Simulation of Finite-Source Retrial Queueing Systems with Impatient Customers Using Different Failure Modes

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. In this paper, a finite-source retrial queueing system is considered with impatient customers and catastrophic breakdowns. The characteristic of the system includes collision which occurs when a new job arrives in the system and the service facility is occupied with a job, they will collide. Both jobs will be forwarded to the virtual waiting room the so-called orbit. Here, the customers initiate other attempts to reach the server after a random time. But they give up retrying after staying in the orbit a while and leave the system which is the impatient attribute of the customers. In case of a negative event, a catastrophic breakdown takes place meaning that all the customers at the server and in the orbit depart from the system. The novelty of this paper is to investigate that feature in a collision environment with impatient customers using different distributions of the service time.

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

1 Introduction

Designing info-communication systems are essential because of understanding how to optimize a system and also how to handle increasing network traffic. Many tools and mechanisms are available for modeling different systems, and among them, one of the most popular ones is retrial queuing systems. To illustrate real-life problems arising in main telecommunication systems, like telephone switching systems, call centers, computer networks, and computer systems, retrial queues can be effectively applied. In many publications, retrial-queuing systems with repeated calls are utilized to depict their models like in [2,5,6,9]. The specialty of retrial queuing systems relies on the orbit which is assumed to be a virtual waiting room with enough capacity to take in every customer. In this way, a job - whose service can not start - is not lost and may launch

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

A. Dudin et al. (Eds.): ITMM 2021, CCIS 1605, pp. 16–27, 2022.
https://doi.org/10.1007/978-3-031-09331-9_2

numerous attempts to get its service requirement. The source is considered to be finite mainly because in many situations a finite number of entities participate in the operation of the system. Naturally, researchers have studied models with an infinite source but these are not suitably describing real-life applications in many cases. Results in connection with finite-source retrial queuing systems can be viewed in [1,14,18,19]. Impatient behaviour is a natural characteristic of the customers provoking earlier departure without obtaining its service demand. This phenomenon is experienced in many fields of our life and here are some examples: healthcare applications, call centers, telecommunication networks. Not to mention all the papers where the behaviour of impatience is intensively examined, see for example [8,11,17]. Real-life systems tend to be subjected to random breakdowns which can be caused by a power outage, human negligence, or other sudden act. Thus, it is important to examine its effect on the operation of the system and the performance measures because it alters significantly the behaviour of a model. Many papers have studied models having service units assumed to be available all the time which is quite unrealistic. These types of systems have been investigated by many authors for example in [4,10,20]. In technologies, like in Ethernet or in communication sessions where the resources are constrained, the probability of collisions of the jobs occurs. Several individuals in the source may commence uncoordinated attempts leading to the interference of the signals resulting in the necessity for retransmissions. Consequently, it is important to include this phenomenon as part of the investigation creating effective policies preventing conflicts and corresponding message delays. Results that are in connection with collisions can be found in the following publication [12,13,15].

The objective of our investigation is to carry out a sensitivity analysis using different distributions of service times on the main performance measures while catastrophic breakdowns eventuate. In the case of these types of events, customers are forced to leave the system due to sudden acts which can be mechanical failures or power outages. Until repair, it is not allowed for any customer to enter the system and detailed studies on catastrophic breakdowns have been examined by several papers. Because we utilize different distributions for the service time of the customers the results are obtained by our simulation program that is based on Simpack [7]. The basic building blocks of the code are used in which we have the opportunity to calculate any desired measure using numerous values of input parameters. Graphical illustrations are provided depicting the effect of different parameters and distributions on the main performance metrics.

2 System Model

A finite-source retrial queueing system of type $M/G/1//N$ is considered with an unreliable service unit, impatient customers, the appearance of collisions, and blocking. This model has one service unit and a finite-source where every individual (altogether N) may generate a request towards the system according to exponential law with parameter λ/N meaning that the inter-arrival times are exponentially distributed with mean λ/N. As there are no queues the service of an arriving job starts immediately following gamma, hypo-exponential, hyper-exponential, Pareto, and lognormal distribution with different parameters but with the same mean and variance value. In the case of a busy server, an arriving customer brings about a collision with the customer under service, and both are moved into orbit. Jobs residing in the orbit after an exponentially random time with parameter σ/N initiate other tries to be engaged with the server. Since random breakdowns emerge the failure time is also an exponential random variable with parameter γ_0 when the server is occupied and with γ_1 if idle. Two scenarios are distinguished:

- general breakdown: the service of a job is interrupted and it is forwarded back to the orbit, other jobs initiated by the individuals of the source can not enter the system until the service unit is functional.
- catastrophic breakdown: the service of a job is interrupted but instead of arriving at the orbit it leaves the system as the others from the orbit, no customers are allowed by the system until the server fully recovers.

The repair process starts instantly upon the failure of the service unit which follows an exponential distribution with parameter γ_2. Customers are characterized by impatience implicating that jobs can decide to leave the system after spending an exponentially distributed time with parameter τ in the orbit. These requests return to the source being unserved. In the paper of [16] similar models are analyzed by an asymptotic method where N tends to infinity this is why rates λ/N and σ/N are used. For example, it was proved that the number of customers in the system follows a normal distribution. All the random variables in the model creation are assumed to be totally independent of each other.

3 Simulation Results

3.1 First Scenario

To obtain the desired results, our self-developed simulation tool was used in which almost all the performance measures can be estimated. Its statistics package utilizes the batch means method where the useful run is divided into a certain number of batches. Batches are long enough in that way sample averages of the batches are approximately independent thus we have a valid estimation. The following article contains more information about that method [3]. 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. The size of a batch used to detect the initial transient duration is 1000.

Table 1 consists of every parameter that is applied for all the following figures. The parameters of service time of the customers can be found at Table 2, every chosen parameter is listed resulting in the same mean and variance in every used distribution. The reason for selecting these values is focusing on the interesting

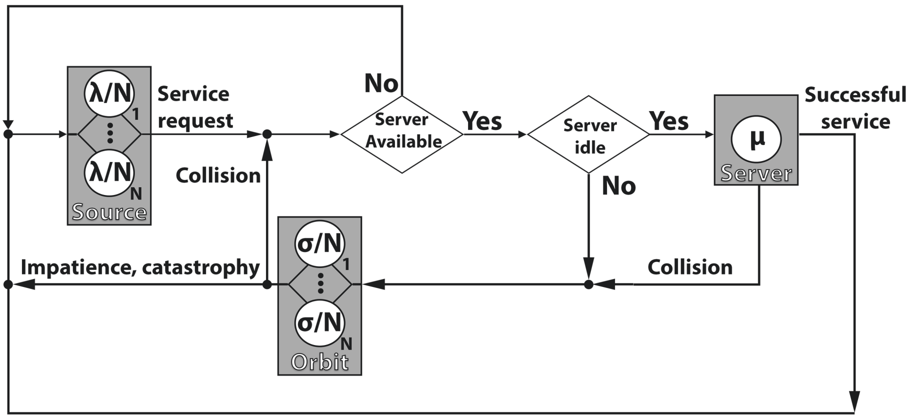

Fig. 1. System model

Table 1. Numerical values of model parameters

N	γ_0	γ_1	γ_2	σ/N	τ
100	0.05	0.05	1	0.05	0.001

situations and it must be noted that this model was tested with other values as well, and in most of the cases, the same phenomenon appeared. It is totally intentional that the squared coefficient of variation is more than one, later on in another scenario we will run the simulations when it is less than one (Fig. 1).

Table 2. Parameters of service time of primary customers

Distribution	Gamma	Hyper-exponential	Pareto	Lognormal
Parameters	$\alpha = 0.054$ $\beta = 0.077$	$p = 0.473$ $\lambda_1 = 1.353$ $\lambda_2 = 1.5$	$\alpha = 2.027$ $k = 0.355$	$m = -1.839$ $\sigma = 1.722$
Mean	0.7			
Variance	9			
Squared coefficient of variation	18.367			

On Fig. 2 and 3 on the X-axes i represents the number of customers located in the system, and on the Y-axes $P(i)$ denotes the probability that exactly i customer are situated at the server and in the orbit altogether. In both Fig. 2 and 3 the distribution of the number of customers in the system is displayed when λ/N is 0.1 using various distributions of service time. Catastrophic breakdown feature is applied and interestingly the mean number of customers in the system differs from each other. In the case of the gamma distribution, customers tend to spend less time in the system compared to Pareto distribution. It is also noticeable that for both types of breakdowns the distribution of the number of customers tends to follow Gaussian distribution.

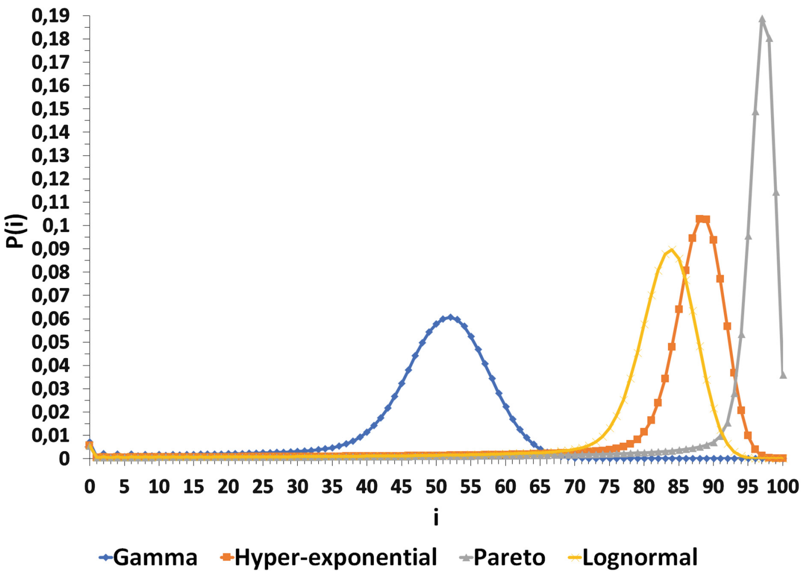

Fig. 2. Distribution of the number of customers in the system

Figure 3 depicts the comparison of different failure modes besides gamma and hyper-exponential distributions. Naturally more customers are in the system using the general breakdown method but the shape of the curves curiously are slightly disparate. In case of catastrophic breakdown, the peak is not that high and the mean number is fewer but other than that curves follow the same tendency.

The mean response time of an arbitrary customer is presented in the function of the arrival intensity of incoming customers in Fig. 4. Even though the mean and the variance are identical huge gaps develop among the applied distributions. With the increment of the arrival intensity, the mean response time of an arbitrary customer increases as well until λ/N equals 0.05 when the maximum is reached then it starts to decrease. The same tendency is observable for the other distributions, as well. The usage of gamma distribution results in a lower mean response time compared to the others, especially versus Pareto distribution.

Figure 5 demonstrates the development of the mean response time of a successfully served customer besides increasing arrival intensity. This measure shows

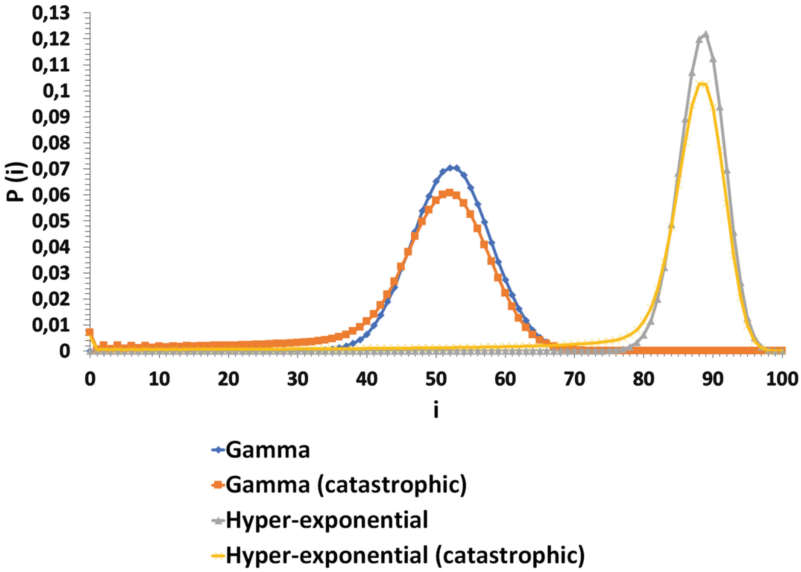

Fig. 3. Comparison of distribution of the number of customers in the system using different failure modes

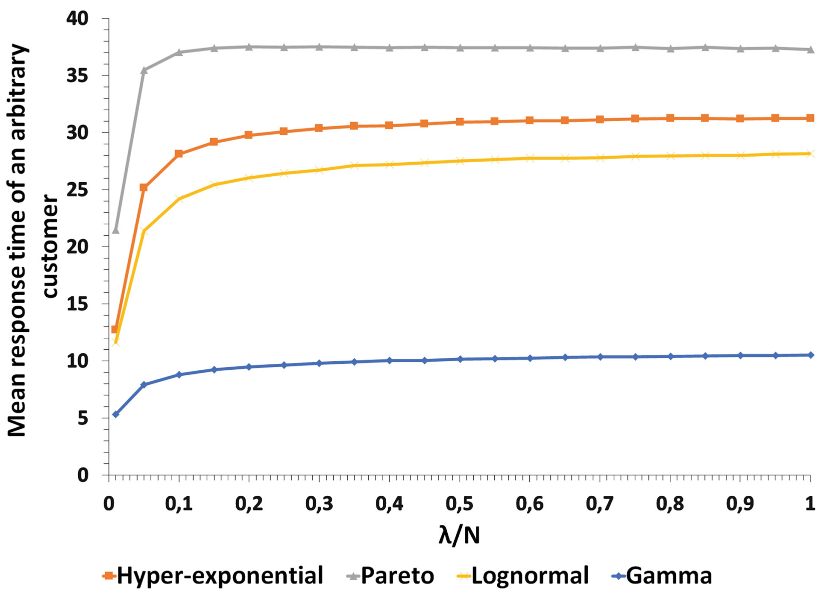

Fig. 4. Mean response time of an arbitrary customer vs. arrival intensity using various distributions.

the average response time of those customers who do not leave the system because of impatience or catastrophic event. As λ/N increases, the value of this performance measure raises as well which is true for every used distribution but the difference is quite high among them. At gamma distribution that value is much fewer than the others especially compared to Pareto distribution.

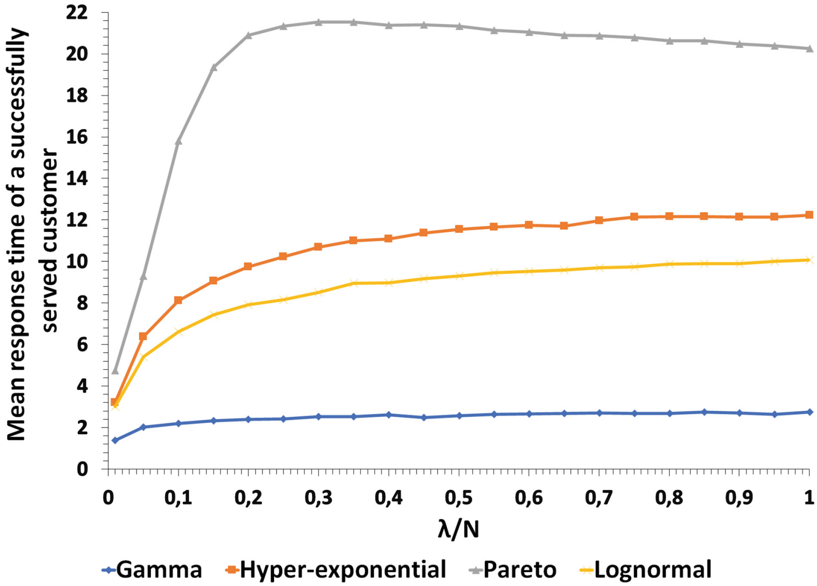

Fig. 5. Mean response time of a successfully served customer vs. arrival intensity using various distributions.

3.2 Second Scenario

In this section after analysing the obtained results of the previous scenario, we were curious to see what happens besides applying another parameter setting on the performance measures. In scenario 1 the squared coefficient of variation was greater than one and in this particular case, the parameters are selected in a way that the squared coefficient of variation is less than one. This also implies that the hyper-exponential distribution can not be used and instead of it we replace it with the hypo-exponential distribution. Table 3 contains the exact values of the parameters of the service time of primary customers in the case of this scenario, the other parameters remain unchanged which is shown in Table 1. Basically, our intention is to check that whether we get back the same tendencies of the previous section or it greatly changes the behaviour of the system and the performance measures with these modified parameters of service time of the incoming customers.

Table 3. Parameters of service time of incoming customers

Distribution	Gamma	Hypo-exponential	Pareto	Lognormal
Parameters	$\alpha = 1.69$ $\beta = 2.41$	$\mu_1 = 2$ $\mu_2 = 5$	$\alpha = 2.64$ $k = 0.435$	$m = -0.589$ $\sigma = 0.682$
Mean	0.7			
Variance	0.29			
Squared coefficient of variation	0.592			

First, we will examine the figures in connection to the steady-state distribution. Analyzing the curves in more detail the obtained values are much closer to each other. As regards the shape of the curves they correspond to normal distribution. The mean number of customers is higher in the case of every distribution compared to the previous section. Not much difference is experienced though. In Fig. 6 regarding the mean values, they are very close to each other as well the shape of the curves, but in this case, the obtained graphs do not tend to correspond to Gaussian distribution.

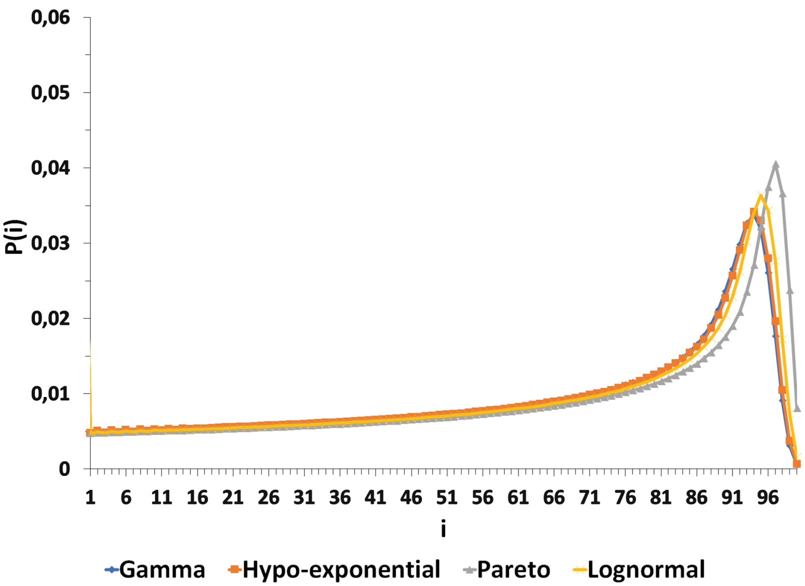

Fig. 6. Distribution of the number of customers in the system using various distributions, $\lambda = 0.1$.

Figure 7 emphasizes the difference between the applied failure modes. The results are depicted when gamma and hypo-exponential distribution are used but it is worth mentioning that the same tendencies occur utilizing the other two remaining ones. The difference is quite obvious even though the peak points are located in the same place but the value of possibility is much higher when catastrophe does not take place.

The next two figures are related to the mean response time of an arbitrary and a successfully served customer. First, in Fig. 8 it can be seen slight differences, in the case of Pareto distribution the values are a little bit higher, otherwise, the graphs almost overlap each other. Here, the same tendency develops as the mean response time increases with the increment of arrival intensity. Obviously, this maximum value feature is a specialty of finite-source retrial queuing systems under a suitable parameter setting.

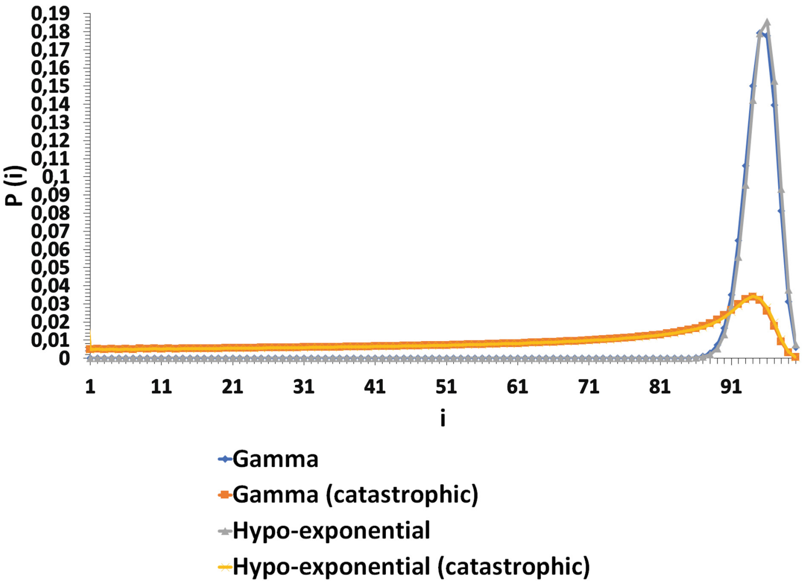

Fig. 7. Distribution of the number of customers in the system using various distributions, $\lambda/N = 0.1$.

Figure 9 demonstrates the comparison of the mean response time of a successfully served customer versus the arrival intensity. Not surprisingly after seeing the curves of the previous figure, the difference in the obtained values are very similar and it can be stated that the same maximum value feature appears in every case. The lowest values are obtained when the service time follows gamma distribution and the highest when Pareto distribution is applied.

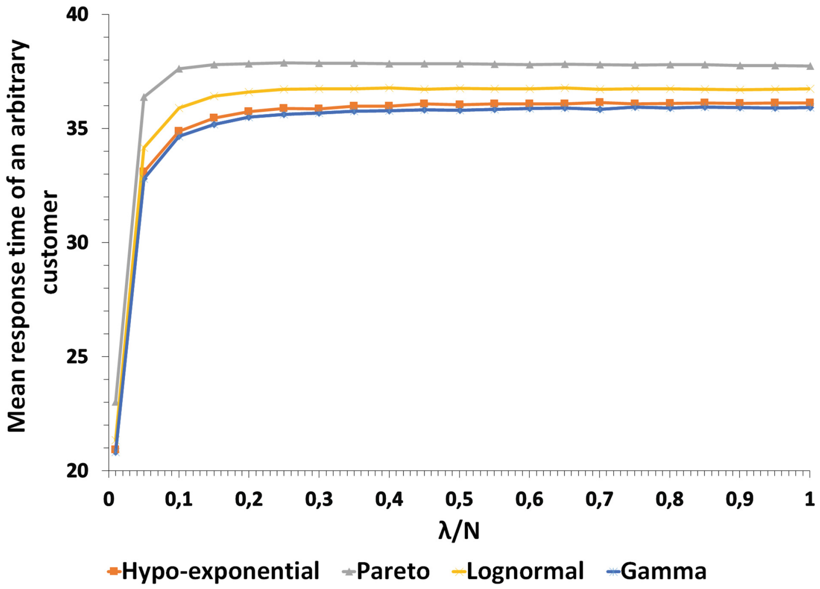

Fig. 8. Mean response time of an arbitrary customer vs. arrival intensity using various distributions.

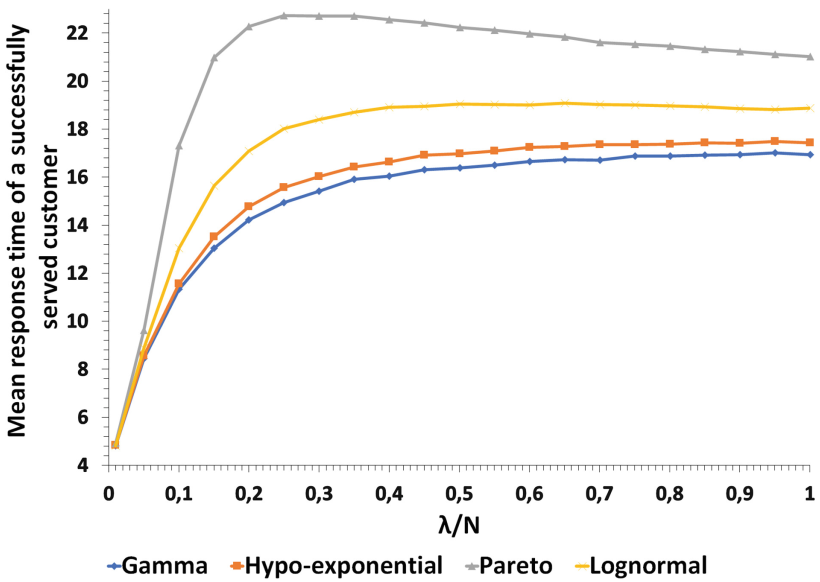

Fig. 9. Mean response time of a successfully served customer vs. arrival intensity using various distributions.

4 Conclusion

We simulated a retrial queueing system of type $M/G/1//N$ with impatient customers in the orbit and with an unreliable server using two different failure mechanisms when blocking is applied. Results are obtained by our program to carry out a sensitivity analysis on different performance measures like the distribution of the number of customers in the system. Under various parameter settings, the most interesting measures were chosen which were graphically illustrated. When the squared coefficient of variation is more than one significant deviation is experienced between the distributions in almost every aspect of the investigated measures. Consistently, it was also revealed that besides catastrophic breakdown less customer is in the system than in the case of a normal breakdown which is an expected phenomenon but the shape of the curves follows the same tendencies. In future works, the authors aim to carry on investigating the effect of catastrophic breakdown in other models and performing sensitivity analysis for other variables like the failure rate or the impatience of the customers.

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, Cham (2008). https://doi.org/10.1007/978-3-540-78725-9
3. 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)
4. Dragieva, V.I.: Number of retrials in a finite source retrial queue with unreliable server. Asia-Pac. J. Oper. Res. **31**(2), 23 (2014). https://doi.org/10.1142/S0217595914400053
5. Falin, G., Artalejo, J.: A finite source retrial queue. Eur. J. Oper. Res. **108**, 409–424 (1998)
6. Fiems, D., Phung-Duc, T.: Light-traffic analysis of random access systems without collisions. Ann. Oper. Res. **277**(2), 311–327 (2017). https://doi.org/10.1007/s10479-017-2636-7
7. Fishwick, P.A.: Simpack: getting started with simulation programming in C and C++. In: 1992 Winter Simulation Conference, pp. 154–162 (1992)
8. Gupta, N.: 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, vol. RTMC, no. 4, May 2012
9. Kim, J., Kim, B.: A survey of retrial queueing systems. Ann. Oper. Res. **247**(1), 3–36 (2015). https://doi.org/10.1007/s10479-015-2038-7
10. Krishnamoorthy, A., Pramod, P.K., Chakravarthy, S.R.: Queues with interruptions: a survey. TOP **22**(1), 290–320 (2012). https://doi.org/10.1007/s11750-012-0256-6
11. 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

12. Kvach, A., Nazarov, A.: Sojourn time analysis of finite source Markov retrial queuing system with collision. In: Dudin, A., Nazarov, A., Yakupov, R. (eds.) ITMM 2015. CCIS, vol. 564, pp. 64–72. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25861-4_6
13. Kvach, A., Nazarov, A.: 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. Materials of the III All-Russian Scientific Conference, vol. 297, pp. 65–70. TSU Publishing House (2015). (in Russian)
14. Lakaour, L., Aïssani, D., Adel-Aissanou, K., Barkaoui, K.: M/M/1 retrial queue with collisions and transmission errors. Methodol. Comput. Appl. Probab. **21**(4), 1395–1406 (2018). https://doi.org/10.1007/s11009-018-9680-x
15. Nazarov, A., Kvach, A., Yampolsky, V.: Asymptotic analysis of closed Markov retrial queuing system with collision. In: Dudin, A., Nazarov, A., Yakupov, R., Gortsev, A. (eds.) ITMM 2014. CCIS, vol. 487, pp. 334–341. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13671-4_38
16. Nazarov, A., Sztrik, J., Kvach, A., Bérczes, T.: Asymptotic analysis of finite-source M/M/1 retrial queueing system with collisions and server subject to breakdowns and repairs. Ann. Oper. Res. **277**(2), 213–229 (2018). https://doi.org/10.1007/s10479-018-2894-z
17. 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
18. 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
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