



# Analyzing the Effect of Catastrophic Breakdowns with Retrial Queues in a Two-Way Communication System

Attila Kuki<sup>(✉)</sup>, Tamás Bérczes<sup></sup>, and János Sztrik<sup></sup>

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

**Abstract.** A two-way communication system is modeled in this paper. A retrial queueing system with a finite and an infinite sources is used in the model. The system has two sources. The first source is finite, the second source is infinite. Jobs from the first source are the primary jobs (requests). They can be called as first order job, as well. Jobs from the second source are the secondary jobs. They can be called as second order job, as well. In case of an idle server, the second order customers are called for service. This situation is said as a special search for customers.

The non-reliable server is subject to random breakdowns. Two types of breakdowns are considered: the regular breakdown, when the first or second order customer under service is sent back to the orbit or the infinite source, respectively, and the catastrophic breakdown, when all of the requests at the server and in the orbit are sent back to the corresponding sources. The novelty of this paper is to investigate the effect of catastrophic breakdown in a two-way communication environment. The goal is to determine the steady-state probabilities and the system characteristics. The system balance equations are formulated for different cases, but the analytic solution is very difficult. A software tool is used instead. Figures illustrate the effect of the system parameters on the performance measures in scenarios of regular and catastrophic breakdowns.

**Keywords:** two-way communication · catastrophic breakdown · retrial queues

## Introduction

For modeling different types of infocommunication and computer sciences, the retrial queueing systems are a useful and effective tool. Results can be found in various publications [1–6]. Several models assume finite sources. It means, that a finite number of population is in connection with the system. In some

---

The research work was supported by the Austro-Hungarian Cooperation Grant No 106öu4, 2020.

cases, these finite source models are more realistic and give a better description of the considered application [2, 7]. In addition, the considered real-life systems are unfortunately non-reliable, that is the server can lose its efficiency or may break down. There is a large literature on this types of system subject to random breakdowns [1–6].

Sometimes the customers can not spend long time in the queue or in the orbit. They leave the system. In order not to lose these customers, a two-way communication system is built up. Here the customers can sign up for a special service. The system in an idle period will call these customers for the service. This field also has a literature. For the most interesting results see, [8–15].

This two-way communication principle can be considered as a special search for customers, as well. With this outgoing call, the organization (business, commercial etc.) can look for customers, send advertisement, and call them a personal encounter in case of interest. This way the efficiency of the system also can be increased, the ratio of the idle periods can be optimized. Results can be found, e.g. in [5, 16–20].

In this paper, a special two-way communication system is considered. It is called searching for customers. Two types of customers are in the system. The organization has a small or large number of basic customers. They are called the primary customers. They make calls (request) towards the business entity. It can be considered as a server in the system. The principle of the retrial queueing system is applied for these customers. In case of an occupied server, the customer is not lost, it can wait and retry its request for the service. The customer keep retrying until a successful service. The first (finite) source contains these customers. During the idle periods of the server, outgoing calls are performed towards the secondary clients. They are in the second (infinite) source. These customers are the secondary customers, and they are reaching the system for some special reasons. For example, answer some promotion or check some personal data. They are called at an idle period of the system, but in the time period until arriving, a primary customer might be arrived. In this case the secondary customer finds the system occupied. A called customer can not be sent away, so it is placed in a priority (non-preemptive) buffer. For the next outgoing call this buffer will be applied.

The non-reliable server might break down. The main interest of this paper is to investigate the effect of a disaster event, which is called as a catastrophic event or breakdown. The most characteristic property of this type of breakdown is, that in case of the disaster event (or negative customer arrival) all of the services are interrupted, the customers from the servers, orbit, priority buffer are sent back to the corresponding source. There is a large literature on this phenomenon. For example, these types of investigations are very effective for describing the behavior of bank teller equipment. Here, several different disaster event can be imagined. For example, some mechanical malfunction, loss of power etc. Sometimes, the disaster is represented as the presence or arrival of a so called negative client. In this case, all of the service is stopped, the system is blocked,

and all of the clients are removed from the system. Results on disaster and negative customers can be seen in [21–25], and reference therein.

The chapters of the paper are organized in the following way. In Sect. 1 the element of the system is given. It contains the stochastic description of the system (Markov analysis), the working modes, scenarios, parameters etc. Section 2 contains the system balance equations. The system probabilities can be calculated from those equation. For calculation the Mosel-2 software is used. From system probabilities the performance measures (mean waiting times, size of orbit etc.) can be obtained. The paper ends with a summary and conclusion.

## 1 Description of the Model

The system is modeled by a single-server retrial queueing system with a finite and an infinite sources. The functionality of the model is displayed in Fig. 1.

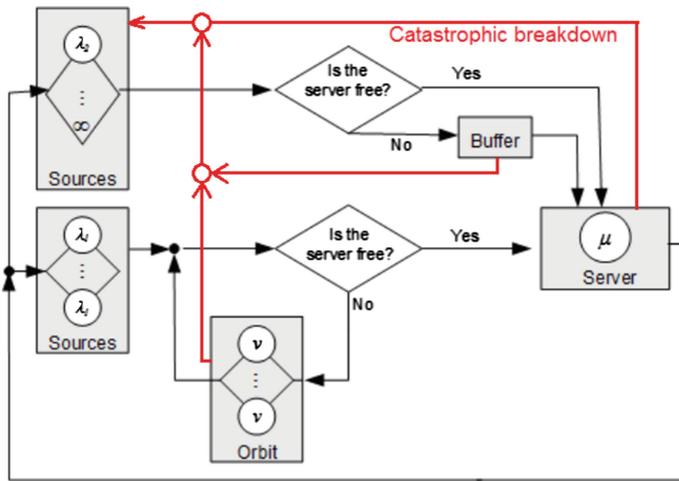


Fig. 1. The system model

The system has two sources. The first one is finite with the first order customers, the number of customers is  $N$ . These customers generate a job towards the server using the exponential law with parameter  $\lambda_1$ . For the first order customers, there is no queue at the server. After the service, the job goes back to the source can generate a new request again. The service time is also exponentially distributed with parameter  $\mu_1$ . When the server is busy, the incoming job is transferred to the orbit. The size of the orbit is  $N$ . From the orbit the jobs after an exponential random time interval with parameter  $\nu$  retry their request to the server until they are served.

The model has second order customers in an infinite number of sources. These customers generate triggered requests only. The idle server makes outgoing calls

towards this infinite source, and the second order customers generate a request to be served. The generation is also exponential with parameter  $\lambda_2$ . The distribution of service times is also exponential with parameter  $\mu_2$ . During the generation time interval of the second order customer, a new customer might arrive to the server. Thus, the called and incoming second order customer may find the server busy. In this case, the following scenarios can be investigated:

- The second order customer is sent back to the infinite source,
- The second order customer is transferred into a priority buffer. In the case of an idle server, a second order customer is called from this buffer. The size of the buffer is one because in the case of an idle server, there is no outgoing call when a customer is in the buffer.

In this model the single server is unreliable, it may be subject to breakdown. Here the regular and the catastrophic breakdowns are considered. In regular breakdown, the server stops working. The breakdown parameters are  $\gamma_0$  and  $\gamma_1$  for idle and busy servers, respectively.  $\gamma_2$  is the parameter of the repair. The behavior of the customers at the time of breakdown is described below. The considered times are exponentially distributed. During the breakdown period, the sources can be blocked (they are not able to generate requests) or non-blocked. In this paper, the non-blocked case is considered. The other breakdown mode is the catastrophic breakdown. This is the situation when a disaster event removes all of the customers from the system (from the orbit, from the buffer, and from the server after interrupting the service). The repair of the system starts immediately. The breakdown parameters are  $\gamma_0$  and  $\gamma_1$  for idle and busy servers, respectively.  $\gamma_2$  is the parameter of the repair. The considered times are exponentially distributed. During the breakdown period, the sources are blocked, they are not able to generate requests.

In case of a regular breakdown, a primary client can find the server down. In this case, it is sent to the orbit. In this server state, a secondary client might reach the system, as well. For this, the unoccupied server performs an outgoing call. The infinite source has an inter arrival time (request generation). This time has an exponential distribution. The parameter of the distribution is  $\lambda_2$ . During the time interval of the generation, the time between the call and the arrival of the secondary client, the server can go wrong. Two different scenarios can be considered here.

- The secondary client is transferred back to the infinite source,
- The secondary client is transferred into a priority buffer. In the case of an idle server, a second order customer is called from this buffer.

In case of a catastrophic breakdown, the sources are blocked. No new request is generated.

When the server is busy, a first order or a second order customer is under service. A breakdown can occur in the busy state, as well. In case of regular breakdown, the behavior of the first order customers can be the followings:

- The primary client is transferred to the waiting facility (orbit),

- The primary client leaves the system and goes back to the source,
- The primary client remains at the system, namely, at the server. The service of the client will be restarting or resuming when the server is up again. (because of the memory-less property of the exponential distribution of the service times, restarting or continuing the service has no difference).

In case of catastrophic breakdown, all of the customers in the system, thus the customer under service are sent back to the sources.

In case of regular breakdown, the behavior of the second order customers there are also some cases to be investigated.

- The secondary client remains at the server. The service will start again when the repair is finished.
- The secondary client leaves the system and goes back to the second source.

In case of catastrophic breakdown, all of the customers in the system, thus the customer under service are sent back to the sources.

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  $t$ :

$$S(t) = \begin{cases} 0, & \text{when no job is at the server} \\ 1, & \text{when the server is working} \\ & \text{with a primary client} \\ 2, & \text{when the server is working} \\ & \text{with a secondary client} \\ 3, & \text{when the server is down} \end{cases}$$

The state-space of the underlying 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.

For buffered and non-buffered models the system balance equations can be formulated. For example, in the non-buffered case when a customer under service is sent back to the corresponding source at a breakdown the equations are the following.

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

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

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

$$[(N - j)\lambda_1 + \mu + \gamma_1] p_{2,j} = \lambda_2 p_{0,j} \tag{4}$$

$$[(N - j)\lambda_1] p_{3,j} = \gamma_0 p_{0,j} + \gamma_1 p_{1,j-1} + \gamma_1 p_{2,j} \quad (5)$$

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

The system balance equations in the case of a catastrophic breakdown look like this:

$$[(N - j)\lambda_1 + \lambda_2 + j\nu + \gamma_0] p_{0,j} = \mu p_{1,j} + \mu p_{2,j} \quad (6)$$

$$[(N - j - 1)\lambda_1 + \mu + \gamma_1] p_{1,j} = (N - j)\lambda_1 p_{0,j} + (j + 1)\nu p_{0,j+1} \quad (7)$$

$$[(N - j)\lambda_1 + \mu + \gamma_1] p_{2,j} = \lambda_2 p_{0,j} \quad (8)$$

$$\gamma_2 p_{3,0} = \gamma_0 p_{0,j} + \gamma_1 p_{1,j} + \gamma_1 p_{2,j} \quad (9)$$

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

The manual solution of the Kolmogorov-equations is very hard. An alternative method has to be found. For calculating the steady-state probabilities, the MOSEL-2 software has been used. Based on the calculated system probabilities the usual performance measures are provided by the software. Using the system probabilities, these measures can be calculated by the following formulas, as well.

– *Utilization 1*

$$U_1 = \sum_{o=0}^N P(1, o) \quad (10)$$

– *Utilization 2*

$$U_2 = \sum_{o=0}^N P(2, o) \quad (11)$$

– *Mean number of customers in the orbit*

$$\bar{O} = \sum_{s=0}^3 \sum_{o=0}^N oP(s, o) \quad (12)$$

– *Mean number of active primary customers*

$$\bar{M} = N - \bar{O} - U_1 \quad (13)$$

– *Mean generation rate of primary customers*

$$\bar{\lambda}_1 = \lambda_1 \bar{M} \quad (14)$$

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

$$\bar{W} = \frac{\bar{O}}{\bar{\lambda}_1} \quad (15)$$

## 2 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 response time are focused.

The manual solution of the Kolmogorov-equations is very hard. There exist alternative methods for performing this task. For calculating the steady-state probabilities, several software tools can be applied. Based on the calculated system probabilities the usual performance measures are usually provided by the software. Because solving directly the balance equations is rather difficult, the MOSEL-2 tool is used. The system equations are solved by the SPNP (Stochastic Petri Net Program). The following figures illustrate the most interesting numerical results. The numerical values of the applied parameters in the model are listed in Table 1.

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

Case studies							
No.	$N$	$\lambda_1$	$\lambda_2$	$\mu$	$\nu$	$\gamma_0$	$\gamma_2$
Fig. 2	100	<i>x - axes</i>	2	3	0.05	0.1	1
Fig. 3	100	<i>x - axes</i>	2	3	0.05	0.01	1
Fig. 4	100	<i>x - axes</i>	2	3	0.05	0.2, 0.5	1
Fig. 5	100	<i>x - axes</i>	2	3	0.05	0.2, 0.5	1
Fig. 6	100	<i>x - axes</i>	2	3	0.05	0.2	1
Fig. 7	100	<i>x - axes</i>	2	3	0.05	0.2, 0.5	1
Fig. 8	100	<i>x - axes</i>	2	<i>special</i>	0.05	0.2	1

The table contains only the idle time breakdown parameters. For calculations, the same values are used for the busy time breakdown parameters. The first two figures compare two different cases for the regular breakdown:

- In case of busy state breakdown, the service of the first order and second order customers are interrupted. The first or second order customer under the interrupted service is sent back to the orbit or to the infinite source, respectively.
- The service of both types of customers is interrupted. The customers are left at the server. When the server is up again, their service will continue or restart. Because of the exponentially distributed service time, this difference - restart or continue - has no effect on the system characteristics.

On Fig. 2 displays the difference of the scenarios mentioned above (leave or remain). The failure rate here is rather high, thus the difference between the scenarios is significant. The interruption is more frequent and the first order

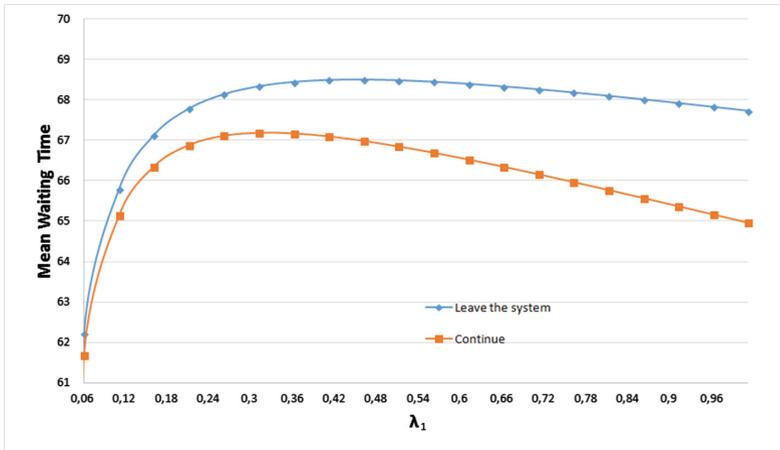


Fig. 2. Mean Waiting Time vs.  $\lambda_1$

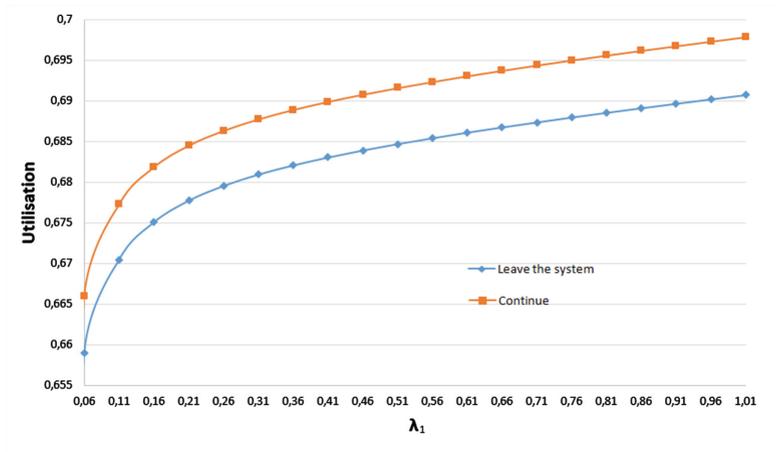


Fig. 3. Utilization vs.  $\lambda_1$

customers are sent back to the orbit more frequently, which results higher waiting times. The waiting time of the ‘leave the system’ case is greater because the first order jobs go to the orbit and they have to try again.



Fig. 4. Mean Waiting Time vs.  $\lambda_1$

Figure 3 displays the utilization in function of the first order generation rate. When the failure rate is small, the difference between the two scenarios is not significant. This figure shows the situation when the failure rate is high, thus the differences in utilization are more significant. The utilization is greater for the ‘Continue’ scenario because after the repair the server state will be immediately busy. While for the other scenario the server will be idle, and a retrial, first or second order generation with an exponentially distributed time interval will take place.

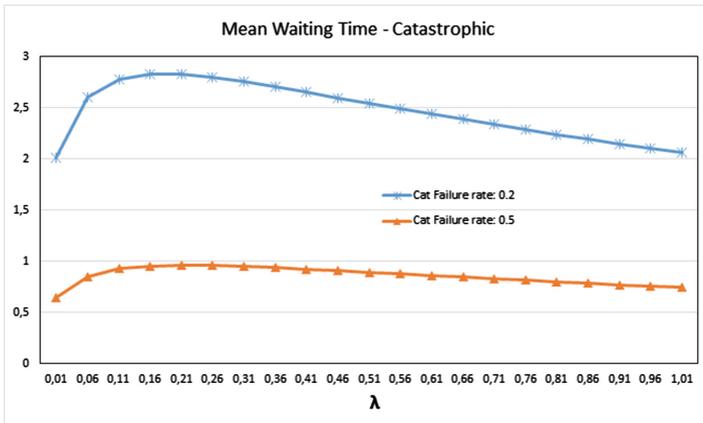


Fig. 5. Mean Waiting Time vs.  $\lambda_1$

In Figs. 4 and 5 the running parameter is the generation rate for first order customers,  $\lambda_1$ , and the considered performance measure is the mean waiting time

in the orbit. Here only the first order customers can take place. In Fig. 4 the case of regular breakdown is displayed for two different failure rates. Comparing the mean waiting times, a reverse effect can be observed. In case of regular breakdown (Fig. 4) for a higher failure rate, the waiting time is higher. The reason is that the customers spend more time in the orbit or at the server in down periods. While in case of catastrophic breakdown (Fig. 5) the waiting time is less for higher breakdown rates. Customers are sent to the sources in case of a breakdown.

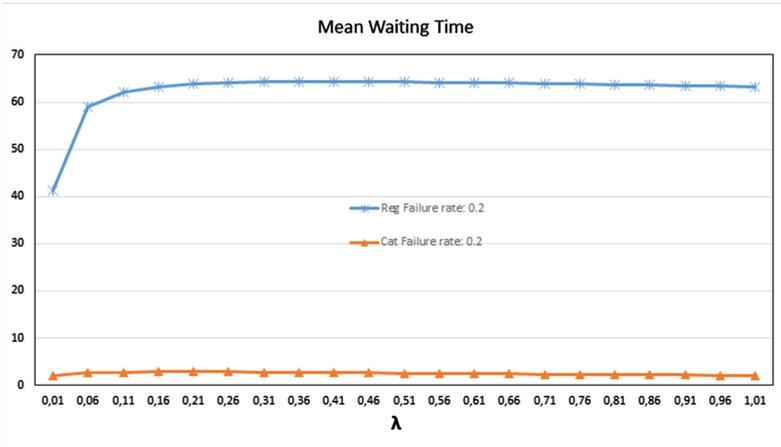


Fig. 6. Mean Waiting Time vs.  $\lambda_1$

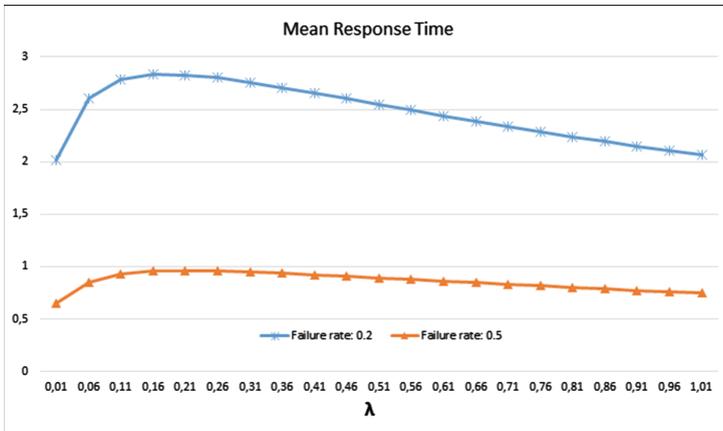


Fig. 7. Mean Response Time vs.  $\lambda_1$

Figure 6 compares the waiting time between the regular and catastrophic breakdown at a given failure rate,  $\gamma_0 = \gamma_1 = 0.2$ . As it was expected, the waiting time is much higher for the regular breakdown.

In Fig. 7 and 8 the catastrophic breakdown is applied. The running parameter is  $\lambda = \lambda_1$ , the first order generation rate. In Fig. 7 the mean response time can be seen for two different failure rates ( $\gamma_0 = \gamma_1$  in this figure). For a higher failure rate, lower response time can be observed because the jobs are more often kicked off to the source.

Figure 8 displays the server utilization. Here, the service rates for first ( $\mu_1$ ) and second order ( $\mu_2$ ) customers are different.  $\mu_1 = 4$  and  $\mu_2 = 2$ . This is the reason, that the utilization is higher in the catastrophic case than in the normal breakdown case.

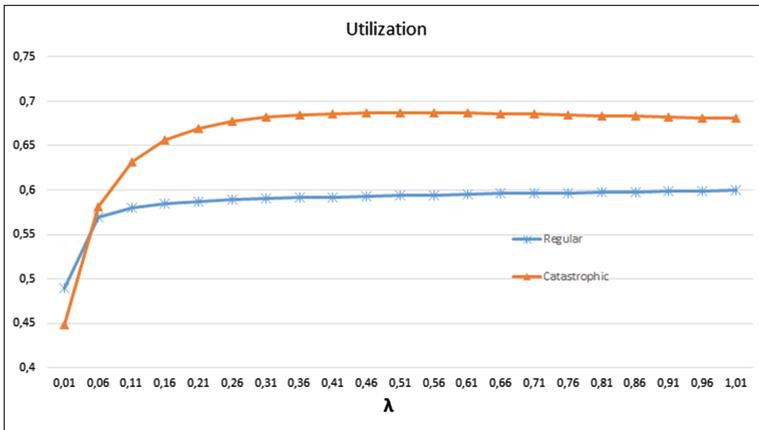


Fig. 8. Server utilization vs.  $\lambda_1$

### 3 Conclusion

A special two-way communication system was investigated here. First order customers come from a finite source, while in the case of an idle server, second order customers can reach the system via a direct call. Different cases can be considered. Failure rates are set to be equal for idle server, for server with first order customer, and for server with second order customer. Two different cases were considered. The system is subject to regular breakdown and catastrophic breakdown. In case of regular breakdown, the “leave the system” and the “remain at server” scenarios were compared. Based on the numeric result, it can be stated, the second scenario is more effective regarding the response times, waiting times, and utilization. When the two breakdown models were compared with different failure rates, the expected reverse effect of the breakdown parameters can be observed.

## References

1. Falin, G.I.: Waiting time in a single-channel queuing system with repeated calls. *Mosc. Univ. Comput. Math. Cybern.* **4**, 66–69 (1977)
2. Artalejo, J.R.: Retrial queues with a finite number of sources. *J. Korean Math. Soc.* **35**, 503–525 (1998)
3. Falin, G.I., Templeton, J.G.C.: *Retrial Queues*. Chapman and Hall, London (1997)
4. Templeton, J.G.C.: Retrial queues. *Top* **7**(2), 351–353 (1999)
5. Gans, N., Koole, G., Mandelbaum, A.: Telephone call centers: tutorial, review, and research prospects. *Manuf. Serv. Oper. Manage.* **5**(2), 79–141 (2003)
6. Artalejo, J.R., Gomez Corral, A.: *Retrial Queueing Systems: A Computational Approach*. Springer, Heidelberg (2008). <https://doi.org/10.1007/978-3-540-78725-9>
7. Falin, G.I., Artalejo, J.R.: A finite source retrial queue. *Eur. J. Oper. Res.* **108**, 409–424 (1998)
8. Falin, G.I.: Model of coupled switching in presence of recurrent calls. *Eng. Cybern.* **17**(1), 53–59 (1979)
9. Dimitriou, I.: A retrial queue to model a two-relay cooperative wireless system with simultaneous packet reception. In: Wittevrongel, S., Phung-Duc, T. (eds.) *ASMTA 2016*. LNCS, vol. 9845, pp. 123–139. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-43904-4\\_9](https://doi.org/10.1007/978-3-319-43904-4_9)
10. Dragieva, V., Phung-Duc, T.: 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*, p. 11. ACM (2016)
11. 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](https://doi.org/10.1007/978-3-319-68069-9_3)
12. 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* (2017)
13. 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](https://doi.org/10.1007/978-3-642-30782-9_2)
14. Sakurai, H., Phung-Duc, T.: Two-way communication retrial queues with multiple types of outgoing calls. *TOP* **23**(2), 466–492 (2014). <https://doi.org/10.1007/s11750-014-0349-5>
15. Sakurai, H., Phung-Duc, T.: Scaling limits for single server retrial queues with two-way communication. *Ann. Oper. Res.* **247**(1), 229–256 (2015). <https://doi.org/10.1007/s10479-015-1874-9>
16. Aguir, S., Karaesmen, F., Akşin, O.Z., Chauvet, F.: The impact of retrials on call center performance. *OR Spectr.* **26**(3), 353–376 (2004)
17. Aksin, Z., Armony, M., Mehrotra, V.: The modern call center: a multi-disciplinary perspective on operations management research. *Prod. Oper. Manag.* **16**(6), 665–688 (2007)
18. Brown, L., et al.: Statistical analysis of a telephone call center: a queueing-science perspective. *J. Am. Stat. Assoc.* **100**(469), 36–50 (2005)
19. Pustova, S.V.: Investigation of call centers as retrial queueing systems. *Cybern. Syst. Anal.* **46**(3), 494–499 (2010)

20. Wolf, T.: System and method for improving call center communications. US Patent App. 15/604,068, 30 November 2017
21. Subramanian, S., et al.: A stochastic model for automated teller machines subject to catastrophic failures and repairs. *Queueing Models Serv. Manage.* **1**(1), 75–94 (2018)
22. Thilaka, B., Poorani, B., Udayabaskaran, S.: Performance analysis for queueing systems with close down periods subject to catastrophe. *Int. J. Pure Appl. Math.* **119**(7), 39–57 (2018)
23. Gupta, U.C., Kumar, N., Barbhuiya, F.P.: A queueing system with batch renewal input and negative arrivals. In: Joshua, V.C., Varadhan, S.R.S., Vishnevsky, V.M. (eds.) *Applied Probability and Stochastic Processes*. ISFS, pp. 143–157. Springer, Singapore (2020). [https://doi.org/10.1007/978-981-15-5951-8\\_10](https://doi.org/10.1007/978-981-15-5951-8_10)
24. Piriadarshani, D., Narasimhan, S., Maheswari, M., James, B., et al.: A retrial queueing system operating in a random environment subject to catastrophes. *Eur. J. Mol. Clin. Med.* **7**(2), 5029–5032 (2020)
25. Ammar, S.I., Zeifman, A., Satin, Y., Kiseleva, K., Korolev, V.: On limiting characteristics for a non-stationary two-processor heterogeneous system with catastrophes, server failures and repairs. *J. Ind. Manage. Optim.* **17**(3), 1057 (2021)