Research and Application Analysis of the Basic Theory of Queuing Theory

Yikuan Xiong [1, *]

[1] University of Sheffield, Sheffield, S10 2TN, United Kingdom
*yxiong34@sheffield.ac.uk

Abstract. Queuing theory is a mathematical discipline that studies the behavior of queues or waiting lines. This article aims to provide a comprehensive examination of the core principles of queuing theory and explore its applications in various industries such as transportation, manufacturing, and telecommunications. In transportation, queuing theory helps in optimizing traffic flow, reducing congestion, and improving overall efficiency. By analyzing arrival rates, service times, and queue lengths, transportation planners can make informed decisions regarding capacity expansion, signal timing, and route planning. Queuing theory also plays a crucial role in managing queues at airports, bus terminals, and train stations, ensuring smooth passenger flow and minimizing waiting times. In manufacturing, queuing theory aids in production planning and control. By studying the interaction between machines, workstations, and buffers, manufacturers can optimize production processes, minimize bottlenecks, and reduce lead times. This leads to increased productivity, cost savings, and improved customer satisfaction. In the telecommunications industry, queuing theory helps in network design and performance analysis. By modeling call arrival rates, call durations, and network capacities, telecommunication providers can dimension their networks to handle peak loads efficiently. Queuing theory also assists in determining the optimal number of service channels, improving call routing strategies, and managing customer service levels. While queuing theory offers significant potential for optimization and efficiency improvements, there are also obstacles associated with its implementation. Real-world systems often involve complex interdependencies and stochastic behavior, making it challenging to develop accurate queuing models.

Keywords: Queuing theory, Common queuing models, Stochastic processes

1 Introduction

Queueing scenarios are a common part of our daily lives. It's easy to spot people lining up at banks for transactions, or observe how restaurants handle seating arrangements using queue numbers. Even cars can be seen waiting their turn at petrol stations. These are tangible queuing phenomena, occurring when the demand for services exceeds the speed of delivery. Yet, queues are not just physical. Many are

P. Kar et al. (eds.), *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*, Advances in Computer Science Research 108,
https://doi.org/10.2991/978-94-6463-300-9_46

hidden, such as those resulting from network or computer congestion. As technology progresses and networks grow, user demands in numerous sectors continue to rise. But resources are not infinite, making careful planning and organization crucial.

With the rapid pace of advancements and growth in networks, meeting the diverse and increasing user demands presents a challenge due to resource limitations. Hence, there's a significant need for effective resource allocation and planning. Queueing theory can provide a solution to these problems. It allows for the analysis of complex queues, offering a way to describe their characteristics. This can lead to more efficient resource management and address issues of poor resource use and service delivery delays [1]. All of this works to benefit users, ensuring a smoother experience in a variety of areas.Queueing theory, which is a significant branch of operations research, employs statistical methods to analyze quantifiable metrics (such as waiting time, queue length, and busy cycle length) based on the arrival of service entities and service times. By organizing or reorganizing the service entities in accordance with this statistical methodology, it aims to improve the structure of service systems. By doing this, the system can satisfy customer service needs while attaining the most cost-effective organization or optimal cost for several variables [2]. In the middle of the 1930s, the mathematical community started to take queueing theory seriously after W. Feller added birth-death processes to the theory of random service systems. Since then, as research has advanced, queueing theory has been widely applied in a variety of industries, including logistics scheduling, urban traffic management, computer storage, banking services, and transportation management in the sea, land, and air domains [3]. Nowadays, the theoretical knowledge and application of queueing theory have become highly mature, and there is a scarcity of research gaps in its theory. However, summarizing the applications of queueing theory in various fields still holds significant value. Therefore, this article aims to systematically review and analyze the applications of queueing theory in several common domains, as well as explore optimization strategies in these domains. The goal is to provide readers with a comprehensive understanding of the practical significance of queueing theory.

This paper provides a literature review of the applications of queueing theory. The second section provides a brief introduction to the fundamental theory of queueing. The third section summarizes the applications of queueing theory in the fields of transportation, manufacturing, and communication. The fourth section analyzes the prospects and challenges of queueing theory in practical applications. Finally, the article concludes with a summary of the entire review.

2 The Basic Theory of Queuing Theory

2.1 Stochastic Processes

Queueing theory heavily relies on two primary stochastic processes: the Markov process and the Poisson process. The Markov process is distinguished by its key property, stating that the future state relies solely on the present state and is independent of any past states. Essentially, it's a sequence of states linked by transition probabilities, with each state transition being probabilistic and dependent only on the current state. This predictive ability enables us to forecast the subsequent state based on the current one and its transition probabilities. Markov processes are frequently employed in queueing theory to dissect the behavior of queueing systems, providing insights into steady-state probability distribution, system stability, and overall efficiency. On the other hand, the Poisson process is utilized to represent situations where events occur independently and at a steady rate. Its defining features include discontinuous occurrences over time and the mutual independence of the number of events within a given time span. Within queueing theory, the Poisson process is regularly used to model arrival processes. This aids in analyzing and improving the performance and efficiency of queueing systems.s.

2.2 Queuing Models

In queueing theory, a queueing model is often represented using the notation X/Y/Z/A/B/C. Here, X symbolizes the arrival process, Y represents the service time distribution, Z is the total number of servers, A indicates the capacity of the queueing space, B is the count of objects being served, and C stands for the service rule [4]. This notation was formulated by British mathematician D.G. Kendall, who employed the concept of embedded Markov chains, based on American research, to classify and elucidate queueing systems.

Generally, only the first three variables are considered. As per K. Q. Zhang et al., exponential, Poisson, and Erlang distributions are commonly utilized in queueing theory, potentially serving as parameters for X and Y. Frequently employed queueing models include the M/M/1 and M/M/C models [1]. In the M/M/1 queueing model, three assumptions are made: 1) User arrival times follow a Poisson distribution, 2) Service times adhere to an exponential distribution, and 3) There exists only one server. In the M/M/C queueing model, the following assumptions are made: 1) Multiple servers are present, 2) Each server's work is independent with an equal service rate, and 3) If all servers are occupied upon a customer's arrival, the customer joins a queue and waits. This model follows a first-come, first-served discipline. The M/D/1 and M/Ek/1 models are also frequently encountered. In the M/D/1 model, the service time remains constant for a single server queue, analogous to waiting for toll collection at a single queue toll booth. The M/Ek/1 model signifies a single server queue, where the service time follows an Erlang distribution..

2.3 Steady-State and Transient Analysis

In queueing theory, both transient and steady-state analyses are essential for studying queueing systems. Transient analysis focuses on the system's behavior during specific moments or the initial phase, considering factors such as initial conditions and state transitions. It provides insights into the system's initial response, performance during transition phases, and time-varying behavior. In contrast, steady-state analysis examines the system's long-term behavior and stability, assuming the system has reached a steady-state where the state distribution and performance metrics become constant. It helps evaluate average performance measures, such as waiting time and queue length, under stable operating conditions. Both transient and steady-state analyses offer valuable perspectives for understanding and optimizing queueing systems, allowing researchers to gain comprehensive insights into system dynamics and performance characteristics.

3 Application Research

3.1 Applications in Transportation

Security check at airport. Queueing theory can be employed in the context of airport security inspections to optimize the security systems. The analysis conducted by L. Qinghe, L. Shengyi, and Z. Xiansheng focused on the Documental check and Baggage and Body screening subsystems of the security system [5]. An analysis of passenger processing times collected from the security system was performed, and SPSS software was used to analyze the data. Based on the analysis, queueing models were developed for the two subsystems, and the average waiting lengths for the entire system were calculated. The relationship between the average waiting times for the two subsystems was plotted using MATLAB, allowing the identification of any design flaws in the overall system. Once the flaws were identified, improvement proposals for the security system were put forward, and these proposals were further validated using MATLAB.

Additionally, Y. Zhu et al. directly analyzed the number of open security lanes for the security checkpoint [6]. Research data from various time periods at a specific airport were collected to establish a queueing model for the airport's security checkpoint. By calculating a series of system metrics, the optimal number of open security lanes for each period was determined.

Application for bus station. The application of queueing theory can be used to optimize the design of bus stops. L. Wei, K. Chen, and Z. Xie analyzed the characteristics of bus operations and passenger waiting at bay-type bus stops [7]. By introducing the basic parameters of queueing theory models and establishing a queueing model based on state changes, they proposed optimized designs for bay-type bus stops and multi-service desk guidance modes. The feasibility of the proposed solutions was demonstrated through a case study at the Damao Station in Chongqing City.

Y. Meng, T. Li, and L. Wang simulated bus vehicles and intermediate bus stops as a queueing system with multiple-channel service desks [8]. The theoretical calculations of berth numbers and accommodated routes at intermediate bus stops provided a basis for optimizing the design of bus stations. An example analysis was conducted at the Qinggong Market Station in Xi'an City, where the results were used to optimize the design of bus stops and effectively control the queue length of vehicles at the station, thus improving the station's capacity.

Queueing theory can also be applied to analyze the placement of bus stops. H. Zhang analyzed the impact of bus stop locations on the capacity of expressway exits [9]. By considering two scenarios of bus stop placement upstream and downstream of the exit, the intermittent acceptance theory and queueing theory were employed to establish capacity models for the expressway exit in each scenario. The models were validated through simulations, providing a basis for effective traffic design and management measures.

Dredge the traffic flow. W. Luqiao developed a queuing model for signalized intersections and proposed a signal phase timing algorithm based on queuing networks [10]. The algorithm combines vehicle passage rate models and vehicle arrival rates to efficiently compute the optimal signal timing plan that maximizes intersection throughput. Furthermore, a global road network signal queuing model was established, and a coordination control timing algorithm based on queuing networks was proposed. This algorithm adjusts signal durations based on the vehicle arrival rates at each intersection and the service status of each approach, resulting in a signal timing plan that maximizes throughput. These methods provide effective and efficient solutions for traffic flow management in areas experiencing stable population growth and alleviating the increasing pressure on urban road infrastructure.

3.2 Applications in Manufacturing

Queuing analysis of production line. In practical production scenarios, delays in the production process often occur due to various factors such as material shortages and equipment failures, leading to queuing phenomena. The utilization of queuing theory not only helps identify bottlenecks within the production line but also provides insights into optimizing production processes, resource allocation, and quality control. S. Z. Yu, L. Q. Tao, and L. X. Ting specifically focused on the challenge of optimizing decision-making in remanufacturing systems, where the quality of waste or used parts is highly uncertain [11]. Taking the example of remanufacturing the spindle in waste machine tools, they developed a queuing network model based on queuing theory. This model was utilized to study and optimize the production process, and its effectiveness was confirmed through simulation experiments. Their research provides a novel approach to enhance decision-making and optimize remanufacturing systems.

Optimization of inventory management. Queueing theory can be used to study optimization strategies in inventory management. B. Han conducted an analysis on the optimization control of inventory based on stochastic processes [12]. Using queuing theory and storage theory, a cost model for production-inventory systems under the S-policy was established, and an analysis of the optimal S-policy was conducted based on the model. The feasibility of the proposed optimal strategy was validated through case studies.

Y. Qin focused on production companies that provide products and value-added services and studied the performance and inventory optimization problems of production-inventory systems considering service time [13]. Through the process of model construction, analysis and solution, cost function establishment, and strategy optimization, a research methodology based on queuing theory for analyzing production-inventory systems considering service time was developed.

Optimization of production logistics. For manufacturing companies, considerable research has been conducted on resource scheduling, but there has been relatively less research on the configuration of production logistics transportation resources. Queueing theory is a powerful research tool for optimizing the configuration of production logistics resources. An optimization model for the configuration of transportation resources in a logistics system, with the total operational cost of the logistics system as the optimization objective, was established by A. Cheng et al [14]. The model was solved using genetic algorithms, and the effectiveness of this approach was validated through a case study in the chemical industry. Queueing theory can also be applied to the optimization analysis of logistics equipment.

An optimization analysis of inbound unloading equipment was conducted by B. Han [12]. Firstly, a cost model based on queueing theory was established to obtain the cost function, and then the optimal service rate of the equipment was determined using the Newton iteration method. Through case analysis, the relationship between the unloading rate and cost was elucidated, providing operational and management recommendations for decision-makers.

3.3 Applications in Telecommunications

Optimization of data center. The increasing demand for internet services and cloud computing today has led to a rapid increase in the number of servers running in data centers. As a result, energy consumption in data centers has become a significant concern.

Queuing theory can be used to analyze and optimize data center strategies, aiming for high efficiency and low energy consumption. D. Chen conducted an analysis and research on existing energy-saving strategies for cloud data centers [15]. By analyzing the arrival patterns of user request tasks in a cloud environment, a queuing model for cloud resource allocation was established, and the performance of the model was quantitatively analyzed. A task scheduling model for cloud computing systems was designed. Through comparative experiments, the practicality and effectiveness of the energy-saving queuing task allocation strategy were validated.

In another study, a management strategy was proposed by S. Jian et al [16]. This strategy dynamically opens/closes a portion of servers to minimize data center energy consumption while meeting performance requirements based on waiting time. The minimum number of servers to be opened was determined using queuing theory. The feasibility of this strategy was verified through simulation experiments, providing an approach to balance energy consumption and performance in heterogeneous data centers.

Optimization of network performance. In recent years, with the widespread adoption and rapid development of smart mobile devices, the communication demand for mobile services has grown exponentially. Optimizing network performance using limited spectrum resources has become one of the pressing issues.

Queuing theory can be used to evaluate and optimize wireless network performance. J. Fang conducted research on the performance analysis and resource optimization methods for Device-to-Device (D2D) heterogeneous cellular networks based on stochastic geometry and queuing theory [17]. The adaptive control mechanism of transmission rates was analyzed using the M/G/1 queuing model. The spatial temporal traffic and transmission models of D2D heterogeneous networks were obtained, and performance analysis and optimization were carried out for D2D cellular networks with retransmission mechanisms. Additionally, considering the priority of different types of services, a queuing model with a dynamic priority switching strategy was established to analyze network performance based on birth-death processes. Through simulations, the feasibility of optimizing network performance using the retransmission mechanism and dynamic priority switching strategy was demonstrated.

Furthermore, L. Lei analyzed and researched the wireless network communication performance of Code Division Multiple Access (SCMA) and Non-Orthogonal Multiple Access (NOMA) using stochastic geometry theory and queuing theory [18]. A network performance evaluation method was designed, considering the dynamic and time-invariant characteristics of services, in the context of large-scale randomly deployed networks.

4 Application Prospects and Challenges

4.1 Future Application Prospects of Queuing Theory

The potential future applications of queueing theory are wide-ranging and provide ample room for innovation across diverse fields: Artificial Intelligence and Machine Learning: The fusion of queueing theory with AI and machine learning can lead to the development of intelligent queuing systems. By leveraging real-time data and predictive models, these systems can optimize queuing procedures, forecast customer influx, streamline resource allocation, automate scheduling, and deliver personalized services. Such applications can enhance efficiency, reduce waiting times, and improve overall service quality. Blockchain Technology: The integration of queueing theory with blockchain technology could give rise to distributed queuing systems. Blockchain's hallmark features of decentralization, security, and transparency ensure fair and trustworthy queuing procedures. This innovative approach could prove invaluable in scenarios like ticket sales, resource allocation, identity verification, and more. Internet of Things and Edge Computing: Queueing theory can find use in resource scheduling and optimization within IoT devices and edge computing environments. The ability to monitor and analyze real-time data paves the way for intelligent device scheduling and prioritization, promoting efficient resource utilization and service delivery.

These novel applications marry queueing theory with cutting-edge technologies and sectors, providing fresh opportunities to grapple with intricate real-world challenges. As technology continues to advance and queueing theory research progresses, we can anticipate more innovative applications that deliver efficient, intelligent, and sustainable solutions across a variety of industries.

4.2 Challenges Faced by Queuing Theory in Practical Applications

Despite the broad application of queueing theory across diverse fields, it does face challenges, primarily in its analysis process and the objects it attempts to analyze. In the realm of analysis, many queueing theory models are predicated on the assumption of independence, where arrival and service processes are considered independent. In practical scenarios, these processes can be influenced by various factors like interconnected arrivals or correlated service times, which may surpass the framework of queueing theory. Additionally, certain factors often get disregarded for the sake of simplifying simulation calculations. For example, studies employing queueing theory for traffic signal scheduling often neglect the influence of vehicle arrivals and departures at particular locations, such as residential areas or office buildings, on downstream traffic. Such oversights limit the real-world applicability of the queueing model. Furthermore, queueing theory often employs static models for analyzing systems, assuming unchanged parameters and conditions throughout the study. However, queueing systems often face time variations, seasonal demands, or particular events which demand dynamic considerations not fully accommodated by static models.

As for the objects under analysis, queueing theory commonly assumes the system operates under steady-state conditions, with consistent arrival and service rates. However, many real-world systems inherently have instabilities, such as during system startup and shutdown phases or unforeseen events. In these instances, queueing theory models may fall short and necessitate further refinement. Additionally, when handling complex or large-scale systems like massive data centers, cloud computing environments, or expansive networks, traditional queueing theory methods can prove challenging or even impractical.

5 Conclusion

This comprehensive review delves into the core principles of queueing theory and its diverse applications across various domains. It starts with an elucidation of the fundamental tenets of queueing theory, including stochastic processes, typical queueing models, and both transient and steady-state analyses. The focus then shifts towards the systematic analysis of contemporary applications of queueing theory, predominantly within the realms of transportation, production, and communication. Insights drawn from a plethora of case studies demonstrate the immense potential of queueing theory in enhancing decision-making processes and facilitating system optimization. Moreover, the review goes on to explore the future prospects of queueing theory, highlighting its untapped capacities in emerging fields. It further outlines the challenges frequently encountered in the application of queueing theory, discussing the limitations inherent to the research methodology and the subjects examined. These constraints underscore the necessity for innovative approaches in expanding the theory's usability. Despite its extensive deployment in numerous disciplines, the application of queueing theory is not without limitations in its present form. These include difficulties in modeling and predicting real-world scenarios due to inherent complexity and variability. Furthermore, the development of more effective computational tools and methods for managing the increased complexity of problems in queuing theory is a clear requirement. In summary, while queueing theory offers significant benefits, its current limitations underscore the need for ongoing research to further its development and maximize its potential benefits. These limitations and future prospects for research form a critical part of this comprehensive review, thereby contributing to the broader field of queueing theory and its applications.

References

1. Zhang, K. Q., et al. (2021). A Review of Research on Service Resource Availability Based on Queueing Theory. Computer Science, 48(1), 1-10.
2. Zhou, H. N., et al. (2020). A Review of Academic Research on the Application of Queueing Theory in China. China Storage & Transport, 2020(4), 1-10.
3. Qu, W. (2012). Application Research of Queueing Theory in Supermarket Cashier Management Queue Service System. Master's thesis, Harbin Institute of Technology.
4. Jianwen, T. (2019). Study on the Volume Prediction and Queueing Theory Model of TJG Container Intermodal Transportation. Master's thesis, Beijing Jiaotong University.
5. Qinghe, L., Shengyi, L., & Xiansheng, Z. (2018). Optimization of airport security systems based on queueing theory. In Scientific and Technological Innovation (Vol. 2018, No. 6, pp. 1-6). Nanjing, China.
6. Zhu et al. (2019). Optimization of airport security based on queueing theory. In Shandong Youth (Vol. 2019, No. 2, pp. 1-6). Wuhan, China.
7. Wei, L., Chen, K., & Xie, Z. (2018). Dynamic Queuing Theory and Its Optimization for Large Bus Station. Journal of Chongqing Jiaotong University (Natural Sciences), 37(8).
8. Meng, Y., Li, T., & Wang, L. (2018). Optimization Design of Bus Mid-Stop Stations based on Queueing Theory. Science & Technology Information, 16(14).
9. Zhang, H. (2021). Capacity Calculation of Expressway Exit under the Influence of Bus Stop. Science Technology and Engineering, 21(6).
10. Luqiao, W. (2021). Research on Traffic Signal Scheduling Methods Based on Queueing Theory. Master's thesis, Xi'an Technological University, Xi'an, China.

11. Yu, S. Z., Tao, L. Q., & Ting, L. X. (2020). Remanufacturing Production Optimization of Waste or Used Parts Based on Queuing Theory. Modular Machine Tool & Automatic Manufacturing Technique, 2020(11), 1-10.
12. Han, B. (2014). Optimization of Inbound Logistics Equipment for Manufacturing Enterprises Based on Queuing Theory. Master's thesis, Tianjin University.
13. Qin, Y. (2019). Analysis of Production-Inventory Systems Considering Service Time and Optimization of (s, S) Policy. Ph.D. dissertation, Yanshan University.
14. Cheng, A., et al. (2015). Optimal Configuration of Transportation Resources Based on Queuing Theory in a Production Logistics System. Journal of Guangdong University of Technology, 32(4), 1-10.
15. Chen, D. (2020). Energy-Efficient Management Strategy for Cloud Data Centers Based on Queueing Theory. Master's thesis, Chongqing University of Posts and Telecommunications.
16. Jian, S., et al. (2018). A Strategy for Queuing Theory-Based Performance and Energy Management in Heterogeneous Data Centers. Journal of University of Electronic Science and Technology of China, 47(2), 1-10.
17. Fang, J. (2020). Performance Analysis and Optimization of D2D Heterogeneous Cellular Networks Based on Queuing Theory. Ph. D. dissertation, Nanjing University of Posts and Telecommunications.
18. Lei, L. (2019). Performance Study of Wireless Networks Based on Stochastic Geometry and Queuing Theory. Ph.D. dissertation, Xidian University.