Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/54/20241227

Research on the queuing theory in practical applications

Yuheng Li

Department of Software Engineering, Shandong University, Jinan, 250100, China

liyuheng@mail.sdu.edu.cn

Abstract. In the face of mounting global urbanization and digitization trends, the need for advanced tools for city planners and system managers becomes increasingly paramount to ensure seamless infrastructure operations. Among the arsenal of available tools, queuing theory emerges as a standout, offering invaluable predictions and strategies for a broad spectrum of situations. This article delves into the nuanced applications of queuing theory, with a specific lens on network communications and urban space planning. Drawing from a rich tapestry of academic sources, the narrative weaves together core principles to shape models that mirror real-world situations. At the heart of this exploration lies a deep dive into solutions that tackle network delay challenges, fine-tuning techniques for 6TiSCH resource allocation, and the subtle art of queue design at railway ticket counters. These instances highlight the adaptability and immediate relevance of queuing theory across various sectors. Those in the fields of design, system architecture, and urban planning will find this read enlightening. By leveraging the insights offered, decision-makers can pave the way for optimized system functionalities and heightened user experiences, vital in an era dominated by urban sprawl and digital transformation.

Keywords: Queuing Theory, Application, Literature Review.

1. Introduction

The origins of queuing theory can be traced back to analyses and studies related to the incoming call systems of telephone exchanges. Today, the application of queuing theory spans a diverse array of sectors, from telecommunications, computing, and transportation to logistics, warehouse management, and the layout of hospitals and factories. Urban planners harness queuing theory to refine public transportation systems—streamlining bus and train schedules, reducing wait times, and enhancing passenger throughput.

While there's no shortage of literature dedicated to the practical applications of queuing theory, there's a noticeable gap in research that critically analyzes, evaluates, and addresses the challenges in these applications. This paper seeks to delve into the intricacies and challenges that arise when queuing theory intersects with real-world scenarios, and when it's combined with other theoretical frameworks. By providing such an analysis, the intention is to offer clarity and guidance for scholars venturing into similar realms of research. Drawing from real-world use cases of queuing theory, this paper amalgamates findings from prior studies, offering fresh perspectives and novel insights.

DOI: 10.54254/2755-2721/54/20241227

2. Theoretical Foundations

2.1. Definition and Historical Evolution of Queuing Theory

A.K. Erlang's seminal 1909 paper, "Probability and the Telephone," is widely regarded as the inaugural publication in the realm of queuing theory [1]. This discipline, a specialized subset of operations research, delves deep into the intricacies of analyzing and modeling queues, aiming to enhance and fine-tune system performance. At its core, queuing theory seeks to pinpoint the ideal system configuration, encompassing factors like server count, system capacity, and the orchestration of arrivals and departures. Such a configuration would ideally curtail wait times, amplify throughput, and elevate overall efficacy.

Several pivotal concepts underpin queuing theory:

Arrival Rate: The frequency of work items entering the system.

Serving Rate: The speed at which tasks are processed by the server.

Queue Discipline: A guideline dictating the sequence in which tasks are addressed from the queue.

System Capacity: The ceiling on the number of tasks the system can accommodate.

Utilization: The proportion of time the server dedicates to task processing.

Little's Law: An elemental equation elucidating the relationship between average wait duration, the count of tasks within the system, and server utilization.

Queuing Model: An abstract, mathematical depiction of a queuing system's dynamics, examples being the M/D/1 and G/G/1 models.

By meticulously dissecting these principles and harnessing mathematical simulations to mirror queuing system behaviors, queuing theory furnishes invaluable perspectives. These insights aid in crafting systems that adeptly cater to client requirements, concurrently slashing lags and bolstering output.

2.2. Other Theories Discussed in This Study

Quality of Experience (QoE) represents the comprehensive assessment of customer satisfaction and interaction with a service or product, along with its respective provider, such as a webpage or television broadcast [2].

3. System Analysis and Empirical Research

3.1. Application of Queuing Theory in Network Communications

3.1.1. The Necessity for Queuing Theory in Contemporary Network Communications. Queuing theory plays a vital role in network communication performance as it provides a mathematical framework for analyzing and modeling the behavior of packets in communication networks. This allows network designers to optimize the network structure, reduce error rates [3], and improve resource utilization [4]. For example, network designers can use flow modeling to represent flow patterns in a network, which helps understand the dynamics of packet arrival and departure. This information is critical for designing efficient protocols and algorithms that can handle traffic efficiently.

3.1.2. Case Studies and Analysis: Queuing Theory in Network Communications

3.1.2.1. Use Queuing Theory to Calculate Appropriate Server Performance to Cope with Network Delays. In the process of network users obtaining services, users are always concerned about the waiting time of services, such as the loading of web pages and videos. For some services, even a pause event of just a few seconds can significantly degrade user-perceived quality of experience (QoE) [5]. Therefore, network designers need to reduce the possible waiting time for users. The process of the server processing requests issued by users can be modeled using queuing theory.

DOI: 10.54254/2755-2721/54/20241227

$$\pi_i = \pi_0 \rho^i$$

$$\pi_i = (1-\rho)\rho^i$$

$$\rho = \frac{\lambda}{\mu}$$

$$E[N] = \sum_{i=1}^{\infty}(i-1)\pi_i = (1-\rho)\rho\sum_{i=1}^{\infty}(i-1)\rho^{i-1} = (1-\rho)\rho\left(\sum_{i=1}^{\infty} i\rho^{i-1} - \sum_{i=1}^{\infty}\rho^{i-1}\right) = (1-\rho)\rho\left(\frac{1}{(1-\rho)^2} - \frac{1}{1-\rho}\right) = \frac{\rho^2}{1-\rho} \quad (1)$$

The waiting time of a job in the queue can be obtained by Little's law (Formula 3)

$$E[T_q] = \frac{E[N]}{\lambda} = \frac{\rho^2}{(1-\rho)\lambda} = \frac{\rho}{1-\rho}\cdot\frac{1}{\mu} \quad (2)$$

The average waiting time in Figure 2 is about 9 times that of Figure 1. This means that the average wait time for users in Figure 2 is ten times longer than that of users in Figure.

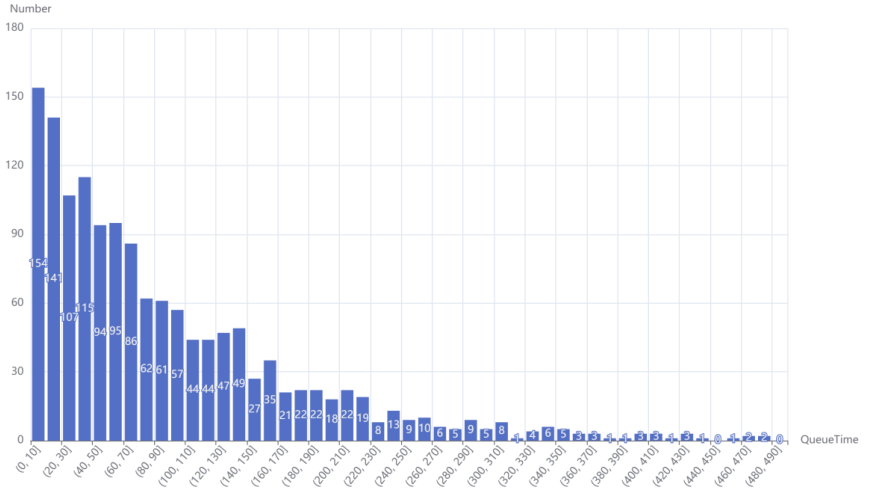

Figure 1. Statistics on waiting time, 3000 jobs are simulated, equipment utilization $\rho=\frac{1}{2}$ (Photo/Picture credit: Original).

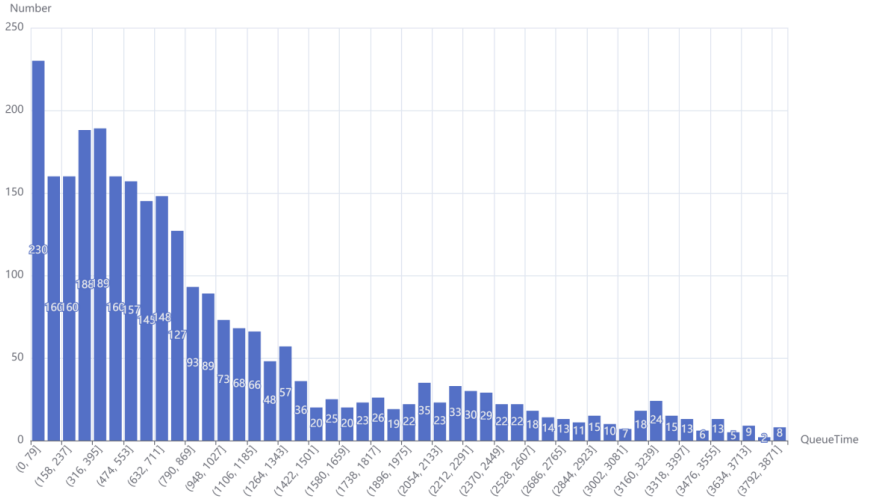

Figure 2. Statistics on waiting time, 3000 jobs are simulated, equipment utilization $\rho=\frac{9}{10}$, the average service rate μ of the server is the same as Figure 1 (Photo/Picture credit: Original).

DOI: 10.54254/2755-2721/54/20241227

When user requests to the network are close to the server's service capabilities, it means that users need to wait longer for their requests, which will greatly affect the user's mood. At the same time, in reality, the storage space of the server is limited. If there are too many jobs in the queue, data will overflow and cause failure [6]. Network designers need to reasonably plan server performance or add new servers based on user response to waiting time and the number of network requests to ensure that the network can run smoothly and stably.

3.1.2.2. Using Queuing Theory to Solve the Problem of Resource Quantity Calculation for Resource Scheduling in 6TiSCH Network. In 2013, the IETF initiated the 6TiSCH standards working group with the intent of integrating IPv6 into the realm of low-power, short-range wireless technology. This initiative aimed to address challenges associated with IP access for industrial field nodes and ensuring the reliability and predictability of communications within resource-constrained environments. The efficacy of the 6TiSCH network is significantly influenced by its scheduling methods. Consequently, academic circles have introduced autonomous resource scheduling algorithms, notably Orchestra and ALICE. The subsequent section will delve into a resource quantity computation solution, termed MY_SF, tailored for resource scheduling within the 6TiSCH network. This solution, rooted in queuing theory, was pioneered by the China University of Posts and Telecommunications [7]. Considering the volume of data packets and spatial dimensions, reference can be made to both formula 4 and formula 5.

$$\alpha_n \begin{cases} \alpha, n = 0, 1, 2, \ldots, L_{q-1} \\ 0, n \geq L_q \end{cases}$$

$$\beta_n \begin{cases} n\beta, 0 \leq n < N_{tx} \\ N_{tx}\beta, N_{tx} \leq n \leq L_q \end{cases} \tag{3}$$

Similarly, the entry and exit of data packets have Markov properties, so formula 6 can be obtained.

$$\begin{cases} \alpha_0 p_0 = \beta_1 p_1, n = 0 \\ (\alpha_n + \beta_n) P_n = \alpha_{n-1} P_{n-1} + \beta_{n+1} P_{n+1}, n > 0 \end{cases} \tag{4}$$

According to Equations 4, 5, and 6, the probability of state n (Formula 7) and the probability of state n when the amount of data inputs and outputs reaches a balanced state (Formula 8) can be obtained.

$$p_n = \frac{\alpha_{n-1}\alpha_{n-2}\ldots\alpha_0}{\beta_n\beta_{n-1}\ldots\beta_1} p_0$$

$$p_n = \begin{cases} \frac{\rho^n}{n} p_0, 0 \leq n < N_{tx} \\ \frac{\rho^n}{N_{tx} N_{tx}^{n-N_{t\alpha}}} p_0, N_{tx} \leq n \leq L_q \end{cases} \tag{5}$$

According to formulas 4, 5, 7, and 8, the probability that there is without any data packet in the buffer queue p_0 can be obtained (formula 9).

$$p_0 = \left(\sum_{n=0}^{N_{tx}-1} \frac{\rho^n}{n!} + \frac{\rho^{N_{ax}}\left(1-\left(\frac{\rho}{N_{tx}}\right)^{L_q - N_{ax}+1}\right)}{N_{tx}!\left(1-\frac{\rho}{N_{tx}}\right)}\right)^{-1} \tag{6}$$

The average cache queue length E (L_{avg}) can be also derived from Formula 8 (Formula 10).

DOI: 10.54254/2755-2721/54/20241227

$$
\begin{aligned}
\mathrm{E}\left(L_{avg}\right) &= \sum_{n=N_{ax}}^{L_q} (n - N_{tx})\, p_n \\
&= \sum_{n=N_{ax}}^{L_q} n\, p_n - N_{tx} \sum_{n=N_{tx}}^{L_q} p_n \\
&= \sum_{n=0}^{L_q} n\, p_n - \sum_{n=0}^{N_{ax}-1} np_n - N_{tx}\left(1 - \sum_{n=0}^{N_{ax}-1} p_n\right) \\
&= E(L_q + N_{tx}) - \sum_{n=0}^{N_{ax}-1} (n - N_{tx})p_n - N_{tx}
\end{aligned}
\tag{7}
$$

The average amount of packets $\mathrm{E}(\mathrm{L_{avg}} + \mathrm{L_q})$ in a node is as follows (Formula 11):

$$
\mathrm{E}\left(\mathrm{L_{avg}} + N_{tx}\right) = \mathrm{E}\left(L_{avg}\right) + N_{tx} + p_0 \sum_{n=0}^{N_{ox}-1} \frac{(n - N_{tx})\rho^n}{n} \tag{8}
$$

The average amount of occupied units $\bar{s}$ is as follows (Formula 12):

$$
\begin{aligned}
\bar{s} &= \mathrm{E}(\mathrm{L_{avg}} + N_{tx}) - \mathrm{E}(L_{avg}) \\
\lambda_e &= \lambda(1 - p_k) \\
W_s &= \frac{\mathrm{E}(\mathrm{L_{avg}} + N_{tx})}{\lambda_e} \\
W_q &= \frac{\mathrm{E}(\mathrm{L_{avg}})}{\lambda_e} = W_s - \frac{1}{\beta}
\end{aligned}
\tag{9}
$$

Through the above formula, network designers can predict many attributes of the system. This attributes can guide network designers to better design network. By modeling queues and analyzing their behavior, network administrators can guarantee a certain level of quality of service for various applications. Simultaneously, queuing theory can be used to determine whether new calls should be accepted into the network. By analyzing the current state of the network and the expected load on the network, queuing models can help network administrators make decisions about call acceptance control to ensure that the network does not become overloaded and provide poor service to existing calls. However, network administrators must also pay attention to the actual situation when applying queuing theory. Network administrators must apply queuing theory correctly based on specific needs.

3.2. Application of Queuing Theory in Public Place Planning

The Imperative of Queuing Theory in Urban Planning. Queuing theory stands as a crucial instrument in urban planning, adept at streamlining and optimizing pedestrian flow across various public domains such as transport hubs, retail centers, medical facilities, and amusement parks. This mathematical approach empowers urban designers to capitalize on spatial efficiency, guiding them toward optimal designs that not only host maximal foot traffic but also circumvent potential overcrowding or lag. Furthermore, queuing theory plays a pivotal role in curtailing wait times by spotlighting congestion points and fine-tuning the operational capacity of structures such as ticket booths, security checks, and attraction lines. Real-world Insights: The Role of Queuing Theory in Urban Spaces. An imperative for railway station architects and administrators is the judicious design of ticket booths to avert undue delays for patrons. Consequently, a nuanced model becomes indispensable for a holistic grasp of the dynamics at play during the conceptualization and governance of such ticket counters. Research conducted at the NSG-3 railway stations over a seven-day span [8] facilitated data accumulation, which upon analysis,

DOI: 10.54254/2755-2721/54/20241227

aligned seamlessly with a single-server queuing paradigm. This indicates that railway design teams and overseers could potentially harness the M/G/1 model as a planning foundation. Nonetheless, the scope of this study was confined to a select few regions in Southern India, necessitating a more expansive on-site evaluation before embracing this model universally. Harmonizing Queuing Theory with Complementary Schemas in Urban Design. Strategies for Incorporating Queuing Theory into Urban Spaces. Prior to deploying any design solution, urban planners should possess a comprehensive understanding of the specific ecosystem they're dealing with. This entails a thorough examination of factors like patron influx rates, service intervals, and overall capacity to pinpoint the primary contributors to gridlocks and lags [9]. Distinct regions, and even varying timeframes within them, will exhibit disparate patterns of arrivals and service distributions. This necessitates a tailored approach where designers actively engage in field assessments to ascertain the most apt model and the associated parameters for their unique situation [10].

4. Conclusion

This article introduces a series of modeling optimization strategies rooted in queuing theory tailored to specific real-world scenarios. While designers can select models that seem to best match their current conditions, it's imperative to recognize that disparities often exist between theoretical models and the intricacies of real-life situations. Thus, it falls upon the designer to refine these models, ensuring they are more congruent with the multifaceted realities they encounter. Additionally, owing to the constraints in the scope of this article, the exploration into the specific applications of queuing theory might not sufficiently address the myriad challenges present in everyday production and living conditions. Such constraints shouldn't be viewed as limitations but rather as starting points. There exists a vast expanse of opportunities for future researchers and scholars to delve deeper into this domain. By further investigating and encapsulating the myriad directions and nuances of queuing theory, they can bring forth more comprehensive solutions and insights that can bridge the gap between theoretical ideals and the ever-evolving dynamics of real-world scenarios.

References

[1] Asmussen, S., & Boxma, O.J. (2009). Editorial introduction. Queueing Syst, 63(1), 1.

[2] Le Callet, P., Möller, S., & Perkis, A. (Eds.). (2013). Qualinet White Paper on Definitions of Quality of Experience. European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Version 1.2.

[3] Prados-Garzon, J., Ameigeiras, P., Ramos-Munoz, J. J., Navarro-Ortiz, J., Andres-Maldonado, P., & Lopez-Soler, J. M. (2021). Performance Modeling of Softwarized Network Services Based on Queuing Theory With Experimental Validation. IEEE Transactions on Mobile Computing, 20(4), 1558-1573.

[4] Zeng, Q., Chen, Z., Jia, Y., & Wang, M. (2021). Dynamic Resource Sharing for Non-preemptive M/M/1/1 Queueing System: An Age of Information Perspective. 2021 IEEE/CIC International Conference on Communications in China (ICCC), 659-663.

[5] Hossfeld, T., Egger, S., Schatz, R., Fiedler, M., Masuch, K., & Lorentzen, C. (2012). Initial delay vs. interruptions: Between the devil and the deep blue sea. 2012 Fourth International Workshop on Quality of Multimedia Experience, 1-6.

[6] Akbarzadeh, O., Hamzehei, S., Amer, A., Fasihihour, N., & Karami, M. (2022). Analyzing the Network System Performance Based on the Queuing Theory Concept. 2022 International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI), 1-6.

[7] Zhang, Y., Guo, J., Cai, Y., & Wu, Y. (2022). Research on Autonomous 6TiSCH Network Resource Demand Calculation Based on Queuing Theory. 2022 5th International Conference on Artificial Intelligence and Big Data (ICAIBD), 583-588.

[8] Nair, A. M., K.S, S., & Ushakumari, P. V. (2021). Application of Queuing Theory to a Railway ticket window. 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM), 154-158.

DOI: 10.54254/2755-2721/54/20241227

[9] Cejka, J., & Šedivý, J. (2021). Discussion of operational transport analysis methods and the practical application of queuing theory to stationary traffic. Transportation Research Procedia, 53, 196-203.

[10] Shortle, J. F., Thompson, J. M., Gross, D., & Harris, C. M. (2018). Fundamentals of queueing theory (Vol. 399). John Wiley & Sons.