Russian Academy of Sciences (RAS)
V.A. Trapeznikov Institute of Control Sciences of RAS (ICS RAS)
Peoples' Friendship University of Russia (RUDN University)
Institute of Information and Communication Technologies
of Bulgarian Academy of Sciences (Sofia, Bulgaria)
National Research Tomsk State University (NR TSU)
Research and development company
"Information and networking technologies"

DISTRIBUTED COMPUTER AND COMMUNICATION NETWORKS: CONTROL, COMPUTATION, COMMUNICATIONS (DCCN–2024)

DCCN
2024

PROCEEDINGS

OF THE XXVII INTERNATIONAL SCIENTIFIC CONFERENCE

Russia, Moscow, September 23–27, 2024

Moscow
Peoples' Friendship University of Russia Russia Named After Patrice Lumumba
2024

DISTRIBUTED COMPUTER AND COMMUNICATION NETWORKS: CONTROL, COMPUTATION, COMMUNICATIONS (DCCN-2024)

9 785209 118770

Российская академия наук (РАН)

**Институт проблем управления им. В.А. Трапезникова
Российской академии наук (ИПУ РАН)**

Российский университет дружбы народов (РУДН)

**Институт информационных и телекоммуникационных технологий
Болгарской академии наук (София, Болгария)**

Национальный исследовательский Томский государственный университет (НИ ТГУ)

**Научно-производственное объединение
«Информационные и сетевые технологии» («ИНСЕТ»)**

РАСПРЕДЕЛЕННЫЕ КОМПЬЮТЕРНЫЕ И ТЕЛЕКОММУНИКАЦИОННЫЕ СЕТИ: УПРАВЛЕНИЕ, ВЫЧИСЛЕНИЕ, СВЯЗЬ (DCCN-2024)

Материалы XXVII Международной научной конференции

Россия, Москва, 23–27 сентября 2024 г.

Под общей редакцией
д.т.н. *В.М. Вишневского* и д.т.н. *К.Е. Самуйлова*

**Москва
Российский университет дружбы народов
им. Патриса Лумумбы
2024**

Russian Academy of Sciences (RAS)
V.A. Trapeznikov Institute of Control Sciences of RAS (ICS RAS)
Peoples' Friendship University of Russia (RUDN University)
Institute of Information and Communication Technologies
of Bulgarian Academy of Sciences (Sofia, Bulgaria)
National Research Tomsk State University (NR TSU)
Research and development company
"Information and networking technologies"

DISTRIBUTED COMPUTER AND COMMUNICATION NETWORKS: CONTROL, COMPUTATION, COMMUNICATIONS (DCCN-2024)

Proceedings of the XXVII International Scientific Conference

Russia, Moscow, September 23–27, 2024

Under the general editorship
of D.Sc. *V.M. Vishnevskiy* and D.Sc. *K.E. Samouylov*

Moscow
Peoples' Friendship University of Russia
Named After Patrice Lumumba
2024

УДК 004.7:004.4:621.391:007:001(063)
ББК 32.973.202:32.968+32.81
Р24

П о д о б щ е й р е д а к ц и е й
д.т.н. *В.М. Вишневского* и д.т.н. *К.Е. Самуйлова*

Р24 **Распределенные компьютерные и телекоммуникационные сети : управление, вычисление, связь (DCCN-2024) = Distributed computer and communication networks : control, computation, communications (DCCN-2024)** : материалы XXVII Международной научной конференции. Россия, Москва, 23–27 сентября 2024 г. / под общ. ред. В. М. Вишневского и К. Е. Самуйлова. – Москва : РУДН, 2024. – 319 с. : ил.

В научном издании представлены материалы XXVII Международной научной конференции «Распределенные компьютерные и телекоммуникационные сети: управление, вычисление, связь» по следующим направлениям:

- Алгоритмы и протоколы телекоммуникационных сетей;
- Управление в компьютерных и инфокоммуникационных системах;
- Анализ производительности, оценка QoS / QoE и эффективность сетей;
- Аналитическое и имитационное моделирование коммуникационных систем последующих поколений;
- Эволюция беспроводных сетей в направлении 5G;
- Технологии сантиметрового и миллиметрового диапазона радиоволн;
- RFID-технологии и их приложения;
- Интернет вещей и туманные вычисления;
- Системы облачного вычисления, распределенные и параллельные системы;
- Анализ больших данных;
- Вероятностные и статистические модели в информационных системах;
- Теория массового обслуживания, теория надежности и их приложения;
- Высотные беспилотные платформы и летательные аппараты: управление, передача данных, приложения.

В материалах научной конференции DCCN-2024, подготовленных к выпуску к.ф.-м.н. Д.В. Козыревым, обсуждены перспективы развития и сотрудничества в этой сфере.

Сборник материалов конференции предназначен для научных работников и специалистов в области управления крупномасштабными системами.

Текст воспроизводится в том виде, в котором представлен авторами.

ISBN 978-5-209-11877-0

UDC: 004.94

Investigation of M/G/1//N system with collisions, unreliable primary and a backup server

Ádám Tóth [1] and János Sztrik[1]

[1]University of Debrecen, Debrecen H-4032 Egyetem tér 1., Hungary

{toth.adam,sztrik.janos}@inf.unideb.hu

Abstract

This paper explores a finite-source retrial queueing system featuring collisions of the requests, unreliability of the primary server and a backup server. In the case of collisions, wherein a new job arriving when the service facility is occupied results in a collision, sending both jobs to a virtual waiting room, termed the orbit. In the orbit, customers initiate further attempts to access the server after a random time interval. In the event of a breakdown, the customer at the server is forwarded to the orbit. The novelty of this study lies in implementing a backup facility when the primary server is unreachable and carrying out a sensitivity analysis employing various service time distributions of the primary customers. We investigated a scenario where the most important performance measures are visually represented highlighting the observed disparities.

Keywords: Simulation, Queueing system, Finite-source model, Sensitivity analysis, Backup server, Unreliable operation, Collision.

1. Introduction

In the contemporary context characterized by escalating traffic volumes and expanding user bases, the analysis of communication systems or the design of optimal configurations poses a formidable challenge. Given the pivotal role of information exchange across all spheres of life, it becomes imperative to develop or adapt mathematical and simulation models for telecommunication systems to align with these evolving dynamics. Retrial queues emerge as potent and apt tools for modeling real-world scenarios encountered in telecommunication systems, networks, mobile networks, call centers, and analogous domains. A plethora of scholarly works, exemplified by references such as [1] and [2], have been dedicated to investigating various manifestations of retrial queuing systems characterized by recurrent calls.

In certain contexts, researchers postulate the perpetual availability of service units, yet operational interruptions or unexpected events may occur, resulting in the rejection of incoming customers. Devices deployed across diverse industries are susceptible to malfunctions, rendering the presumption of their infallible operation overly sanguine and impractical. Likewise, within wireless communication environments, diverse factors can impinge upon transmission rates, precipitating interruptions during packet delivery. The inherent unreliability of retrial queuing systems significantly influences system functionality and performance metrics. Concurrently, halting production entirely is unviable, as it may engender delays in order fulfillment. Therefore, amidst such occurrences, machines or operators endowed with lower processing capacities may continue operating to sustain smoother functionality. Moreover, the authors investigate the viability of incorporating a backup server capable of delivering services at a diminished rate in instances of primary server unavailability. Numerous recent scholarly works have extensively examined retrial queuing systems featuring unreliable servers, as exemplified by references such as [3].

In service-oriented domains, service providers frequently encounter operational disruptions stemming from various factors, including database accessibility issues hindering the fulfillment of customer requests. In response to such disruptions, service providers commonly resort to contingency measures such as activating backup systems or eliciting additional information from customers to facilitate resolution.

In technological contexts such as Ethernet networks or constrained communication sessions, the occurrence of job collisions is probable. Multiple entities within the source may initiate asynchronous attempts, causing signal interference and necessitating retransmissions. Hence, it is imperative to incorporate this phenomenon into investigations aimed at devising effective policies to mitigate conflicts and associated message delays. Publications addressing results related to collisions include [4], and [5].

The aim of our study is to conduct a sensitivity analysis, employing diverse service time distributions of the primary server, to assess the main performance metrics under scenarios involving a backup facility. During failure periods of the primary server, the service of the customers is traversed to the backup service facility and until restoration, new customers are permitted to reach the backup unit or the orbit if it is busy. Our investigation emphasize the effect of a backup service unit and the results are obtained through simulation using Simpack [6]. The simulation program is developed upon fundamental code elements enabling the computation of desired metrics across a range of input parameters. Graphical representations are provided to elucidate the impact of different parameters and distributions on the primary performance indicators.

2. System model

We examine a finite-source retrial queueing system characterized by type $M/G/1//N$, incorporating an unreliable primary service unit, occurrences of collisions, and a backup service unit. This model features a finite source, with each of the N individuals generating requests to the system according to an exponential distribution with parameter λ. Arrival times follow an exponential distribution with a mean of $\lambda * N$. With no queues present, service for arriving jobs commences immediately following a gamma, hypo-exponential, hyper-exponential, Pareto, or lognormal distribution, each with distinct parameters but equivalent mean and variance values. In instances of server congestion, an arriving customer triggers a collision with the customer currently under service, resulting in both being transferred to the orbit. Jobs residing in the orbit initiate further attempts to access the server after an exponentially distributed random time with parameter σ. Additionally, random breakdowns occur, with failure times represented by exponential random variables. The failure time has a parameter of γ_0 when the server is occupied and γ_1 when idle.

Upon the failure of the service unit, the repair process commences immediately, with the duration of the repair following an exponential distribution characterized by parameter γ_2.

In the event of a busy server experiencing a failure, the customer is promptly transitioned to the orbit. Despite the unavailability of the service unit, all customers in the source retain the capability to generate requests, albeit directed towards the backup server, which operates at a reduced rate characterized by an exponentially distributed random variable with parameter μ during periods of primary server unavailability. The backup server is assumed to be reliable and operates solely in the absence of the primary server. When the backup server is occupied, incoming requests are directed to the orbit. The phenomenon of collision does not occur in front of the backup service unit. The model assumes complete independence among all random variables during its formulation.

3. Simulation results

We employed a statistical module class equipped with a statistical analysis tool to quantitatively estimate the mean and variance values of observed variables using the batch mean method. This method involves aggregating n successive observations from a steady-state simulation to generate a sequence of independent samples. The batch mean method is a widely utilized technique for establishing confidence intervals for the steady-state mean of a process. To ensure that the sample averages are approximately independent, large batches are necessary. Further details on the batch mean method can be found in [7]. Our simulations were conducted with a confidence level of 99.9%, and the simulation run was terminated once the relative half-width of the confidence interval reached 0.00001.

N	γ_0	γ_1	γ_2	σ	μ
100	0.1	0.1	1	0.05	0.6

Table 1. Numerical values of model parameters

In this section, our objective was to determine the service time parameters for each distribution in a manner that ensures equal mean values and variances. Four distinct distributions were examined to assess their influence on performance metrics. Specifically, the hyper-exponential distribution was selected to ensure a squared coefficient of variation greater than one. The input parameters of the various distributions are presented in Table 2, while Table 1 provides the values of other relevant parameters.

Table 2. Parameters of service time of primary customers

Distribution	**Gamma**	**Hyper-exponential**	**Pareto**	**Lognormal**
Parameters	$\alpha = 0.011$ $\beta = 0.011$	$p = 0.494$ $\lambda_1 = 0.989$ $\lambda_2 = 1.011$	$\alpha = 2.005$ $k = 0.501$	$m = -2.257$ $\sigma = 2.125$
Mean	1			
Variance	90.25			
Squared coefficient of variation	90.25			

Figure 1 depicts the correlation between the mean response time of customers and the arrival intensity. The Pareto distribution exhibits the highest mean response time, while the distinctions among the other distribution types become more apparent. Notably, the gamma distribution stands out by yielding the lowest mean response time. An intriguing observation is that, as the arrival intensity increases, the mean response time initially rises but subsequently decreases after reaching a specific threshold. This behavior is a characteristic feature of retrial queuing systems with a finite source, and it tends to manifest under appropriate parameter configurations.

Figure 2 illustrates the utilization of the service unit in relation to the arrival rate of incoming customers. Despite possessing identical mean and variance values, notable distinctions are observed among different distributions. As the arrival rate escalates, the utilization of the service unit correspondingly rises. Specifically, the utilization rate is lower with the gamma distribution compared to other distributions, particularly evident with the hyper-exponential distribution. Interestingly, in the case of Pareto distribution the tendency is reversed as the utilization of the primary service unit starts to decrease besides increasing arrival intensity.

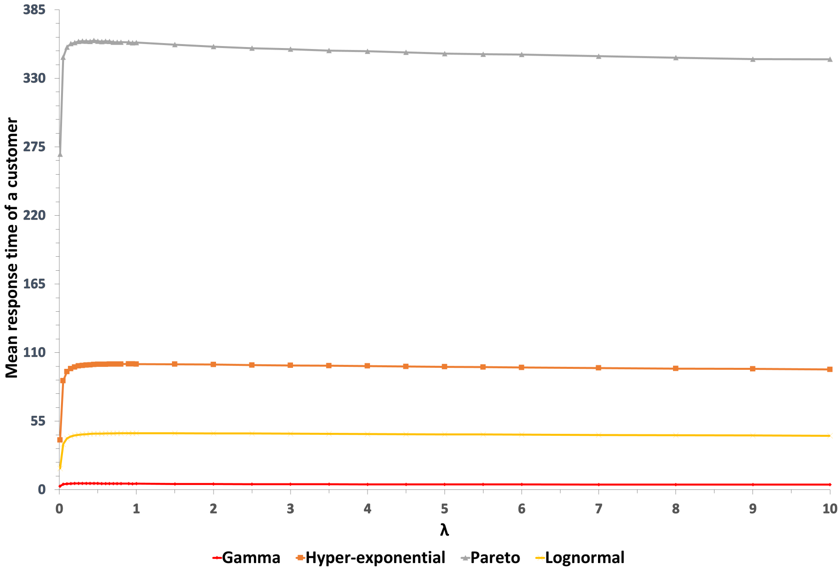

Fig. 1. Mean response time vs. arrival intensity

4. Conclusion

We conducted simulations of a retrial queuing system following the $M/G/1//N$ model, incorporating an unreliable primary server and a backup service unit. Our program was utilized to perform a sensitivity analysis on various performance metrics, including the mean response of times of the customers. From a multitude of parameter configurations, the most relevant measures were selected and graphically depicted. Notably, when the squared coefficient of variation exceeds one, significant deviations are observed among distributions across multiple aspects of the investigated metrics. In future studies, the authors intend to further explore the impact of server blocking, impatience of the customers in alternative models and conduct sensitivity analyses for other variables, such as failure rates.

REFERENCES

1. V. I. Dragieva, Number of retrials in a finite source retrial queue with unreliable server., Asia-Pac. J. Oper. Res. 31 (2) (2014) 23. doi:10.1142/S0217595914400053.
2. D. Fiems, T. Phung-Duc, Light-traffic analysis of random access systems without collisions, Annals of Operations Research (2017) 1–17.

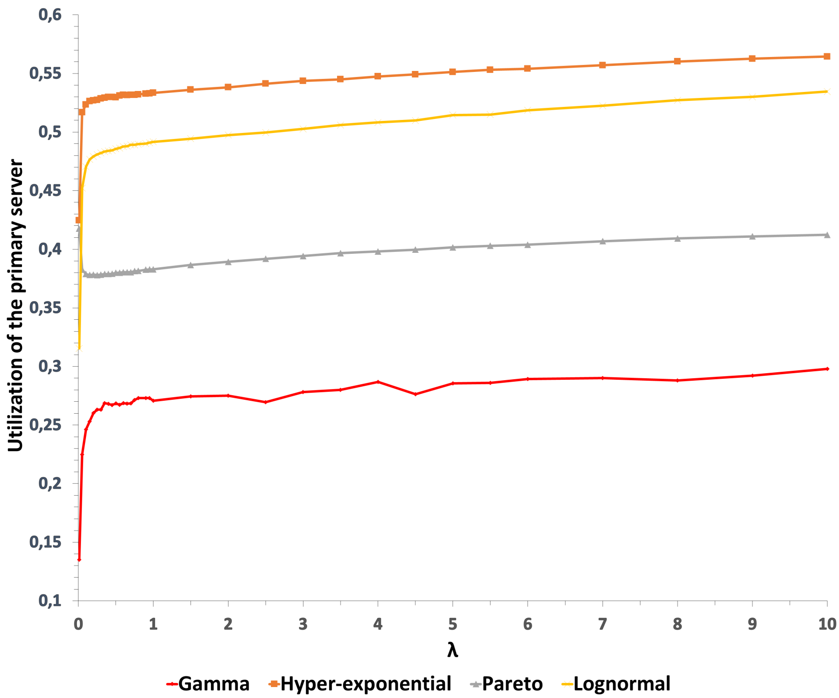

Fig. 2. Comparison of utilization

3. N. Gharbi, B. Nemmouchi, L. Mokdad, J. Ben-Othman, The impact of breakdowns disciplines and repeated attempts on performances of small cell networks, Journal of Computational Science 5 (4) (2014) 633–644.
4. A. Kvach, A. Nazarov, Sojourn Time Analysis of Finite Source Markov Retrial Queuing System with Collision, Springer International Publishing, Cham, 2015, Ch. 8, pp. 64–72.
5. A. Nazarov, A. Kvach, V. Yampolsky, Asymptotic Analysis of Closed Markov Retrial Queuing System with Collision, Springer International Publishing, Cham, 2014, Ch. 1, pp. 334–341.
6. P. A. Fishwick, Simpack: Getting started with simulation programming in c and c++, in: In 1992 Winter Simulation Conference, 1992, pp. 154–162.
7. E. J. Chen, W. D. Kelton, A procedure for generating batch-means confidence intervals for simulation: Checking independence and normality, SIMULATION 83 (10) (2007) 683–694.