Investigation of a Finite-Source Retrial Queueing System with Two-Way Communication, Catastrophic Breakdown and Impatient Customers Using Simulation

János Sztrik, Ádám Tóth(✉), Ákos Pintér, and Zoltán Bács

University of Debrecen, University Square 1, Debrecen 4032, Hungary
{sztrik.janos,toth.adam}@inf.unideb.hu, apinter@science.unideb.hu, bacs.zoltan@econ.unideb.hu

Abstract. An M/M/1//N retrial queueing system with two-way communication to the infinite source and impatient customers in the orbit is considered in the paper. There is a finite source in which the primary or regular customers are coming, while requests from the infinite source are the secondary customers. Because of not having waiting queues, the service of an arriving, primary customer begins immediately. Otherwise, in the case of a busy server, the primary customers are forwarded to the orbit waiting an exponentially distributed random time to try to reach the service unit. When the service unit is in an idle state, it may call a customer from the infinite source for service. All requests possess an impatience property resulting in an earlier departure from the system through the orbit if they wait too much for being served. Besides, the service unit is supposed to break down according to several distributions which have a specialty in removing all the customers located in the system. In the case of a faulty state, blocking is applied not allowing the customers into the system until the service unit fully recovers. This work concentrates on examining the effect of those distributions on several performance measures like the distribution of the number of customers in the system, and the probability of a primary customer departing because of catastrophe. The obtained results are graphically realized to show the differences and curiosities among the used parameter settings of the various distributions.

Keywords: Simulation · Catastrophic breakdown · Retrial queuing system · Collision · Impatience · Sensitivity analysis

1 Introduction

In many fields of our life, the phenomenon of waiting appears making inevitably creating queueing systems handling for example increasing network traffic

V. M. Vishnevskiy et al. (Eds.): DCCN 2022, LNCS 13766, pp. 231–242, 2022.
https://doi.org/10.1007/978-3-031-23207-7_18

in many info-communication systems. Throughout the years researchers have designed numerous tools and mechanisms which are suitable for modeling various organizations and one prime instance is retrial queueing systems that are capable of depicting arising real-life problems in telecommunication schemes like telephone switching systems, call centers, computer networks, and computer systems. Several publications exist where researchers exploit the advantages of retrial-queuing systems with repeated calls using for their models like in [2,7,8,12].

In the case of a retrial queueing system, a virtual waiting room is taken into consideration which is called the orbit meaning that whenever a service of a job can not start because of failure or occupation of the server it remains in the system. In the orbit, these customers have the opportunity to be at the service facility after a random time. The population of the customers is finite as the probability that a server calls a customer from the orbit is not very small and under such circumstances, it is more suitable to examine models with retrial queues. Exploring the available literature many papers have applied infinite and finite source queueing systems for example in [1,15,17,19].

It is also interesting to observe how the feature of two-way communication is used in papers in many fields of life. Its popularity originates from its usefulness to model applicable systems and creating real-life applications. One prime example can be mentioned in the topic of telecommunication, especially in call centres where agents may be occupied with other particular labor during an inactivity period like selling, advertising, and promoting products besides handling the calls of the customers. Optimizing the utilization of the service units or agents is always pivotal to increasing the efficiency of such systems. To mention some works about applying two-way communication schemes here are some instances [5,18].

Waiting is a natural occurrence in many aspects of our life and people experience the annoyance of having to wait in queues which is not a satisfying act. This may result in earlier departures of the requests leaving the system without being served which is called impatience. This behaviour is experienced in many territories like in healthcare applications, call centers, or telecommunication networks. This phenomenon is extensively studied in numerous articles and different types of impatience are distinguished: balking customers choose to avoid entering the system if the size of the queue is large, and jockeying customers can change their positions among the queues to get the service earlier or reneging customers decide to leave the queues if they have waited a specific long time there. About the investigation of this behaviour here are some examples: [10,14,16].

Random breakdowns and malfunctions occur in real-life scenarios caused by a power outage, human negligence, or other catastrophic activity. This greatly changes the operation of the system and the performance measures thus its investigation is necessary. In many cases, the service units are presumed to be accessible all the time which is not realistic. Systems with random failures have been investigated by many researchers for example in [6,13,20]. However, there are certain situations where the effect of breakdown is different types of failures can be investigated. For instance, power outages or mechanical failures may cause catastrophic events in which all the customers in the system are removed. This is known as a negative customer and it takes out every other request from service upon its arrival. This eventuates a disaster event because it also breaks down the service unit and in this case, every customer is forwarded back to the source. Papers in connection with negative customers can be found in [3,11].

The aim of this work is to realize a sensitivity analysis using various distributions of failure time on several performance measures while the departure of customers may happen. The results are obtained by our stochastic simulation program using the basics of SimPack [9] which contains the basic building blocks of a simulation model. This gives us the opportunity to model any type of queueing system to create any type of simulation model and we can calculate any performance measure using arbitrary random number generators for the desired random variable. The presented curves highlight both the effect of disaster events and the impatience of the customers applying various parameter settings and these figures concentrate on the interesting phenomena of these systems. The table of input parameters and graphical illustrations of the results are included demonstrating the influence of the used distributions on the main performance metrics.

2 System Model

A finite-source retrial queueing system of type $M/M/1//N$ is regarded as an unreliable service unit and impatient customers (see Fig. 1). This model has a service unit and exactly N individual resides in the finite-source in which request generation (primary request) is proceeded towards the system according to exponential law with parameter λ. This means that the inter-arrival times are exponentially distributed with mean λ. If an arriving job finds the server in an idle state then its service starts immediately which is an exponentially distributed variable with μ_1. Otherwise, in vain of a queue, jobs are not lost but remain in the system being forwarded to the orbit which is a virtual waiting space. From there these requests after an exponentially random time with parameter ν retry to reach the service unit. After spending futilely an exponentially distributed time with rate τ in the orbit a customer may choose to leave the system without being served so in other words, every request has an impatience characteristic. It is assumed random breakdowns take place according to various distributions like gamma, hyper-exponential, Pareto, hypo-exponential, and lognormal. The parameters are chosen in a way that a real sensitivity analysis would be accomplished. In these occurrences, disaster events develop resulting in interruption of the service of a job and with the customers, in the orbit, they all depart from the system. Blocking is applied during faulty periods so no customers are allowed by the system until the server fully recovers.

The repair process begins to be executed after the failure of the service unit which happens according to an exponential distribution with parameter γ_2. Two-way communication was also introduced in our model, when the server becomes free it may call a request from an infinite source (secondary customer) after an exponentially random time with parameter λ_2. That type of customer occupies instantly the service unit if it is not busy upon its arrival, otherwise, it is forwarded to a special buffer where it waits there until the server turns idle. At that moment a secondary customer automatically enters the service facility. In the case of a catastrophic event, every primary job returns to the finite-source, and every secondary customer exits from the system including the one who is under service. The service time of the secondary customer also follows an exponential distribution with the rate of μ_2. Every appeared arbitrary variable in the model construction is supposed to be independent of each other.

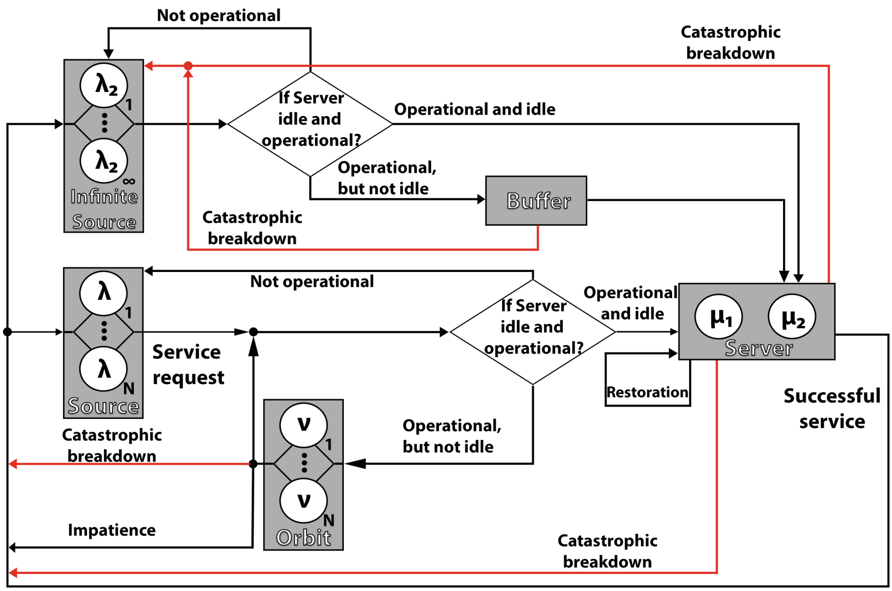

Fig. 1. System model.

3 Simulation Results

3.1 First Scenario

As mentioned earlier SimPack was used as a base of our program and we used a statistical package that was responsible for obtaining the desired performance measures. This utilizes the batch means method where the running period is split into batches (altogether S) and in every batch, $s = R - M/S$ are executed, in which M denotes the warm-up period observations at the beginning of the simulation which are rejected and R the length of the simulation. After the initial period, the sample average of the whole run is calculated, and what is really important here is to have long enough batches and approximately independent sample averages of the batches. Here is an article containing more information about the used process [4].

The confidence level of 99.9% is employed throughout the simulations and 0.00001 is the amount of the relative half-width of the confidence interval to pause the actual simulation sequence. The size of the batch in the initial transient period can not be too small therefore its value is set to 1000.

In Table 1 the used values of input parameters are presented.

Table 1. Numerical values of model parameters

N	ν	μ_1	μ_2	τ	γ_2	λ_2
100	0.01	1	2.5	0.05	1	0.5

The next table (Table 2) consists of the parameters of failure time, every chosen parameter is according to have the same mean and variance value in that way a valid comparison is achieved. The simulation program was tested by many parameter values and the reason for selecting these values to focus on interesting situations besides that it is worth mentioning that almost the same phenomenon appeared in this particular setting. The squared coefficient of variation is more than one in this scenario which is totally intentional to check the influence of peculiar random variables.

Table 2. Parameters of failure time

Distribution	Gamma	Hyper-exponential	Pareto	Lognormal
Parameters	$\alpha = 0.31225$ $\beta = 0.05588$	$p = 0.36197$ $\lambda_1 = 0.12955$ $\lambda_2 = 0.22835$	$\alpha = 2.1455$ $k = 2.9835$	$m = 1.00278$ $\sigma = 1.19819$
Mean	5.558			
Variance	100			
Squared coefficient of variation	3.2024857438			

In Fig. 2 i represents the number of primary customers in the system on the X-axes, and $P(i)$ denotes the probability that exactly i primary customers are situated at the server and in the orbit altogether on the Y-axes. The distribution of the number of primary customers in the system is displayed when λ is 0.11 using various distributions of failure time. The mean number of primary customers in the system differs from each other greatly. In the case of the gamma distribution, customers tend to spend more time in the system compared to Pareto distribution. It is also noticeable that the highest probability is 0 and this can be explained by the fact that during faulty periods customers are not allowed to enter and for every catastrophic breakdown the system is emptied.

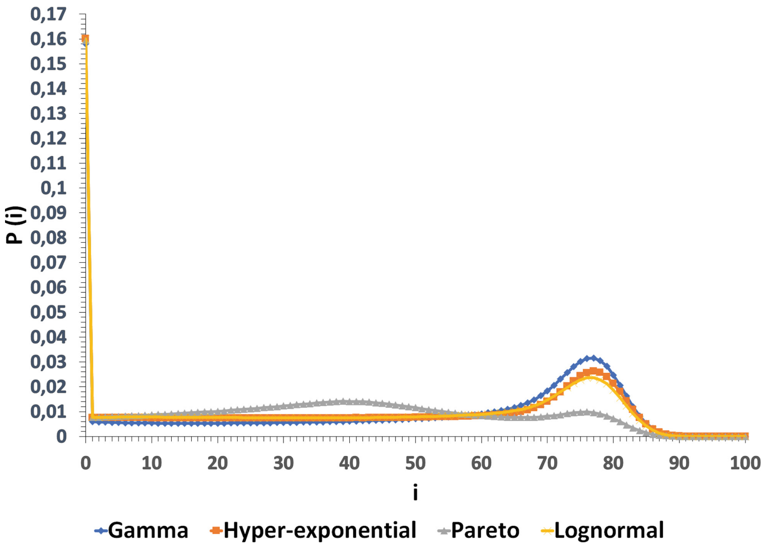

Fig. 2. Distribution of the number of primary customers in the system.

The expected response time of an arbitrary primary customer is presented in the function of the arrival intensity of incoming primary request in Figs. 3 and 4. Even though the mean and the variance value are equal to each other, huge gaps develop among the applied distributions, at gamma the highest average response times can be observed compared to the others (Fig. 3). Also with the increment of the arrival intensity, the expected response time of an arbitrary primary customer starts to increase then after a certain intensity arrival ($\lambda = 0.07$) it decreases except in the case of Pareto. The effect of impatience of the primary customer on the average response time is visible in Fig. 4 using the gamma distribution. Naturally, with the increment of rate τ the mean sojourn time decreases and this tendency is exactly the same for the other distributions. Although the mean operation time is 5.558 the values of the expected response times are higher than that which is quite interesting. Our intuition is that this can be explained by that the variance is quite high resulting in many small operation times and in most of them no customer can enter because they are so small. But there are several high operation periods in which it is very probable that many jobs enter and spend relatively a high amount of time.

Figure 5 demonstrates the development of the probability that a primary customer departs because of impatience besides increasing arrival intensity. At $\tau = 0.0001$ the probability of departure caused by impatience is basically 0 and as the rate of impatience increases this measure increases as well. At $\tau = 0.1$ almost half of the incoming primary customers leave the system earlier willingly.

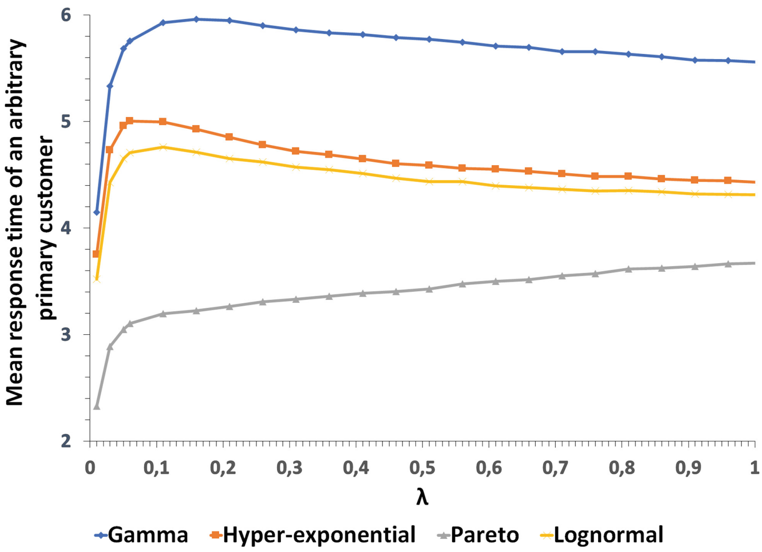

Fig. 3. Average response time of an arbitrary primary customer.

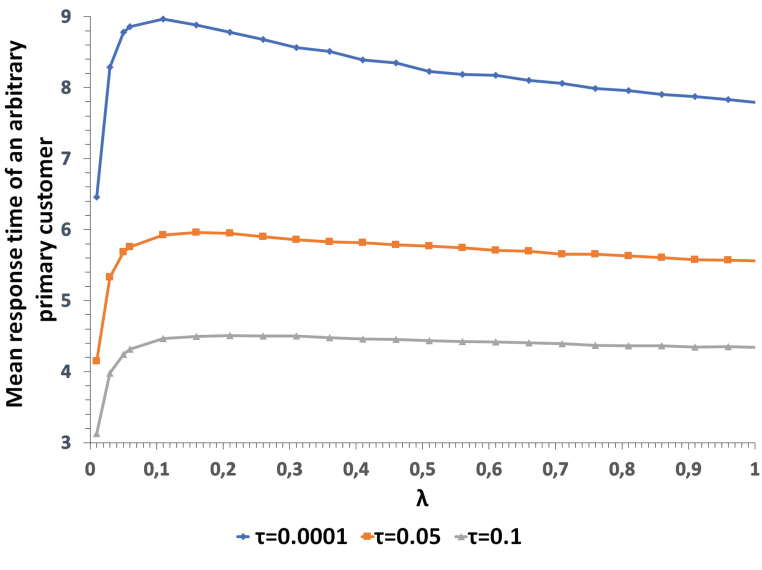

Fig. 4. Average response time of an arbitrary primary customer using different values of impatience.

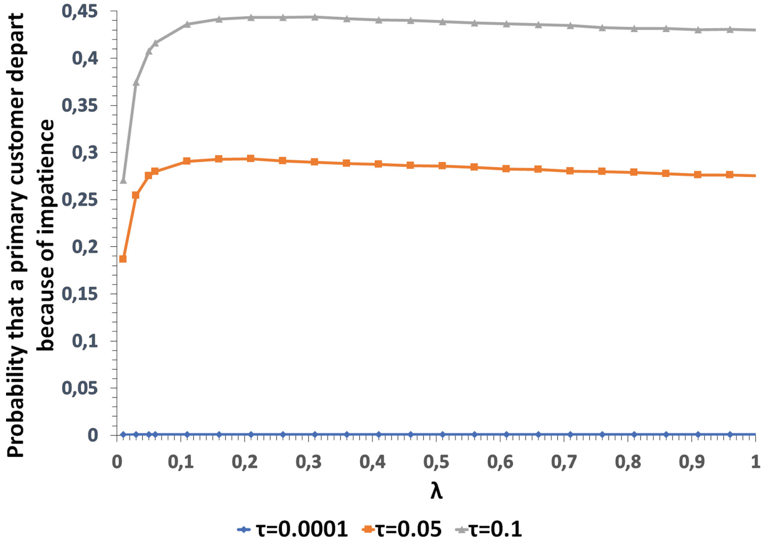

Fig. 5. Probability of departure of a primary customer due to impatience.

3.2 Second Scenario

Examining the results and phenomena of the first section we were excited to see how another parameter setting modifies the behaviour of the performance measures. In this second scenario, the parameters were chosen to have the squared coefficient of variation being less than one. Hypo-exponential replaces the hyper-exponential distribution in which the squared coefficient of variation is always more than one. We will analyze the same measures as in the previous section but the new parameters of failure time which are presented in Table 3. The mean is the same but the variance is different in this case less than compared in the last section. All the remaining parameters are unchanged see Table 1.

Table 3. Parameters of failure time

Distribution	Gamma	Hypo-exponential	Pareto	Lognormal
Parameters	$\alpha = 1.2321$	$\mu_1 = 0.2$	$\alpha = 2.494$	$m = 1.4235$
	$\beta = 0.2205$	$\mu_2 = 1.7$	$k = 3.3478$	$\sigma = 0.7709$
Mean	5.588			
Variance	25.3460207612			
Squared coefficient of variation	0.811634349			

Figure 6 is in connection to the steady-state distribution of primary customers in the system. Analyzing the curves the obtained values are very near to each other, and the mean numbers are much closer than in the first section. However, there is a similarity as well, the most probable state is 0 and it is around 0.16 too. Likewise, these results were obtained besides the arrival intensity of 0.11.

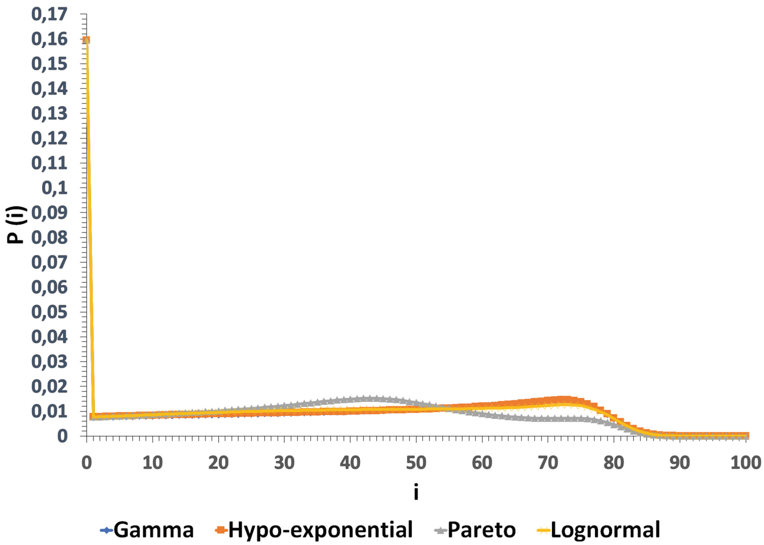

Fig. 6. Distribution of the number of customers in the system.

The next two figures (Fig. 7 and 8) present the average response time of an arbitrary primary customer versus the arrival intensity. On the first one (Fig. 7) the comparison of the used distribution can be seen where the difference is quite small among the values but the tendency remains the same, the highest amount is observed at gamma and the lowest at Pareto distribution. But none of them reach the mean value of operation time and based on our assumption this is because the variance is lower in this parameter setting. Figure 8 demonstrates the influence of impatience when the failure time follows hypo-exponential distribution however the same phenomenon is true for the others. Commonly, when rate τ begins to rise it makes the average response time lessens.

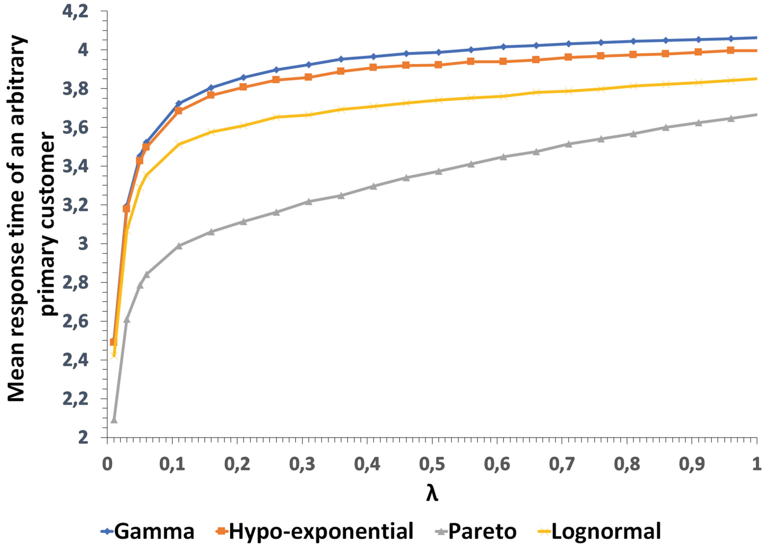

Fig. 7. Average response time of an arbitrary primary customer.

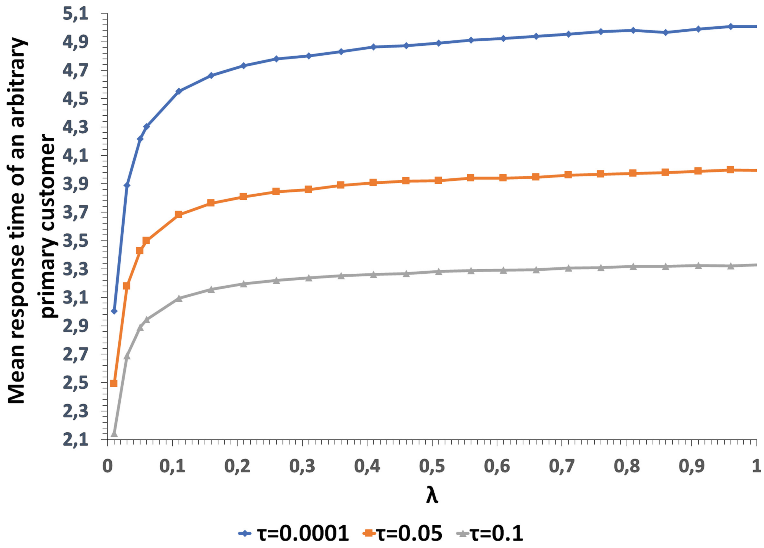

Fig. 8. Average response time of an arbitrary primary customer using different values of impatience.

4 Conclusion

A finite-source retrial queueing system is introduced with a non-reliable server, a two-way communication scheme, and impatient customers. We investigated several scenarios where different parameters are used to carry out a sensitivity analysis to figure out how the different performance measures develop. The results are obtained by our simulation program and several graphical figures depict the effect of using various distributions of failure time on the expected response time of primary and primary departed customers, on the probability of departure because of impatience, or on the distribution of customers in the system. In our figures, the differences were clearly seen among the values of several performance measures when the squared coefficient of variation is greater than one showing how pivotal applying a distribution can be and substantially slightly when it is less than one. The curves also reveal the impact of impatience on reducing the average response time of a primary customer. In the future we plan to continue our research work, examining other types of finite-source retrial queuing systems with two-way communication or adding another service unit for backup purposes.

References

1. Ali, A.A., Wei, S.: Modeling of coupled collision and congestion in finite source wireless access systems. In: 2015 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1113–1118. IEEE (2015)
2. Artalejo, J., Corral, A.G.: Retrial Queueing Systems: A Computational Approach. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78725-9
3. Chakravarthy, S.R., Subramanian, S.: A stochastic model for automated teller machines subject to catastrophic failures and repairs. SIMULATION **1**(1), 75–94 (2018)
4. Chen, E.J., Kelton, W.D.: A procedure for generating batch-means confidence intervals for simulation: checking independence and normality. SIMULATION **83**(10), 683–694 (2007)
5. Dragieva, V., Phung-Duc, T.: Two-way communication $M/M/1//N$ retrial queue. In: Thomas, N., Forshaw, M. (eds.) ASMTA 2017. LNCS, vol. 10378, pp. 81–94. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61428-1_6
6. Dragieva, V.I.: Number of retrials in a finite source retrial queue with unreliable server. Asia Pac. J. Oper. Res. **31**(2), 1–23 (2014). https://doi.org/10.1142/S0217595914400053
7. Falin, G., Artalejo, J.: A finite source retrial queue. Eur. J. Oper. Res. **108**, 409–424 (1998)
8. Fiems, D., Phung-Duc, T.: Light-traffic analysis of random access systems without collisions. Ann. Oper. Res. **277**(2), 311–327 (2019)
9. Fishwick, P.A.: SimPack: getting started with simulation programming in C and C++. In: 1992 Winter Simulation Conference, pp. 154–162 (1992)
10. Gupta, N.: Article: A view of queue analysis with customer behaviour and priorities. In: IJCA Proceedings on National Workshop-Cum-Conference on Recent Trends in Mathematics and Computing 2011, RTMC(4), May 2012
11. 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

12. Kim, J., Kim, B.: A survey of retrial queueing systems. Ann. Oper. Res. **247**(1), 3–36 (2016)
13. Krishnamoorthy, A., Pramod, P.K., Chakravarthy, S.R.: Queues with interruptions: a survey. TOP **22**(1), 290–320 (2014)
14. Kumar, R., Jain, N., Som, B.: Optimization of an $M/M/1/N$ feedback queue with retention of reneged customers. Oper. Res. Decis. **24**, 45–58 (2014). https://doi.org/10.5277/ord140303
15. Lakaour, L., Aïssani, D., Adel-Aissanou, K., Barkaoui, K.: $M/M/1$ retrial queue with collisions and transmission errors. Methodol. Computi. Appl. Probab. **21**(4), 1395–1406 (2018)
16. Panda, G., Goswami, V., Datta Banik, A., Guha, D.: Equilibrium balking strategies in renewal input queue with Bernoulli-schedule controlled vacation and vacation interruption. J. Ind. Manag. Optim. **12**, 851–878 (2015). https://doi.org/10.3934/jimo.2016.12.851
17. Peng, Y., Liu, Z., Wu, J.: An M/G/1 retrial G-queue with preemptive resume priority and collisions subject to the server breakdowns and delayed repairs. J. Appl. Math. Comput. **44**(1–2), 187–213 (2014). https://doi.org/10.1007/s12190-013-0688-7
18. Sztrik, J., Tóth, Á., Pintér, Á., Bács, Z.: The simulation of finite-source retrial queueing systems with two-way communication and impatient customers. In: Vishnevskiy, V.M., Samouylov, K.E., Kozyrev, D.V. (eds.) DCCN 2021. LNCS, vol. 13144, pp. 117–127. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-92507-9_11
19. Takeda, T., Yoshihiro, T.: A distributed scheduling through queue-length exchange in CSMA-based wireless mesh networks. J. Inf. Process. **25**, 174–181 (2017)
20. Tóth, A., Sztrik, J., Pintér, A., Bács, Z.: Reliability analysis of finite-source retrial queuing system with collisions and impatient customers in the orbit using simulation. In: 2021 International Conference on Information and Digital Technologies (IDT), pp. 230–234 (2021). https://doi.org/10.1109/IDT52577.2021.9497567