



Modeling of a Two-Way Communication System with a Special Searching for Customers

Attila Kuki^(✉), Tamás Bérczes, János Sztrik, and Ádám Tóth

Faculty of Informatics, University of Debrecen, Debrecen, Hungary
{kuki.attila,berczes.tamas,sztrik.janos,
toth.adam}@inf.unideb.hu

Abstract. In this paper a special system with two-way communication is modelled by a finite and an infinite sources queueing system with retrial. The system is unreliable, 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 this unreliable system with different idle and busy breakdown intensity and different service rates for first and second order customers. In case of a busy server for first order customers, they can retry their requests. In case of an idle server, the second order customers are called for service. The effect of the breakdown and repair intensities are also investigated. The system balance equations are formulated, and the steady state probabilities can be obtained. Here the MOSEL-2 tool is used for these calculations. By the help of these probabilities the common performance measures are calculated and displayed.

Keywords: Retrial queues · Two-way communication · Unreliable system · Searching for customers

1 Introduction

Queueing theory has been investigated since decades for modeling of various problems of computer science, telecommunication systems, etc. As the complexity of the considered systems has been increased rapidly, developing new approaches of queueing models were necessary. Mainly the telephone switching centers motivated a new model. It was the retrial queueing systems. In case of busy lines or operators, the incoming call is not lost, but it is redirected to a virtual waiting facility, to the orbit, and it can retry the call again. These types of models were investigated by Falin, Templeton, Artalejo and more authors [3, 4, 10, 13, 14, 21]. Real-life situations require, that in the models the customers generate their calls or request from a finite number of population. These demands lead to study the finite source models [3, 12]. Furthermore, the considered real-life systems are unfortunately unreliable, that is the server or other parts of the

systems can lose their efficiency or may breakdown. These types of unreliable systems were investigated e.g. in [9, 22, 23, 25].

An other 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). Demands based on real applications 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 [11]. So far several authors have investigated this type of models [7, 8, 15–17, 19, 20].

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. See, e.g. in [1, 2, 6, 14, 18, 24].

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 Sect. 2 the model definition, the underlying Markovian process with 2 dimensions and the applied parameters are described. In Sect. 3 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.

2 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 Fig. 1.

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 of a single first order customer is λ_1 . 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_1 . When the server is busy, the job is transferred to the orbit. The maximum size of the orbit is N . From the orbit the jobs after a random (exponential, with parameter ν) time keep retrying their request to the server until they are served.

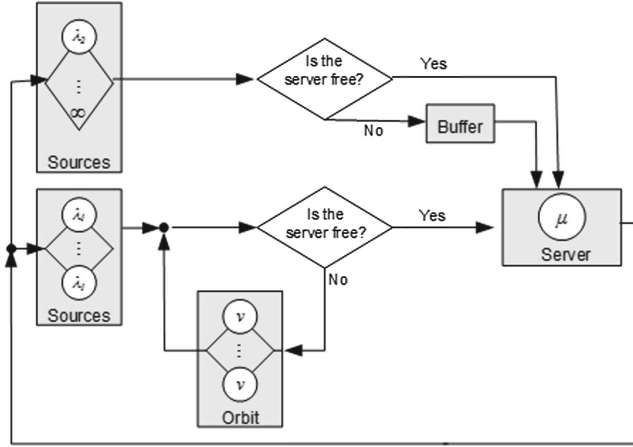


Fig. 1. The system model

The system has an infinite number of sources, as well. They are the second order customers. The idle server makes a call towards this infinite source, and the jobs in the source generate a request. The distribution of the inter-generation times are exponential, with parameter λ_2 . Here λ_2 is the generation rate from the infinite source. In case, when the server is idle at the time of arrival of a second order customer, the service starts immediately. The service times are exponentially distributed with parameter of μ_2 . 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. When the server is up, it will breakdown after a random time with exponentially distribution. The breakdown intensities are γ_0 for the idle server and γ_1 for the busy server. In case of a breakdown, a repair process starts immediately. The repair time is exponentially distributed with parameter γ_2 . 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. The first order customers can be transferred to the orbit or to the source, or the jobs may remain at

the server. The service will continue after the repair. The second order customers also may remain at the server or may 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 . From here, the state space representation of the Markovian-process $(S(t), O(t))$ can be described as a set of $\{0, 1, 2, 3\} \times \{0, 1, 2, \dots, N\}$ 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:

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

$$[(N - j)\lambda_1 + \lambda_2 + j\nu + \gamma_0] p_{0,j} = \mu_1 p_{1,j} + \mu_2 p_{2,j} + \gamma_2 p_{3,j}$$

$$[(N - j - 1)\lambda_1 + \mu_1 + \gamma_1] p_{1,j} \\ = (N - j)\lambda_1 p_{0,j} + (j + 1)\nu p_{0,j+1}$$

$$[(N - j)\lambda_1 + \mu_2 + \gamma_1] p_{2,j} = \lambda_2 p_{0,j}$$

$$[(N - j)\lambda_1 + \gamma_2] p_{3,j} = \gamma_0 p_{0,j} + \gamma_1 p_{1,j-1} + \gamma_1 p_{2,j}$$

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. The second order customer is sent back to the source. 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:

$$p_{i,j} = \lim_{t \rightarrow \infty} P(S(t) = i, O(t) = j),$$

$$i = 0, 1, 2, 3 \text{ and } j = 0, 1, \dots, N$$

$$[(N-j)\lambda_1 + \lambda_2 + j\nu + \gamma_0]p_{0,j} = \mu_1 p_{1,j} + \mu_2 p_{2,j} + \gamma_2 p_{3,j}$$

$$[(N-j-1)\lambda_1 + \mu_1 + \gamma_1]p_{1,j}$$

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

$$[(N-j)\lambda_1 + \mu_2 + \gamma_1]p_{2,j} = \lambda_2 p_{0,j}$$

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

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*

$$\bar{O} = \sum_{s=0}^3 \sum_{o=0}^N o p_{s,o}$$

– *Average number of active primary users*

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

– *Average generation rate of primary users*

$$\bar{\lambda}_1 = \lambda_1 \bar{M}$$

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

$$\bar{W} = \frac{\bar{O}}{\bar{\lambda}_1}$$

3 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 1. Numerical values of model parameters

Case studies								
No.	N	λ_1	λ_2	μ	ν	γ_0	γ_1	γ_2
Fig. 2	100	<i>x-axes</i>	2	3	0.05	0.01	0.01	1
Fig. 3	100	<i>x-axes</i>	2	3	0.05	0.1	0.01	1
Fig. 4	100	<i>x-axes</i>	2	3	0.05	0.01, 0.1	0.01, 0.1	1
Fig. 5	100	<i>x-axes</i>	2	3	0.05	0.01, 0.1	0.01, 0.1	1
Fig. 6	100	<i>x-axes</i>	2	3	0.05	0.01	0.01	1
Fig. 7	100	<i>x-axes</i>	2	3	0.05	0.1	0.1	1
Fig. 8	100	0.2	2	3	0.05	<i>x-axes</i>	<i>x-axes</i>	1
Fig. 9	100	0.2	2	3	0.05	<i>x-axes</i>	<i>x-axes</i>	1
Fig. 10	100	0.2	2	3	0.05	0.1	0.1	<i>x-axes</i>

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 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 [5]). The following figures illustrates the most interesting numerical results. The numerical values of the applied parameters in the model are listed in Table 1. Most figure compares to different cases:

- 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.

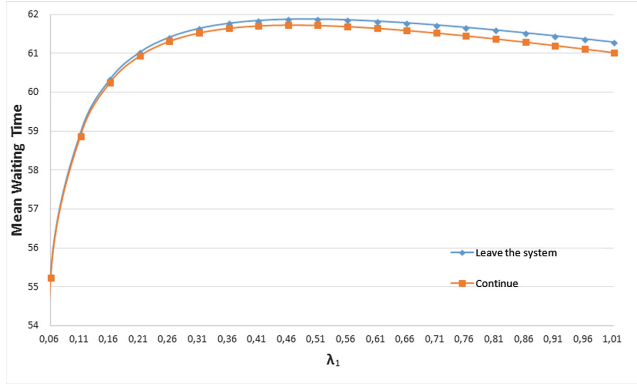


Fig. 2. Mean waiting time vs. λ_1 (Color figure online)

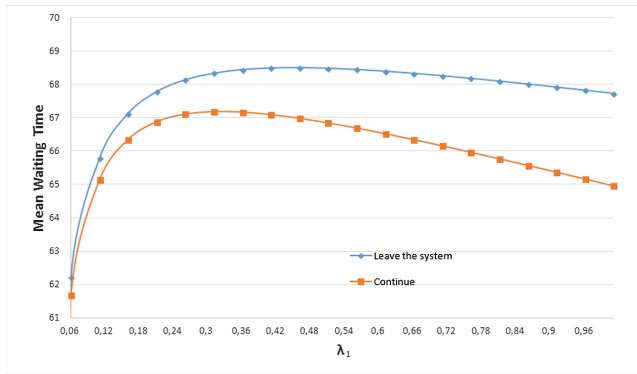


Fig. 3. Mean waiting time vs. λ_1 (Color figure online)

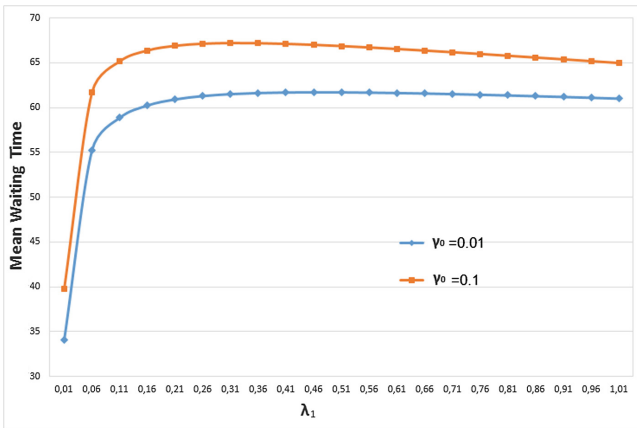


Fig. 4. Mean waiting time vs. λ_1 (Color figure online)

On Fig. 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.

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 Figs. 4 and 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.

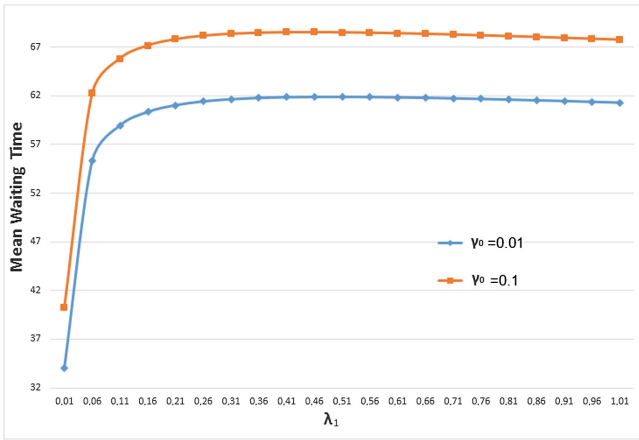


Fig. 5. Mean waiting time vs. λ_1 (Color figure online)

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 Fig. 7 the parameters are the same, but the Failure rate, which is again ten times greater than on the Fig. 6. Consequently, the differences in utilization are more significant.

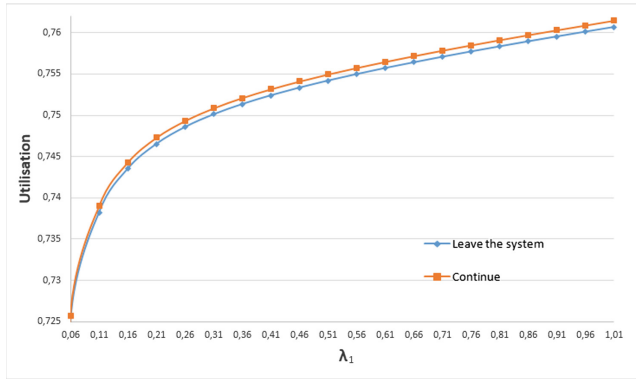


Fig. 6. Utilization vs. λ_1 (Color figure online)

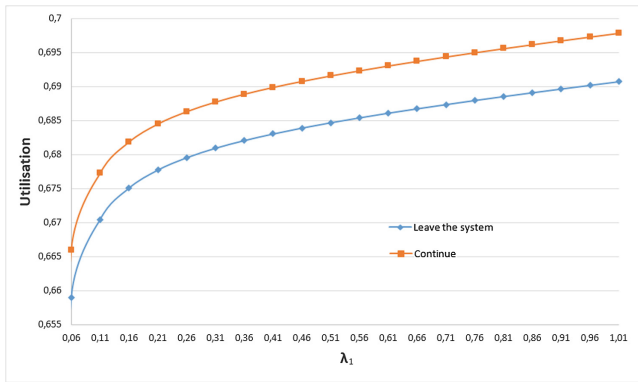


Fig. 7. Utilization vs. λ_1 (Color figure online)

On Figs. 8 and 9 the failure rate, γ_0 and γ_1 is the running parameter. The two parameters move together. On Fig. 8 the mean waiting time is considered. Here the parameter is modified in wider range than on Figs. 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.

On Fig. 9 the failure rate, γ_0 (and γ_1 with the same way) is the running parameter. With higher failure rate the utilization will decrease, and comparing the two scenarios, utilization is higher for the ‘Continue’ scenario.

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.

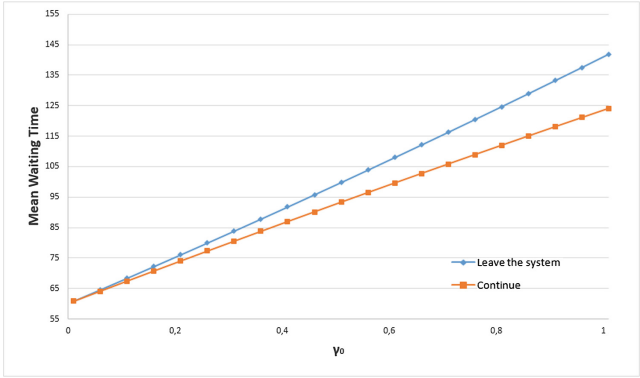


Fig. 8. Mean waiting time vs. γ_0 and γ_1 (Color figure online)

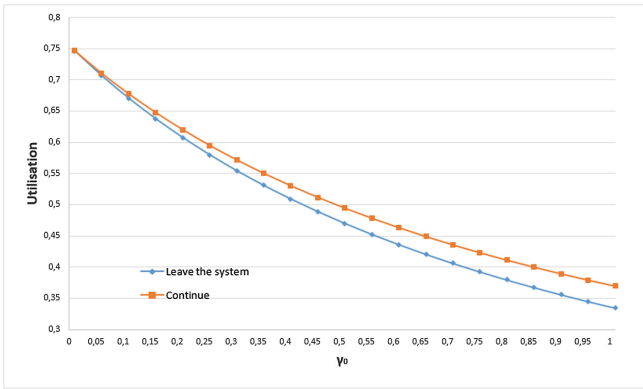


Fig. 9. Utilization vs. γ_0 and γ_1 (Color figure online)

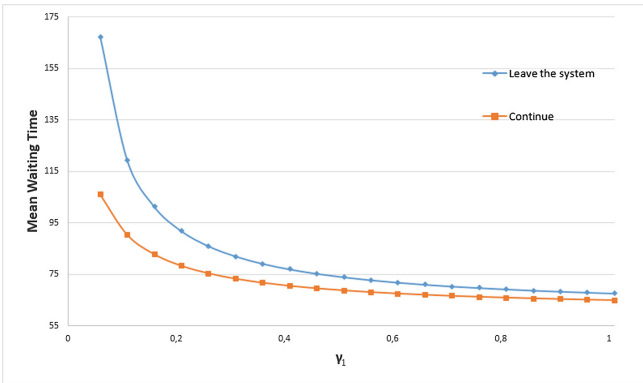


Fig. 10. Mean waiting time vs. γ_2 (Color figure online)

10. Falin, G.I.: Waiting time in a single-channel queuing system with repeated calls. *Mosc. Univ. Comput. Math. Cybern.* **4**, 66–69 (1977)
11. Falin, G.: Model of coupled switching in presence of recurrent calls. *Eng. Cybern.* **17**(1), 53–59 (1979)
12. Falin, G., Artalejo, J.: A finite source retrial queue. *Eur. J. Oper. Res.* **108**, 409–424 (1998)
13. Falin, G., Templeton, J.G.C.: *Retrial Queues*. Chapman and Hall, London (1997)
14. Gans, N., Koole, G., Mandelbaum, A.: Telephone call centers: tutorial, review, and research prospects. *Manuf. Serv. Oper. Manag.* **5**(2), 79–141 (2003)
15. Nazarov, A., Phung-Duc, T., Paul, S.: Heavy outgoing call asymptotics for MMPP/M/1/1 retrial queue with two-way communication. In: Dudin, A., Nazarov, A., Kirpichnikov, A. (eds.) *ITMM 2017. CCIS*, vol. 800, pp. 28–41. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68069-9_3
16. Nazarov, A.A., Paul, S., Gudkova, I., et al.: Asymptotic analysis of Markovian retrial queue with two-way communication under low rate of retrials condition. In: *Proceedings 31st European Conference on Modelling and Simulation*, pp. 687–693 (2017)
17. Phung-Duc, T., Rogiest, W.: Two way communication retrial queues with balanced call blending. In: Al-Begain, K., Fiems, D., Vincent, J.-M. (eds.) *ASMTA 2012. LNCS*, vol. 7314, pp. 16–31. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30782-9_2
18. Pustova, S.: Investigation of call centers as retrial queuing systems. *Cybern. Syst. Anal.* **46**(3), 494–499 (2010)
19. Sakurai, H., Phung-Duc, T.: Two-way communication retrial queues with multiple types of outgoing calls. *Top* **23**(2), 466–492 (2015)
20. Sakurai, H., Phung-Duc, T.: Scaling limits for single server retrial queues with two-way communication. *Ann. Oper. Res.* **247**(1), 229–256 (2016)
21. Templeton, J.: Retrial queues. *Top* **7**(2), 351–353 (1999). <https://doi.org/10.1007/BF02564732>
22. Wang, J., Zhao, L., Zhang, F.: 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*, pp. 169–176. ACM (2010)
23. Wang, J., Zhao, L., Zhang, F.: Analysis of the finite source retrial queues with server breakdowns and repairs. *J. Ind. Manag. Optim.* **7**(3), 655–676 (2011). <https://doi.org/10.3934/jimo.2011.7.655>
24. Wolf, T.: System and method for improving call center communications. US Patent App. 15/604,068, 30 November 2017
25. Zhang, F., Wang, J.: Performance analysis of the retrial queues with finite number of sources and service interruptions. *J. Korean Stat. Soc.* **42**(1), 117–131 (2013). <https://doi.org/10.1016/j.jkss.2012.06.002>