Evsey V. Morozov
Alexander S. Rumyantsev
Oleg V. Lukashenko (Eds.)

Proceedings
of the Third International
Workshop **SMARTY'22**

Stochastic Modeling & Applied Research of TechnologY

Third International Workshop
Petrozavodsk, Karelia
August 21-25, 2022

Ádám Tóth and János Sztrik

Simulation of two-way communication retrial queueing systems with unreliable server and impatient customers in the orbit

Stochastic Modeling and Applied Research of Technology, Vol. 3, Pp. 45-50.

DOI: 10.57753/SMARTY.2023.39.42.006

Citation: Ádám Tóth and János Sztrik (2023). Simulation of two-way communication retrial queueing systems with unreliable server and impatient customers in the orbit. Stochastic Modeling and Applied Research of Technology, 3, 45-50. https://doi.org/10.57753/SMARTY.2023.39.42.006

Simulation of two-way communication retrial queueing systems with unreliable server and impatient customers in the orbit

Ádám Tóth and János Sztrik

Faculty of Informatics, University of Debrecen, Debrecen, Hungary,
toth.adam|sztrik.janos@inf.unideb.hu

Abstract. Models of two-way communication queueing systems play an important role because it is possible to model real-life scenarios utilized in many fields of life like in [2], [5]. It is a natural phenomenon having impatient customers in such systems resulting in earlier departure due to the long wait for being served [4], [3]. For this reason, we consider a two-way communication system with an unreliable server where primary customers may leave the system after residing in orbit for a certain amount of time. The failure of the service unit can occur during its operation or in an idle state, too. One important characteristic of our model is that the server generates requests to the customers from an infinite source in an idle state. These will be the secondary customers and they come into the system if the service unit is available and functional upon their arrival. Otherwise, they return without entering the system. In case of a server failure, every individual in the finite source may continue generating requests but these will be forwarded immediately towards the orbit. The source, service, retrial, impatience, operation, and repair times are supposed to be independent of each other. In this paper, we carry out a sensitivity analysis on some main performance measures to check the effect of different distributions of failure time.

1 Introduction

In recent years, the scrutiny of communication systems has become increasingly crucial, particularly in the realm of distributed computing systems. Given that companies and individuals rely on various telecommunication schemes and devices in their daily lives, there is a growing need to create and model both new and existing communication structures. Queuing systems with repeated calls are well-suited for effectively describing issues that arise in telephone switching systems, call centers, or computer systems. The primary characteristic of retrial queueing systems is that when every service unit is occupied, an incoming customer remains in the system, notably in a virtual waiting room known as the orbit. After a random period, customers in the orbit attempt to enter one of the service units.

Impatience is a common phenomenon in both everyday life and queueing models, and acknowledging this attribute enables a more accurate analysis.

There are numerous situations, particularly in healthcare applications, call centers, and telecommunication networks, where the impact of impatience significantly influences the system's operation. Impatience can be interpreted in different ways: *balking customers* decide not to join the queue if it is too long, *jockeying customers* can move from its queue to another queue if they detect they will get served faster and *reneging customers* leave the queue if they have waited a definite time for service.

In our investigated model customers have a reneging feature.

Another crucial factor is the service unit's availability, as breakdowns and unexpected errors can occur at any time. Many papers simplify the analysis by not considering failures, but such an assumption is quite unrealistic. The failure of the service unit significantly affects the system's characteristics and performance measures, making it worthwhile to examine the impact of server breakdowns [6].The uniqueness of this paper lies in conducting a sensitivity analysis using various failure time distributions on performance measures, such as the mean waiting time of an arbitrary, successfully served, or impatient customer, as well as the total utilization of the service unit, among others. This comparison is executed through a stochastic simulation program based on SimPack, forming the foundation of our program applicable for various algorithms related to discrete event simulation, continuous simulation, and combined (multi-model) simulation. Following the creation of the simulation model, we implemented all system features to calculate and estimate the desired measures using different values of input parameters. The most interesting results were selected for graphical representation, illustrating the impact of different parameter settings for failure time.

2 System model

In this section, we will delve into a more detailed description of the considered retrial queuing model featuring two-way communication. The model involves a finite source with N customers, each capable of generating a request (primary customers) towards the server at a rate λ. Consequently, the inter-request time follows an exponential distribution with the parameter $N\lambda$. Unlike a traditional queue, this model incorporates an orbit (virtual waiting room). The service of a primary customer commences instantly, following an exponential distribution time with the rate μ, if the server is idle. Otherwise, an incoming primary customer is directed to the orbit, where they await an exponentially distributed random time with the parameter σ to occupy the service unit. The server is assumed to experience breakdowns according to gamma, hyper-exponential, Pareto, and lognormal distributions, each with different parameters but the same mean value. Following a server breakdown, the repair process starts immediately, following an exponential distribution with the parameter γ_2.

Each primary customer in our analyzed model possesses an impatience attribute. In this model, a primary customer, after spending a certain time in the orbit, exits the system without receiving the intended service. This departure is modeled as an exponentially distributed random variable with a rate of τ.

The two-way communication feature comes into play when the server becomes idle. In such instances, an outgoing call is initiated after an exponentially distributed period with a parameter of γ. These customers enter the system if the server is neither in an error nor busy state upon their arrival. Otherwise, they are canceled and return to the infinite source.

The service time for outgoing or secondary customers follows a gamma distribution with parameters α_2 and β_2.

In the event of a failure during the service of any customer, primary customers are directed to the orbit, while secondary customers exit the system without completing their service.

3 Simulation results

In Table 1 the values of input parameters are shown used throughout the simulation. To obtain the results a statistics package is utilized to estimate the desired performance measures. Our code uses the method of batch means to collect a certain number of independent samples (batch means) by consecutive n observations of a steady-state simulation. It is one of the most widespread and common methods to define a confidence interval by gathering the steady-state mean of a process. To have the sample averages approximately independent the size of the batches should be carefully selected. By calculating the average of the sample averages of each batch we obtain the final mean value. About this technique you can find detailed information in the following work [1].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.

3.1 First scenario

In the first scenario, we have chosen the hyper-exponential, gamma, lognormal and Pareto distributions to investigate how these distributions of failure time alter the performance measures. In this case, the squared coefficient of variation is greater than one and the variance is relatively high. To accomplish an accurate comparison a fitting process has been done in order to have the same mean and variance value. Table 3 illustrates the exact values of the parameters of the failure time.

N	σ	γ	α_2	β_2	τ	μ	γ_2
100	0.01	0.8	1	2.5	0.001	1	1

Table 1. Numerical values of model parameters

Table 2. Parameters of failure time in scenario number one

Distribution	Gamma	Hyper-exponential	Pareto	Lognormal
Parameters	$\alpha = 0.174$ $\beta = 0.347$	$p = 0.42$ $\lambda_1 = 1.678$ $\lambda_2 = 2.322$	$\alpha = 2.083$ $k = 0.26$	$m = -1.649$ $\sigma = 1.382$
Mean	0.5			
Variance	1.44			
Squared coefficient of variation	5.76			

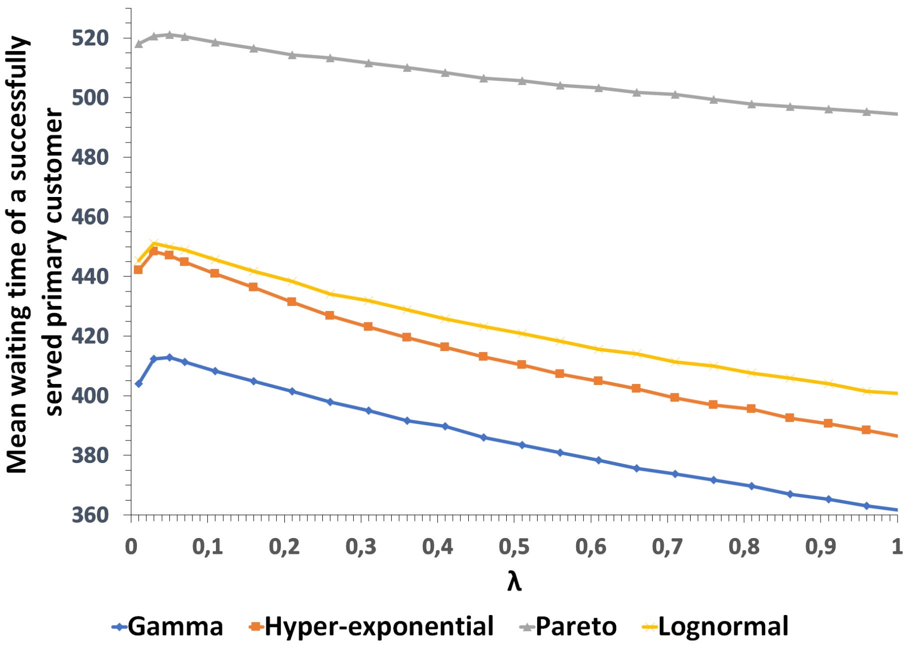

Fig. 1. Mean waiting time of a successfully served customer as the function of arrival intensity

The mean waiting time of a successfully served customer is presented in the function of the arrival intensity of the primary customers in Figure 1. Under this performance measure, we mean those customers who do not leave the system earlier because of impatience. Even though the mean and the variance are identical huge differences can be observed among the applied distributions, especially in the case of gamma and Pareto. As the arrival intensity increases, the average waiting time increases as well until λ equals 0.02 and after this measure starts to decrease. The same tendency occurs in the other distributions, as well. This is a special feature of retrial queueing systems under certain parameter settings.

3.2 Second scenario

In this section, the parameters of failure time are different: the mean value remains the same but the variance is less compared to the previous section. The squared coefficient of variation is still greater than one but now it is much closer to one. We were curious to see how this modification changes the behaviour of the system.

Table 3. Parameters of failure time in scenario number two

Distribution	Gamma	Hyper-exponential	Pareto	Lognormal
Parameters	$\alpha = 0.826$ $\beta = 1.653$	$p = 0.154$ $\lambda_1 = 0.617$ $\lambda_2 = 3.383$	$\alpha = 2.351$ $k = 0.287$	$m = -1.089$ $\sigma = 0.891$
Mean	0.5			
Variance	0.3025			
Squared coefficient of variation	1.21			

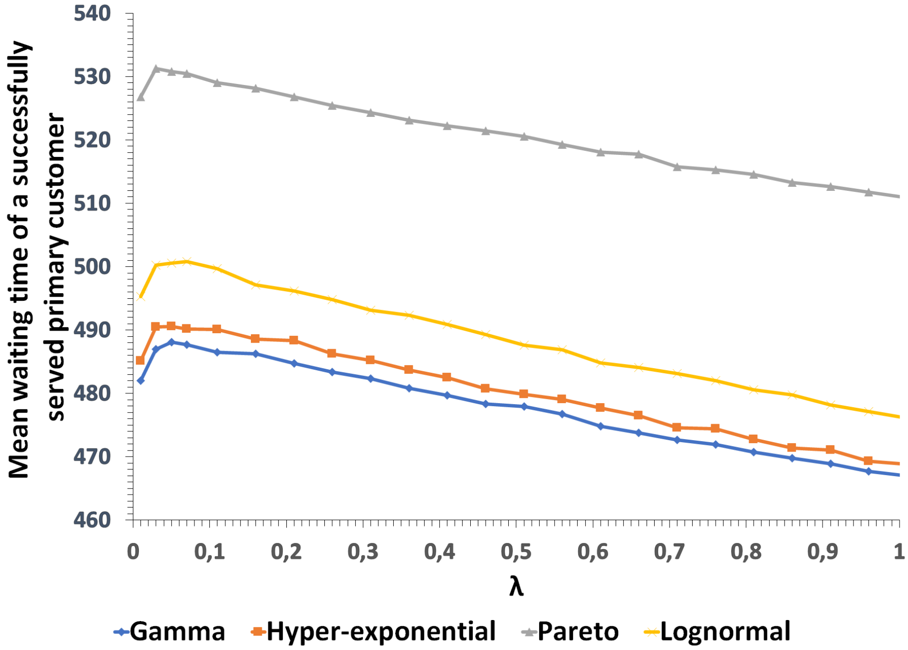

Fig. 2. Mean waiting time of a successfully served customer as the function of arrival intensity

Figure 2 demonstrates the development of the mean waiting time of a successfully served customer besides increasing arrival intensity. In this scenario, the mean value remained the same but the value of variance is much less. The difference in the average mean waiting time among the distributions is not quite significant except for the Pareto where the values are much higher. So it seems that the variance has a significant effect on the performance measures, larger values can result in greater disparities in the performance measures.

4 Conclusion

In this paper, we investigated a queueing system of the type $M/M/1//N$ featuring two-way communication, impatient customers, and an unreliable server. Simulation was employed to accommodate distributions beyond the exponential. The study demonstrated, under two scenarios, the sensitivity of the results to variations in the variance of the failure time while maintaining the same mean value. Additionally, it revealed that the use of the Pareto distribution resulted in the highest mean waiting times for successfully served customers in both scenarios. The authors intend to further delve into this phenomenon, expanding their research by incorporating additional features such as collisions, outgoing calls to customers from the orbit, and conducting further sensitivity analyses on other random variables.

References

1. 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)
2. Dragieva, V.I.: Number of retrials in a finite source retrial queue with unreliable server. Asia-Pac. J. Oper. Res. **31**(2), 23 (2014)
3. Kim, C., Dudin, S., Dudin, A., Samouylov, K.: Analysis of a semi-open queuing network with a state dependent marked markovian arrival process, customers retrials and impatience. Mathematics **7**(8), 715–734 (2019). https://doi.org/https://doi.org/10.3390/math7080715
4. Rakesh, K., Sapana, S.: Transient performance analysis of a single server queuing model with retention of reneging customers. Yugoslav Journal of Operations Research **28**(3), 315–331 (2018). https://doi.org/https://doi.org/10.2298/YJOR170415007K
5. Sztrik, J., Tóth, Á., Pintér, Á., Bács, Z.: Simulation of finite-source retrial queues with two-way communications to the orbit. In: Dudin, A., Nazarov, A., Moiseev, A. (eds.) Information Technologies and Mathematical Modelling. Queueing Theory and Applications. pp. 270–284. Springer International Publishing, Cham (2019)
6. Ádám Tóth, Sztrik, J.: Performance analysis of two-way communication retrial queueing systems with non-reliable server and impatient customers in the orbit. In: Proceedings of 1st Conference on Information Technology and Data Science. pp. 246–257. CEUR-WS (2020), http://ceur-ws.org/Vol-2874/#paper24