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

Ádám Tóth and János Sztrik*

August 21, 2022

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 [1], [2], It is a natural phenomenon having impatient customers in such systems resulting in earlier departure due to the long wait for being served ([3], [4]). For this reason, we consider a two-way communication system with an unreliable server where primary customers may leave the system after residing in the 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 towards the customers from an infinite source in idle state. These will be the secondary customers and they come into the system if the service unit is available and functional upon their arrivals. 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.

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

1 Introduction

In recent years the examination of communication systems is quite crucial and inevitable, especially in the field of distributed computing systems. Basically, companies and individuals use many telecommunication schemes and devices in everyday life, therefore, the creation and modelling of new or existing communication structures are needed, and queuing systems with repeated calls are perfectly suitable for describing arising issues in telephone switching systems, call centers, or computer systems. The main feature of retrial queueing systems is that if every service unit is occupied an incoming customer remains in the system notedly in a virtual waiting room called the orbit. These customers in the orbit after a random time attempt to enter one of the service units.

Impatience is a natural occurrence in everyday life and in queueing models and taking this property into consideration allows us to lead more precise analysis. A relatively great number of situations exist in healthcare applications, in call centers, and in telecommunication networks where the effect of impatience has a great impact on the operation of the system. 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.

The other key factor is the availability of the service unit which is essential because breakdowns and sudden or unexpected errors could take place at any time. A lot of papers do not take failure into consideration easing the analysis but that assumption is quite unrealistic. The failure of the service unit significantly has an influence on the system's characteristics as well on the performance measures thus it is worth investigating the effect of server breakdowns ([5]).The novelty of the present paper is to demonstrate a sensitivity analysis using various distributions of failure time on the performance measures for example the mean waiting time of an arbitrary, successfully served, and impatient customer or the total utilization of the service unit, etc. The realization of this comparison is made with the help of a stochastic simulation program based on SimPack. It is just the base of our program which can be used to perform various algorithms in connection with discrete event simulation, continuous simulation, and combined (multi-model) simulation. After creating the simulation model we implemented every feature of the system to calculate and estimate the desired measures using disparate values of input parameters. We selected the most interesting results to be depicted graphically demonstrating the effect of the different parameter settings of failure time.

2 System model

In this section, the considered retrial queuing model with two-way communication will be described in a more detailed aspect. There is a finite-source with N customers and each of them is able to generate a request (primary customers) towards the server with rate λ, so the inter-request time follows an exponential distribution with parameter λ. Instead of a queue the model has an orbit (virtual waiting room) so the service of a primary customer starts instantly according to exponential distribution time with the rate μ if the server is idle, otherwise, an incoming primary customer is forwarded to the orbit. Here they wait for an exponentially distributed random time with parameter σ to occupy the service unit. It is assumed to the server breaks down according to gamma, hyper-exponential, Pareto, and lognormal distribution with different parameters but with the same mean value. After a server breakdown, the repair process begins right away and this random variable follows an exponential distribution with parameter γ_2.

Every primary customer is characterized by an impatience property and in our investigated model a primary customer after waiting for some time in the orbit leaves the system without receiving its appropriate service which is also an exponentially distributed random variable with rate τ.

The feature of two-way communication takes place when the server becomes idle, in this case, an outgoing call is performed after an exponentially distributed period with parameter γ. These customers arrive in the system if the server is not in a failed or busy state upon their arrivals, otherwise, they will be cancelled and return to the infinite source. The service time of the outgoing or secondary customers is gamma distributed with parameters α_2 and β_2. If a failure occurs during the service of an arbitrary customer then the primary ones will be placed in the orbit and the secondary ones depart the system without continuing 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. 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	$\boldsymbol{\sigma}$	$\boldsymbol{\gamma}$	$\boldsymbol{\alpha_2}$	$\boldsymbol{\beta_2}$	$\boldsymbol{\tau}$	$\boldsymbol{\mu}$
100	0.01	0.8	1	2.5	0.001	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			

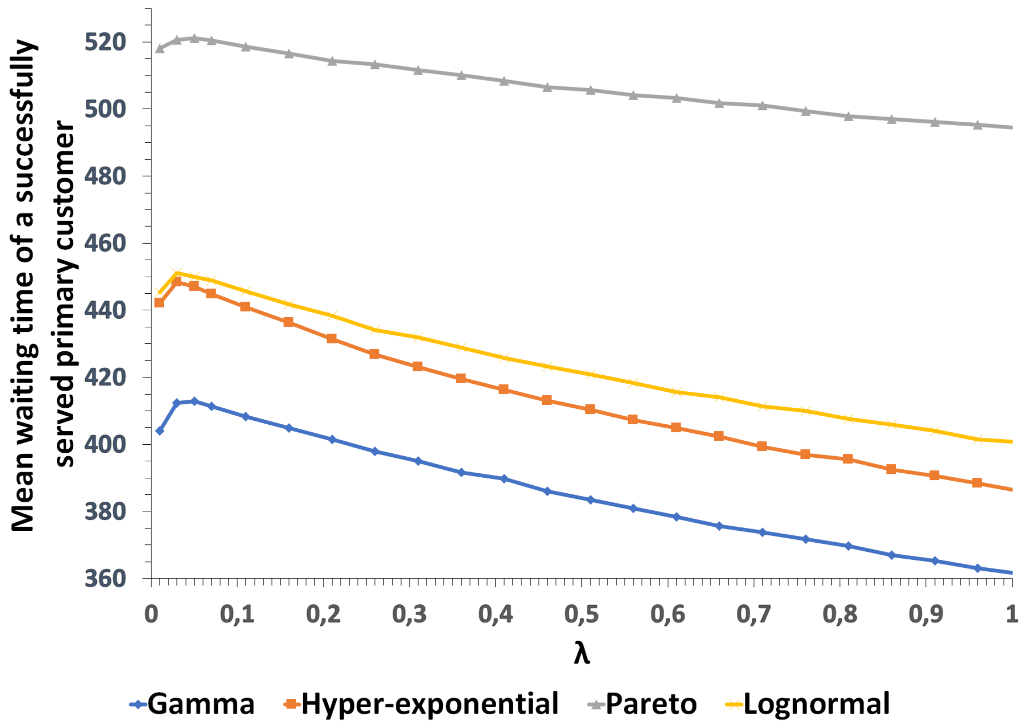

Figure 1: Mean waiting time of a successfully served customer in the function of arrival intensity

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			

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

We examined a queueing system of type $M/M/1//N$ with two-way communication feature, with impatient customers and with an unreliable server in this paper. Simulation has been realized to handle distributions apart from exponential. Under two scenarios, it was displayed how sensitive are the obtained results when the variance of the failure time differs significantly besides the same mean value. It also showed us that by applying Pareto distribution the values of mean waiting time of the successfully served customers are the highest in both scenarios. The authors plan to continue their research work, examining the obtained phenomenon in more detail and expand their model with other features like collisions, outgoing calls toward the

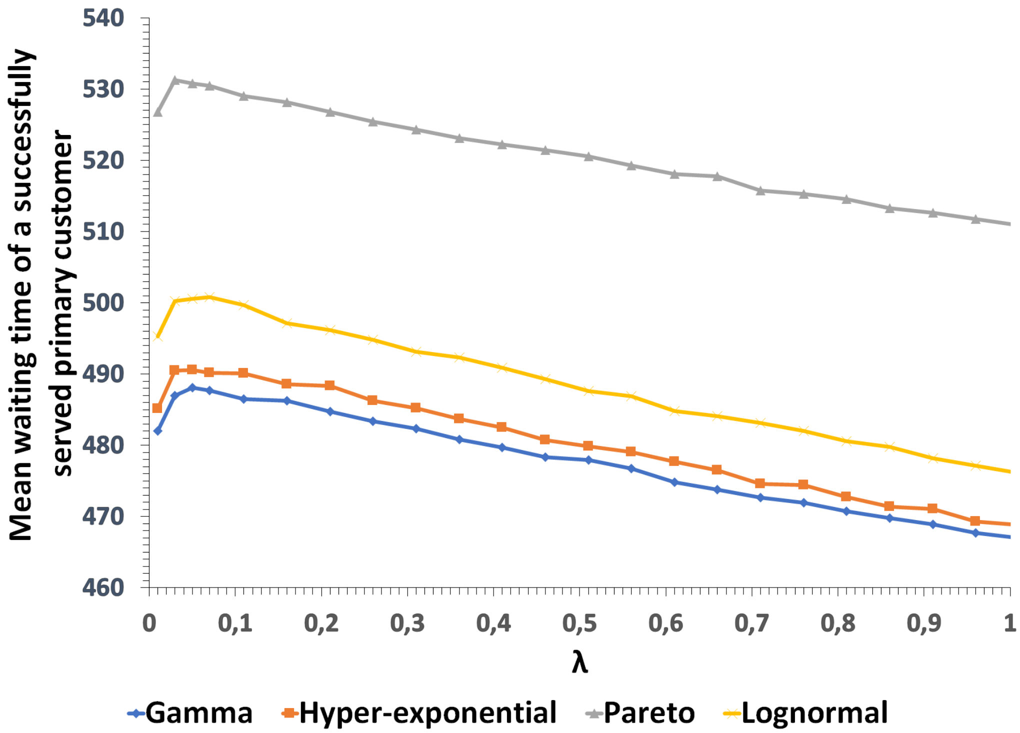

Figure 2: Mean waiting time of a successfully served customer in the function of arrival intensity

customers from the orbit, or carrying out other sensitivity analyses on other random variables.

References

[1] Dragieva Velika, I. Number of retrials in a finite source retrial queue with unreliable server, *Asia-Pacific Journal of Operational Research* Vol. 31, No. 2 (2014), 23

[2] 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. ITMM*

2019. Communications in Computer and Information Science Vol. 1109 (2019), 270–284

[3] Kumar, S., Sharma, S. Transient performance analysis of a single server queuing model with retention of reneging customers, *Yugoslav Journal of Operations Research* Vol. 28, Issue 3 (2018), 315–331

[4] Kim, C. S., 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* Vol. 7 (2019), 715–734

[5] Tóth, Á., Sztrik, J., Performance Analysis of Two-Way Communication Retrial Queueing Systems With Non-Reliable Server and Impatient Customers in the Orbit. *Proceedings of 1st Conference on Information Technology and Data Science*, Debrecen, Hungary (2020), CEUR Workshop Proceedings (CEUR-WS.org) Vol. 2874, (2020), 246–257