

Reliability Analysis of Finite-Source Retrial Queuing System with Collisions and Impatient Customers in the Orbit Using Simulation

Ádám Tóth

University of Debrecen
Debrecen, Hungary
toth.adam@inf.unideb.hu

János Sztrik

University of Debrecen
Debrecen, Hungary
sztrik.janos@inf.unideb.hu

Ákos Pintér

University of Debrecen
Debrecen, Hungary
apinter@science.unideb.hu

Zoltán Bács

University of Debrecen
Debrecen, Hungary
bacs.zoltan@econ.unideb.hu

Abstract—In this paper, we have developed a simulation program to investigate $M/M/1/N$ and $M/G/1/N$ retrial queuing systems with collisions and impatient customers in the orbit. In our model, for lack of waiting queues, the service of an arriving customer begins immediately. Otherwise, it has the ability to bring about a collision in which both the arrived and request under service are forwarded to the orbit where spending some exponentially distributed random time they try to get their service demand to be executed. All requests possess an impatience property resulting in an earlier departure from the system through the orbit if they spend too much time waiting for being served properly. The phenomenon of blocking is applied not allowing the customers into the system while the service unit resides in a faulty condition. The server is supposed to break down according to several distributions and this work concentrates on examining the effect of these distributions on several performance measures like the distribution of the number of collisions and failures of customers. The results are graphically illustrated to experience the difference among the used parameter settings of the various distributions.

Index Terms—retrial queues, finite-source queuing system, server breakdowns and repairs, simulation, impatience.

I. INTRODUCTION

Currently Internet usage increases due to mainly the rapid development of cloud service providers providing highly reliable, scalable, low-cost infrastructure platforms and the great number of devices having the ability to transfer data over a network. Network activity initiated by the users from home and companies escalates in such a way that models of communication systems are needed to fully understand and optimize the operation of network traffic. In real-life scenarios, the usage of retrial queues is an effective tool to cope with issues in telecommunication systems like telephone switching systems, call centers, CSMA-based wireless mesh networks in frame level and computer systems. The following articles express its importance in different fields of informatics [1], [2], [3], [4], [5].

It is not an uncommon feature of retrial queues possessing a virtual waiting area the so-called orbit in the case of modeling such systems. During managing administration duties a customer may wait for its turn to receive its service need when all agents are occupied in a call center. Numerous papers carry out

investigation on models having orbit for example in [3],[4],[5].

It can be observed that in many cases customers may abandon or renege from waiting in a queue or a system that is presented in many situations in our everyday life. For instance, after waiting a specified amount of time customers can hang up waiting for an agent in a call center or customers may decide not to enter a shopping mall if it is so crowded. Every customer has a patience threshold value which depends on many factors, in this model, this relies on the amount of time spent in the system so customers are characterized by renegeing features. Some results in connection with impatience can be viewed in the following articles [6],[7].

In connection to the available channels or facilities, users and sources are fighting for them to fulfill their service needs as soon as possible. Because of the possibility of launching attempts at the same time collisions may take place resulting in the loss of transmission and an unfortunate delay during the provision of services. Steps are needed to be done having an efficient procedure to prevent conflict and corresponding message delay and to ensure retransmission if it is necessary. Many papers have published results including models with collisions in [8],[9],[10],[11].

Unfortunately, service units are subjected to breakdowns which play quite an important role to alter the performance of the system and performance measures. In the literature, some articles neglect failures in the model construction which is quite unrealistic but hardware could fail and networks could have intermittent timeout periods. It is important to have plans for such events experiencing disruptions so the examination of retrial queueing systems with random server breakdowns and repairs is essential, the impact would be crucial for the system if the downtime is too high. Transmission failures, interruptions throughout transferring packets may have severe consequences that can cause the end of a company or big financial losses. Reliability analysis has been carried out by several works like [12],[13],[14],[15],[16],[17],[18],[19].

We examined the operation of type $M/M/1/N$ a retrial queueing system where the service unit is unreliable and customers may depart from the orbit if they spend too much time waiting to be served. The novelty of our work is to

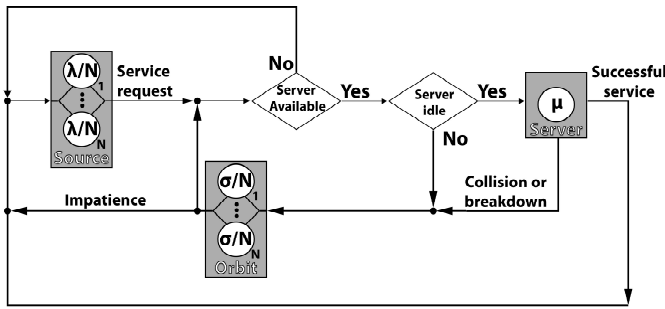


Fig. 1. System model

carry out sensitivity analysis using different distributions of failure and service time on performance measures including the distribution of collisions and the average number of failures of an arbitrary customer. Results were obtained by a self-developed simulation program based on SimPack [20] toolkit that constitutes the base of our model and provides utilities to create a working simulation from a model description. Graphical illustrations are proven to highlight the gathered results to exhibit the effect of the utilized distributions.

II. SYSTEM MODEL

Figure 1 presents the considered model as a type of $M/M/1/N$ finite-source retrial queueing system with an unreliable server, impatient customers in the orbit, and blocking. The finite-source possesses N customers in which each of them generates calls towards the system and this inter-request time follows exponential distribution with parameter λ/N . If the service unit is idle the service of an incoming request begins instantly because there is no waiting queue. This variable is also exponentially distributed with parameter μ . Alternatively, when the server is busy with a call, an arriving customer brings about a collision which results in moving both customers to the orbit, which is a virtual waiting room. Here, the requests spend an exponentially distributed time before launching another attempt to arrive at the service facility. Every customer is characterized by the property of impatience meaning that after a certain amount of exponentially distributed random time with parameter τ it may decide to leave the system earlier without obtaining the appropriate service. The service unit is supposed to break down according to gamma, hypo-exponential, hyper-exponential, Pareto and lognormal distribution selecting the parameters having the same mean value. Promptly, the restoration period initiates upon the breakdown that is also an exponential random variable with parameter γ_2 . In the case of failure occurrence, the service of a customer terminates and it is delivered to the orbit. Throughout this period incoming customers are not capable of entering the system because they are rejected and get back to the source. This is the so-called blocking. Every mentioned variable consisting of this model is assumed to be independent of each other.

III. SIMULATION ENVIRONMENT

We used the simulation approach to acquire all the desired performance measures. As we wanted to depict special measures and some of them are difficult to give exact formulas we chose to develop our own simulation package based on SimPack as it is mentioned earlier. Sometimes the state space of the Markov chain is so huge that makes it impossible to solve the derived steady-state equations which are also valid for the available software packages. These are capable of performing an analytical evaluation of complex systems if the variables in the model construction follow exponential distribution. In the case of simulation, we can utilize other distributions as well which gives us the opportunity to perform a sensitivity analysis. The estimation is carried out by applying a statistical package in which the method of batch means is used. In short, n observations occur in every batch and the useful run contains enough batches to accomplish a valid estimation. The batches should be long enough and approximately independent of each other to increase the accuracy of the achieved results. Among the confidence interval techniques for a steady-state mean of a process, this is the most common technique and the following works [21],[22] include more detailed information about the process. The simulations are performed with a confidence level of 99.9%. The relative half-width of the confidence interval required to stop the simulation run is 0.00001.

IV. SIMULATION RESULTS

A. Different distributions of service time of the customers

In this section, we worked with different distributions of service time of the customers with an unreliable service unit to investigate the effect on some performance measures. Observing various simulation runs with different parameter settings, we selected the scenario with the most interesting results and Table I shows that one. γ_0 and γ_1 are the parameters of failure time when the server is busy and idle, accordingly. In [23],[24] similar systems were treated by an asymptotic method where N tends to infinity that is why we use rates λ/N and σ/N . Table II demonstrates the values of parameters in the case of every distribution including hyper-exponential, gamma, lognormal, and Pareto. These distributions are suitable to perform a valid comparison because we can select parameters in order the squared coefficient of variation would be greater than one meaning that both the mean value and variance are the same. About the fitting process, which is needed to be done, can be found in [25].

TABLE I
NUMERICAL VALUES OF MODEL PARAMETERS

N	γ_0	γ_1	σ/N	τ
100	0.05	1	0.05	0.0005

The steady-state probability of a customer in the orbit is presented in Figure 2 when the arrival intensity equals 0.01, which shows the probability that exactly i customers reside in the orbit in a random time. Having the same mean and variance, the obtained results surprisingly vary significantly

TABLE II
PARAMETERS OF SERVICE TIME OF THE CUSTOMERS

Distribution	Gamma	Hyper-exponential	Pareto	Lognormal
Parameters	$\alpha = 0.0625$ $\beta = 0.0625$	$p = 0.4697$ $\lambda_1 = 0.939$ $\lambda_2 = 1.061$	$\alpha = 2.031$ $k = 0.508$	$m = -1.417$ $\sigma = 1.683$
Mean	1			
Variance	16			
Squared coefficient of variation	16			

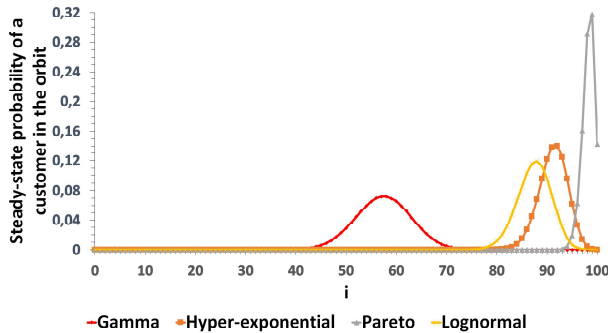


Fig. 2. Distribution of number of customers in the orbit, $\lambda/N = 0.01$

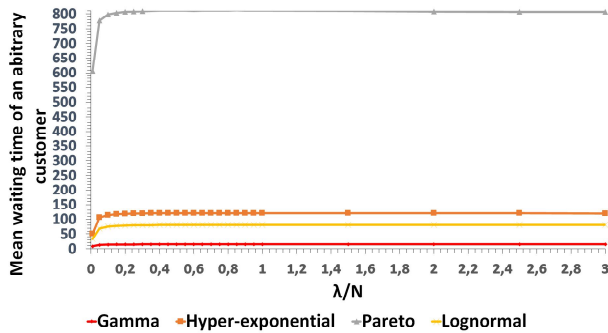


Fig. 3. Mean waiting time vs. arrival intensity

from each other which is especially true in the case of the gamma and the Pareto distribution. This figure ensures us that the selected distribution has a great impact on the operation of the system. The shape of all curves seems to correspond to Gaussian distribution.

Figure 3 shows the mean waiting of an arbitrary customer in the function of the arrival intensity. The same tendency occurs here regarding the results and a huge gap can be observed among the curves which are quite interesting. When Pareto distribution is applied customers spend averagely much more time in the orbit than in the case of other distributions. Because the source is finite the interesting maximum property characteristics appear despite the increasing arrival intensity which is true for every graph.

Figure 4 illustrates the number of collisions an arbitrary customer suffers under service during its residence in the system. Taking a closer look at the graphs the service of the customers is not interrupted by an arriving request most of the time when gamma distribution takes place. This can not be said to the others especially for Pareto distribution where customers undoubtedly experience more collision comparing

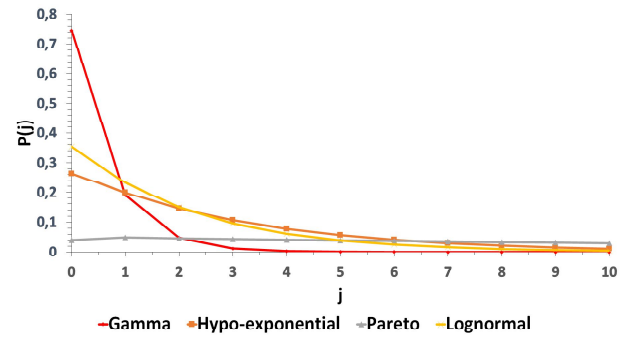
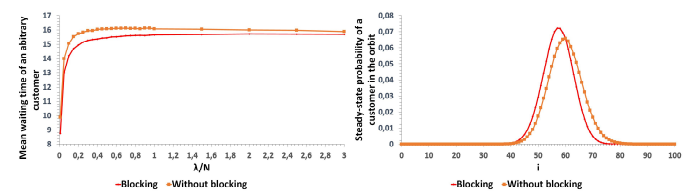


Fig. 4. The distribution of the number of collisions of an arbitrary customer

to the others. Near resemblance is noticeable to geometric distribution scrutinizing closely the results because the division of the consecutive probabilities is almost the same.



(a) Mean waiting time vs arrival inten- (b) Distribution of number of customers in the orbit

Fig. 5. The effect of blocking

Two graphs are presented in Figure 5, which depicts the effect of blocking on two performance measures. Examining closely the curves the expected behaviour is seen, lower values of mean waiting time, and less number of customers are experienced during running the code. However, the same characteristics remain in the case of without blocking as the mean waiting time has maximum value despite the increasing arrival intensity and the number of customers seems to follow normal distribution.

B. Different distributions of failure time of the server

In this scenario, after performing a sensitivity analysis on the service time of a customer we decided to carry out another one on the failure time of the service unit. Basically, we use the same parameter setting that Table II, Table I exhibit but these are the parameters of failure time. The service time of the customers is exponentially distributed with parameter μ . We analyze the same performance measures as in the previous section to see how the different distribution has an influence on the operation of our model.

Figure 6 displays what the probability is if exactly i customers are located in the orbit. The figure is completed with the reliable case when server failures do not occur to see the impact of this feature as well. The difference is not as huge as in the previous section but it is still presented among the utilized distributions and surprisingly we found the lowest mean value at gamma distribution. But similarly

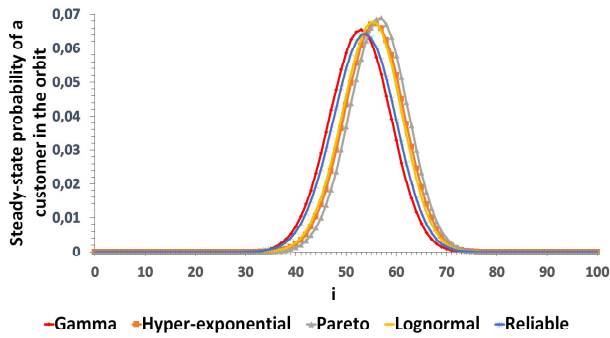


Fig. 6. Distribution of number of customers in the orbit, $\lambda/N = 0.1$

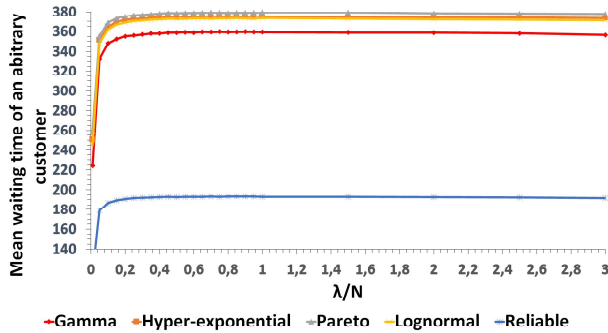


Fig. 7. Mean waiting time vs. arrival intensity

to Figure 2 the obtained results indicate that the distribution of this variable probably is normally distributed.

How the mean waiting time starts to increase and get stagnant after a while can be seen in Figure 7 in the function of arriving intensity. Of course, the mean waiting time is the lowest by far when there is no breakdown but interestingly the disparity among the other graphs is not relatively high except gamma distribution the gathered results are almost the same. Using this parameter setting still results in the phenomenon of maximum property characteristics, after a certain arrival intensity value the mean waiting of an arbitrary customer decreases.

The next figure (Figure 8) represents the distribution of the number of collisions of an arbitrary customer showing that what is the probability of a customer suffering exactly j collision under service. The lines nearly overlap each other illustrating that concerning this aspect there is no distinction among the cases. When the service unit is reliable customers are less exposed to go through collision but other than the curves are almost identical and reveal signs of decreasing tendency as the number of collisions increases.

Lastly, the comparison of the effect of blocking is shown in Figure 9 displaying both the mean waiting time of an arbitrary customer and the number of customers in the orbit. Indubitably, in the case of blocking higher values can be found and in this scenario, the difference is noticeable higher as in the previous section, but naturally, fewer customers could stay in the system if frequent outages take place.

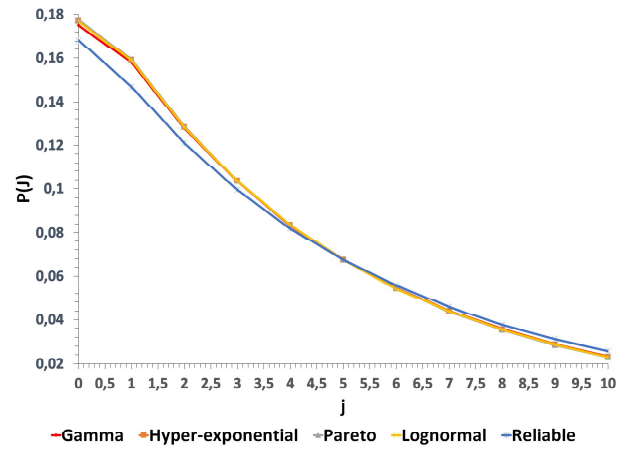
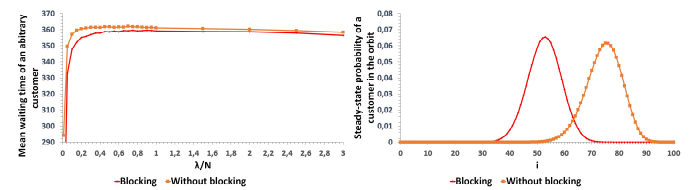


Fig. 8. The distribution of the number of collisions of an arbitrary customer



(a) Mean waiting time vs arrival inten- (b) Distribution of number of customers in the orbit

Fig. 9. The effect of blocking

V. CONCLUSION

A finite-source retrial queueing system is included with a non-reliable server, impatient customers and collisions in this work. In two scenarios we carry out a sensitivity analysis on two different random variables using numerous distributions to investigate the effect on some performance measures. Running our simulation program, graphical illustrations represented that in both scenarios the obtained values significantly differ from each other showing the importance of choosing a distribution having the same mean and variance when the squared coefficient of variation is greater than one. We also examined the effect of reliable operation demonstrating how remarkably changes the system behaviour compared to cases including random breakdowns. We also studied that applying the feature of blocking will lower the mean waiting time and the mean number of customers in the orbit. In the future, the authors want to continue their research work including more components in the model like catastrophic feature, more service units, or including other distributions.

ACKNOWLEDGMENT

The research was supported by the Thematic Excellence Programme (TKP2020-IKA-04) of the Ministry for Innovation and Technology in Hungary.

REFERENCES

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

- [2] G. Falin and J. Artalejo, "A finite source retrial queue," *European Journal of Operational Research*, vol. 108, pp. 409–424, 1998.
- [3] D. Fiems and T. Phung-Duc, "Light-traffic analysis of random access systems without collisions," *Annals of Operations Research*, pp. 1–17, 2017.
- [4] A. Gómez-Corral and T. Phung-Duc, "Retrial queues and related models," *Annals of Operations Research*, vol. 247, no. 1, pp. 1–2, 2016.
- [5] J. Kim and B. Kim, "A survey of retrial queueing systems," *Annals of Operations Research*, vol. 247, no. 1, pp. 3–36, 2016.
- [6] K. Rakesh and S. Sapana, "Transient performance analysis of a single server queueing model with retention of reneging customers," *Yugoslav Journal of Operations Research*, vol. 28, no. 3, pp. 315–331, 2018.
- [7] C. Kim, S. Dudin, A. Dudin, and K. Samouylov, "Analysis of a semi-open queueing network with a state dependent marked markovian arrival process, customers retrials and impatience," *Mathematics*, vol. 7, no. 8, pp. 715–734, 2019.
- [8] A. Kvach and A. Nazarov, "Sojourn time analysis of finite source markov retrial queueing system with collision," in *Information Technologies and Mathematical Modelling - Queueing Theory and Applications*, A. Dudin, A. Nazarov, and R. Yakupov, Eds. Cham: Springer International Publishing, 2015, pp. 64–72.
- [9] A. Kvach, "Numerical research of a Markov closed retrial queueing system without collisions and with the collision of the customers," in *Proceedings of Tomsk State University. A series of physics and mathematics. Tomsk*, ser. Materials of the II All-Russian Scientific Conference, vol. 295. TSU Publishing House, 2014, pp. 105–112, (In Russian).
- [10] A. Kvach and A. Nazarov, "Numerical research of a closed retrial queueing system M/GI/1//N with collision of the customers," in *Proceedings of Tomsk State University. A series of physics and mathematics. Tomsk*, ser. Materials of the III All-Russian Scientific Conference, vol. 297. TSU Publishing House, 2015, pp. 65–70, (In Russian).
- [11] A. Nazarov, A. Kvach, and V. Yampolsky, *Asymptotic Analysis of Closed Markov Retrial Queueing System with Collision*. Cham: Springer International Publishing, 2014, ch. 1, pp. 334–341.
- [12] 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.
- [13] N. Gharbi, B. Nemmouchi, L. Mokdad, and J. Ben-Othman, "The impact of breakdowns disciplines and repeated attempts on performances of small cell networks," *Journal of Computational Science*, vol. 5, no. 4, pp. 633–644, 2014.
- [14] A. Krishnamoorthy, P. K. Pramod, and S. R. Chakravarthy, "Queues with interruptions: a survey," *TOP*, vol. 22, no. 1, pp. 290–320, 2014.
- [15] J. Roszik, "Homogeneous finite-source retrial queues with server and sources subject to breakdowns and repairs," *Ann. Univ. Sci. Budap. Rolando Eötvös, Sect. Comput.*, vol. 23, pp. 213–227, 2004.
- [16] J. Sztrik, B. Almási, and J. Roszik, "Heterogeneous finite-source retrial queues with server subject to breakdowns and repairs," *Journal of Mathematical Sciences*, vol. 132, pp. 677–685, 2006.
- [17] 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.
- [18] 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.
- [19] Á. Tóth, T. Bérczes, J. Sztrik, and A. Kvach, "Simulation of finite-source retrial queueing systems with collisions and a non-reliable server," in *International Conference on Distributed Computer and Communication Networks*. Springer, 2017, pp. 146–158.
- [20] P. A. Fishwick, "Simpack: Getting started with simulation programming in c and c++," in *In 1992 Winter Simulation Conference*, 1992, pp. 154–162.
- [21] E. J. Chen and W. D. Kelton, "A procedure for generating batch-means confidence intervals for simulation: Checking independence and normality," *SIMULATION*, vol. 83, no. 10, pp. 683–694, 2007.
- [22] A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*. McGraw-Hill Education, 1991.
- [23] A. Nazarov, J. Sztrik, and A. Kvach, "A survey of recent results in finite-source retrial queues with collisions," in *Information Technologies and Mathematical Modelling. Queueing Theory and Applications*. Springer, 2018, pp. 1–15.
- [24] A. Nazarov, J. Sztrik, A. Kvach, and T. Bérczes, "Asymptotic analysis of finite-source M/M/1 retrial queueing system with collisions and server subject to breakdowns and repairs," *Annals of Operations Research*, vol. 277, no. 2, pp. 213–229, Jun 2019.
- [25] J. Sztrik, Á. Tóth, Á. Pintér, and Z. Bács, "Simulation of finite-source retrial queues with two-way communications to the orbit," in *Information Technologies and Mathematical Modelling. Queueing Theory and Applications*, A. Dudin, A. Nazarov, and A. Moiseev, Eds. Cham: Springer International Publishing, 2019, pp. 270–284.