$	2	3	0.05	0.01, 0.1	1
Fig. 6	100	$x-axes$	2	3	0.05	0.01	1
Fig. 7	100	$x-axes$	2	3	0.05	0.1	1
Fig. 8	100	0.2	2	3	0.05	$x-axes$	1
Fig. 9	100	0.2	2	3	0.05	$x-axes$	1
Fig. 10	100	0.2	2	3	0.05	0.1	$x-axes$

There exist several methods to calculate the system measures. Solving directly the balance equations is rather difficult in most cases. Effective software tools can be used to get the steady-state system probabilities. From these probabilities

 978-1-7281-1401-9/19/$31.00 ©2019 IEEE

the performance measures can be computed directly or by the help of the considered tool. In this paper the MOSEL-2 tool is used. This is not a simulation tool. The system equations are build up and solved by one of the utilities developed for MOSEL-2. Here the SPNP (Stochastic Petri Net Program) is used (see in [27]. The following figures illustrates the most interesting numerical results. The numerical values of the applied parameters in the model are listed in Table I. Most figure compares to different cases:

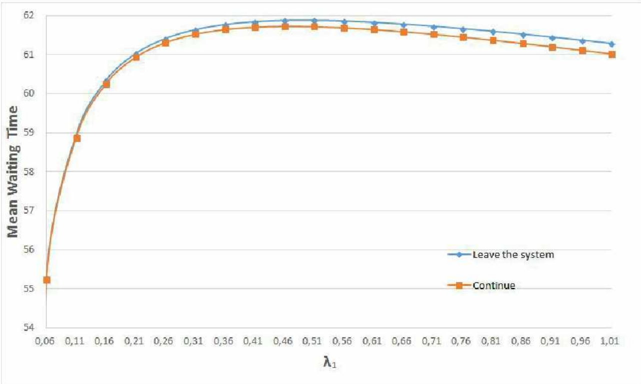

Fig. 2. Mean Waiting Time vs. λ_1

- In case of busy state breakdown, the first order and second order customers are interrupted. The first order customers are sent back to the orbit, the second order customers are sent back to the source. On figure these cases are denoted with blue lines dotted by diamonds.
- The service of both types of customers are interrupted. The customers are left at the server. After the repair their service will continue or restart. Because of the exponentially distributed service time, this difference - restart or continue - has no effect to the system characteristics. On figure these cases are denoted with orange lines dotted by squares.

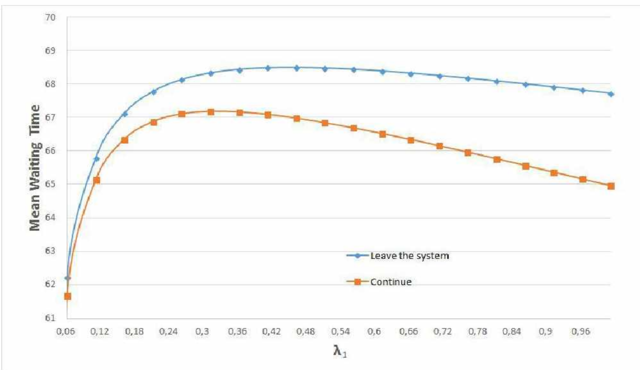

Fig. 3. Mean Waiting Time vs. λ_1

On Figure 2 the running parameter (values of x-axes) is the first order generation rate λ_1. The failure rate is small for this figure. There are not so significant differences between lines. The waiting time of the 'leave the system' case is greater, because the first order jobs goes to the orbit and they have to try again.

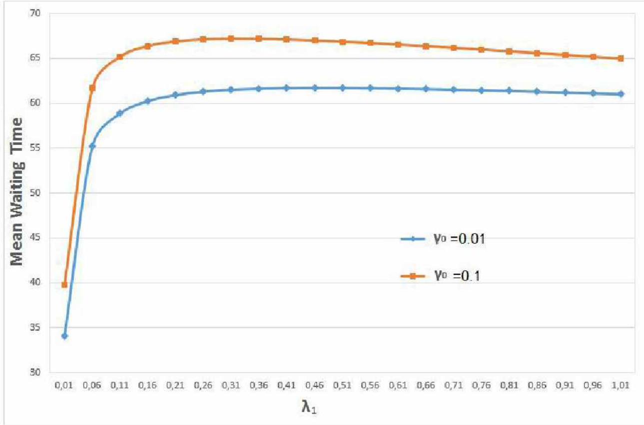

Fig. 4. Mean Waiting Time vs. λ_1

Figure 3 displays the same situation with ten times greater failure rate, which will cause a much more significant deviance between the cases. The interruption is more frequent and the first order customer are sent back to the orbit more frequently, which results higher waiting times. The two considered failure rates are compared on Figure 4 and Figure 5 for 'Continue' and for 'Leave the system' scenarios, respectively. The expected results can be seen, the waiting times are higher for greater values of failure rates.

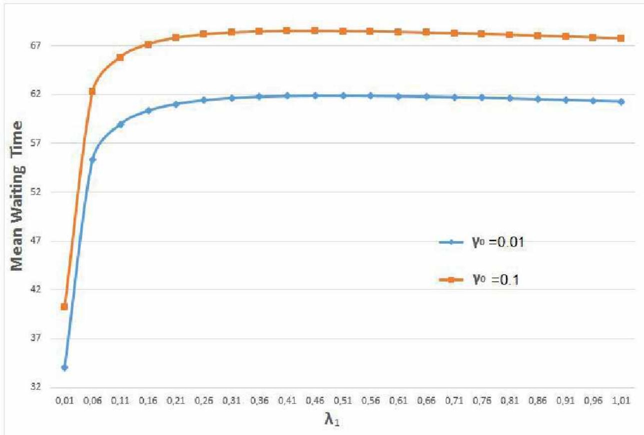

Fig. 5. Mean Waiting Time vs. λ_1

Figure 6 shows the utilization in function of the first order generation rate. The failure rate here is small, so the differences between the two scenarios are also small. The utilization is greater for the 'Continue' case, because after the repair the server state will be busy, immediately. While for the other scenario the server will be idle, and an exponential retrial, first or second order generation will take place. For Figure 7 the parameters are the same, but the Failure rate, which is again ten times greater than on the 6. Consequently, the differences in utilization are more significant.

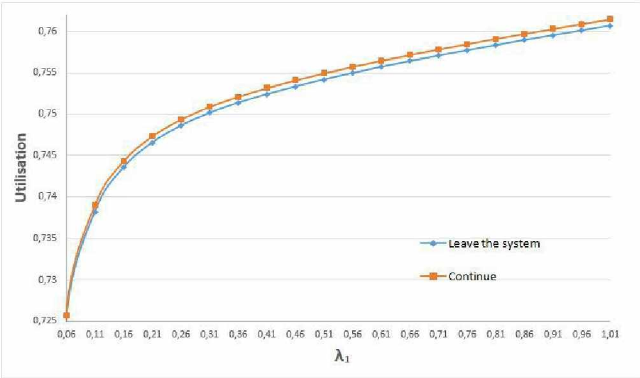

Fig. 6. Utilization vs. λ_1

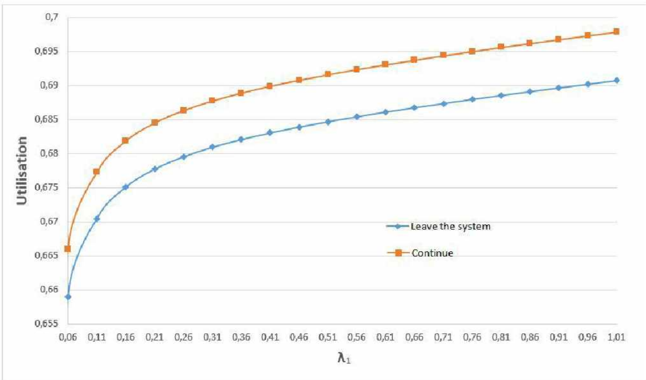

Fig. 7. Utilization vs. λ_1

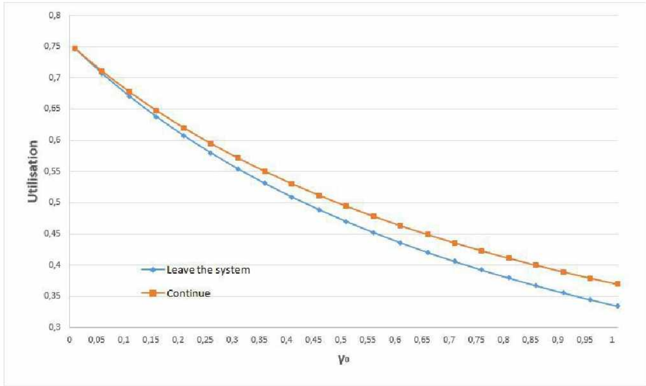

Fig. 9. Utilization vs. γ_0

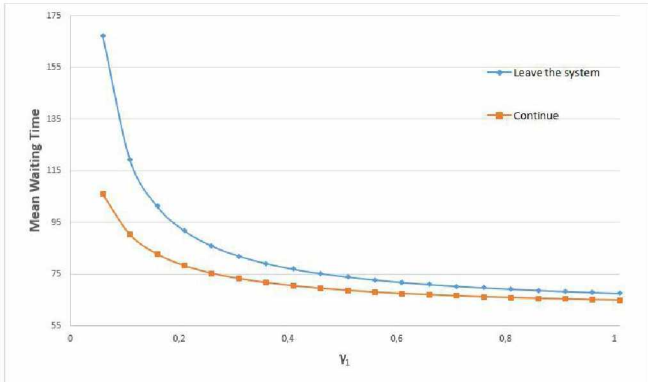

Fig. 10. Mean Waiting Time vs. γ_1

On Figures 8 and 9 the failure rate, γ_0 is the running parameter. On Figure 8 the mean waiting time is investigated. Here the parameter is modified in wider range than on Figures 2 and 3, but the tendencies are the same. The higher the failure rate is, the higher the waiting times are. Additionally, waiting times for 'Leave the system' scenario is higher, as well. Figure 9 displays the result of the utilities under the same parameters. With higher failure rate the utilization will decrease, and comparing the two scenarios, utilization is higher for the 'Continue' scenario.

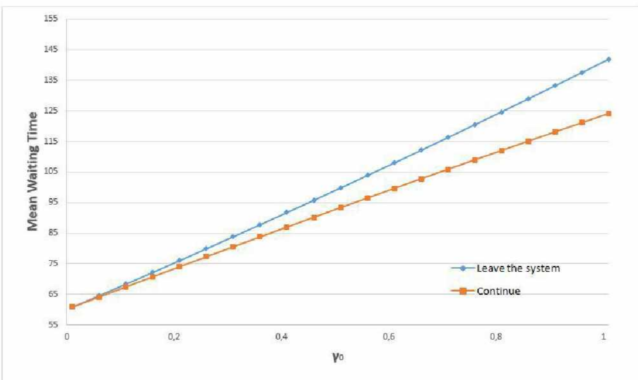

Fig. 8. Mean Waiting Time vs. γ_0

Figure 10 investigates the effect of repair rate. Since the repair rate and the average repair time are reciprocal values, higher repair rate means shorter repair time. According this, it can be seen, that for higher repair rate the waiting times will decrease. Comparing the two cases, 'Leave the system' has greater waiting times.

IV. Conclusion

A special two-way communication system was investigated here. First order customers come from a finite source, while in case of an idle server, second order customers are able to reach the system via a direct call. Different cases can be considered. For simplicity, the service rates for the first and the second order customers were supposed to be the same. Similarly, instead of different failure rate for idle server, server with first order customer, and the server with second order customer a common failure rate was considered. Of course, these distinctions can be easily made in the model, if it would be necessary. The main focus was to compare the 'Continue' and the 'Leave the system' scenarios. Based on the results displayed on the figures above, it can be stated, that the system performance (in waiting times and utilization) is better for the 'Continue' case. For the results, the buffered case of the second order customers was considered. It is closer to the real life situation. When a customer is called for service from the

978-1-7281-1401-9/19/$31.00 ©2019 IEEE

outside world, and in the meantime the server becomes busy, give the chance for the called customer to be served.

Acknowledgment

The research work of János Sztrik and Ádám Tóth was supported by the Austrian-Hungarian Bilateral Cooperation in Science and Technology project 2017-2.2.4-TeT-AT-2017-00010.

The research work of Attila Kuki and Tamás Bérczes was supported by the construction EFOP-3.6.3-VEKOP-16-2017-00002. The project was supported by the European Union,co-financed by the European Social Fund.

This publication is the result of the project implementation: Exhibition and Special Discussion Section on Information and Digital Technologies. Reg. no. 21830315. The project is co-financed by the Governments of Czechia, Hungary, Poland and Slovakia through Visegrad Grants from International Visegrad Fund. The mission of the fund is to advance ideas for sustainable regional cooperation in Central Europe.

References

[1] G. I. Falin, "Waiting time in a single-channel queuing system with repeated calls," *Moscow University Computational Mathematics and Cybernetics*, vol. 4, pp. 66–69, 1977.

[2] J. Artalejo, "Retrial queues with a finite number of sources," *Journal of the Korean Mathematical Society*, vol. 35, pp. 503–525, 1998.

[3] G. Falin and J. G. C. Templeton, *Retrial Queues*. Chapman and Hall, London, 1997.

[4] J. Templeton, "Retrial queues." *Top*, vol. 7, no. 2, pp. 351–353, 1999.

[5] N. Gans, G. Koole, and A. Mandelbaum, "Telephone call centers: Tutorial, review, and research prospects," *Manufacturing & Service Operations Management*, vol. 5, no. 2, pp. 79–141, 2003.

[6] J. Artalejo and A. G. Corral, *Retrial Queueing Systems: A Computational Approach*. Springer, 2008.

[7] G. Falin and J. Artalejo, "A finite source retrial queue," *European Journal of Operational Research*, vol. 108, pp. 409–424, 1998.

[8] J. Wang, L. Zhao, and F. Zhang, "Performance analysis of the finite source retrial queue with server breakdowns and repairs," in *Proceedings of the 5th International Conference on Queueing Theory and Network Applications*. ACM, 2010, pp. 169–176.

[9] ——, "Analysis of the finite source retrial queues with server breakdowns and repairs," *Journal of Industrial and Management Optimization*, vol. 7, no. 3, pp. 655–676, 2011.

[10] F. Zhang and J. Wang, "Performance analysis of the retrial queues with finite number of sources and service interruptions," *Journal of the Korean Statistical Society*, vol. 42, no. 1, pp. 117–131, 2013.

[11] V. I. Dragieva, "Number of retrials in a finite source retrial queue with unreliable server." *Asia-Pac. J. Oper. Res.*, vol. 31, no. 2, p. 23, 2014.

[12] G. Falin, "Model of coupled switching in presence of recurrent calls," *Engineering Cybernetics*, vol. 17, no. 1, pp. 53–59, 1979.

[13] I. Dimitriou, "A retrial queue to model a two-relay cooperative wireless system with simultaneous packet reception," in *International Conference on Analytical and Stochastic Modeling Techniques and Applications*. Springer, 2016, pp. 123–139.

[14] V. Dragieva and T. Phung-Duc, "Two-way communication M/M/1 retrial queue with server-orbit interaction," in *Proceedings of the 11th International Conference on Queueing Theory and Network Applications*. ACM, 2016, p. 11.

[15] A. Nazarov, T. Phung-Duc, and S. Paul, "Heavy outgoing call asymptotics for MMPP/M/1/1 retrial queue with two-way communication," in *Information Technologies and Mathematical Modelling. Queueing Theory and Applications: 16th International Conference, ITMM 2017, Named After AF Terpugov, Kazan, Russia, September 29-October 3, 2017, Proceedings*, vol. 800. Springer, 2017, p. 28.

[16] A. A. Nazarov, S. Paul, I. Gudkova *et al.*, "Asymptotic analysis of Markovian retrial queue with two-way communication under low rate of retrials condition," *Proceedings 31st European Conference on Modelling and Simulation,*, 2017.

[17] T. Phung-Duc and W. Rogiest, "Two way communication retrial queues with balanced call blending," in *International Conference on Analytical and Stochastic Modeling Techniques and Applications*. Springer, 2012, pp. 16–31.

[18] H. Sakurai and T. Phung-Duc, "Two-way communication retrial queues with multiple types of outgoing calls," *Top*, vol. 23, no. 2, pp. 466–492, 2015.

[19] ——, "Scaling limits for single server retrial queues with two-way communication," *Annals of Operations Research*, vol. 247, no. 1, pp. 229–256, 2016.

[20] J. R. Artalejo and T. Phung-Duc, "Markovian retrial queues with two way communication," *Journal of industrial and management optimization*, vol. 8, no. 4, pp. 781–806, 2012.

[21] J. Artalejo and T. Phung-Duc, "Single server retrial queues with two way communication," *Applied Mathematical Modelling*, vol. 37, no. 4, pp. 1811–1822, 2013.

[22] S. Aguir, F. Karaesmen, O. Z. Akşin, and F. Chauvet, "The impact of retrials on call center performance," *OR Spectrum*, vol. 26, no. 3, pp. 353–376, 2004.

[23] Z. Aksin, M. Armony, and V. Mehrotra, "The modern call center: A multi-disciplinary perspective on operations management research," *Production and operations management*, vol. 16, no. 6, pp. 665–688, 2007.

[24] 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," *Journal of the American statistical association*, vol. 100, no. 469, pp. 36–50, 2005.

[25] S. Pustova, "Investigation of call centers as retrial queuing systems," *Cybernetics and Systems Analysis*, vol. 46, no. 3, pp. 494–499, 2010.

[26] T. Wolf, "System and method for improving call center communications," Nov. 30 2017, uS Patent App. 15/604,068.

[27] K. Begain, G. Bolch, and H. Herold, *Practical Performance Modeling, Application of the MOSEL Language,*. Kluwer Academic Publisher, Boston, 2001.