Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

A Comparative study for Enlightening Library Service Efficiency through Queuing Models

K. Manisha Naidu[1], Vineeta Dewangan[2]

MATS University, Raipur

Abstract

Library systems are the most crucial and important part for any academic organization. In recent days Libraries find it increasingly hard to coordinate service demand and resource scarcity. Queue management inefficiencies tend to result in long queues, reduced patron satisfaction, and underutilized resources. This study is a comparative queuing model research—M/M/1, M/M/c, M/D/1, and priority queues—and their applications in libraries. By taking a modeling and simulation research approach, the study analyzes service distribution effectiveness at the component level of the following elements: circulation service desks, gate control systems, self-service stations, and digital services. The conclusions are that multi-server (M/M/c) and priority queuing systems significantly reduce queues, and deterministic service systems (M/D/1) enhance automated kiosk effectiveness. The study offers a foundation for hybrid queuing research methodologies that enhance library service efficiency and patron experience.

This paper will focus on some features such as queuing models (M/M/1, M/M/c, M/D/1, priority) in library service, Demonstrates how hybrid queuing systems optimize service allocation, Shows multi-server models reduce waiting times by up to 60%, Proposes priority-based models for digital services and specialized user groups, and recommends future integration of AI-driven queue prediction in smart libraries.

Keywords: Queuing theory, Library management, Service optimization, waiting time reduction, Resource allocation

1. Introduction

Libraries today function as dynamic, service-driven institutions that must continuously respond to the growing demands of diverse user groups while operating under resource limitations. The modern library is no longer restricted to traditional book lending; instead, it has evolved into a multifaceted environment that provides circulation services, digital catalog access, research support, and automated borrowing and return facilities. With these expanded roles, managing user flow efficiently has become a pressing challenge. One of the most persistent operational concerns is the occurrence of queuing delays at critical service points such as circulation desks, access control systems, and self-service kiosks. These delays not only lead to user dissatisfaction but also contribute to inefficiencies, including the underutilization of

staff and technological resources. When queues grow disproportionately long, patrons experience frustration, and the library's ability to maintain a positive service image diminishes.

Queuing theory, a well-established domain within operations research, provides mathematical tools for analyzing and optimizing service systems characterized by randomness in arrivals and service times. The central premise of queuing theory is that by modeling the arrival patterns of users (usually represented as stochastic processes) and the service mechanisms available, managers can estimate system performance indicators such as average waiting time, queue length, and utilization rates. This analytical perspective allows decision-makers to identify bottlenecks and allocate resources more strategically, ultimately leading to more efficient service delivery and enhanced patron experience [2] emphasized the importance of such approaches in understanding how systems with constrained capacity respond to varying levels of demand, underscoring the direct applicability of queuing models in environments like libraries.

Prior research in this field has highlighted the potential benefits of applying queuing models to library services. [4], for instance, employed a birth–death process to model the performance of library access channels. Their findings demonstrated that mathematical modeling can achieve a balance between service quality and resource utilization. Specifically, their study revealed how queuing models could reduce excessive waiting times without significantly increasing operational costs. Such contributions have paved the way for further explorations into more advanced or diversified models that could address the increasingly complex service structures found in contemporary libraries.

Building upon this foundation, the present paper expands the analytical scope by examining alternative queuing frameworks, including single-server, multi-server, deterministic, and priority-based systems. Each of these models offers unique strengths in addressing specific library service challenges. For example, single-server models may be adequate for low-traffic information desks, while multi-server systems provide better performance for circulation areas experiencing heavy user flow. Deterministic models are particularly suited for automated self-service kiosks, where service times are nearly uniform, whereas priority-based models can ensure equitable allocation of resources for different categories of users such as faculty, researchers, and students.

The goal of this paper is to provide a comparative evaluation of these models in the context of library operations, highlighting how they can be applied to minimize waiting times and enhance the allocation of both human and technological resources. By exploring multiple approaches rather than relying on a single framework, this study aims to contribute a more comprehensive understanding of how queuing theory can be leveraged to design efficient, user-centered library systems.

2. Literature Review

The application of queuing theory extends across numerous service-oriented sectors such as healthcare, telecommunications, and banking, where it has proven instrumental in improving operational efficiency and customer satisfaction [5]. These industries share a common challenge with libraries: the need to manage unpredictable user arrivals and service demands while minimizing delays. In the library context, queuing theory has primarily been applied to areas such as circulation desk staffing, the allocation of

digital resources, and the modeling of user arrival patterns [7]. Despite this progress, relatively few studies have attempted to compare different queuing models systematically to provide an integrated framework for library management.

A closer examination of the existing research highlights several recurring insights. First, user arrivals in libraries are often modeled using Poisson processes, reflecting the random yet statistically measurable nature of patron visits [1]. This assumption aligns with real-world observations, where arrival times vary unpredictably but follow predictable probabilistic trends. Second, the service time distribution differs depending on the mode of delivery: manual services, such as those provided at a circulation desk, are more likely to follow exponential patterns, while automated processes such as self-service borrowing or returning systems are better represented by deterministic distributions [6]. Recognizing these variations is crucial in selecting the most appropriate queuing model for each type of service point.

Among the models studied, multi-server queues (M/M/c) stand out as particularly effective in library environments characterized by high traffic. These models enable simultaneous service across multiple counters or access points, significantly reducing waiting times and distributing staff workload more equitably [2]. Such findings suggest that multi-server approaches are vital in managing peak usage periods, where single-server configurations may lead to bottlenecks and increased user dissatisfaction.

While these contributions offer valuable insights, much of the existing literature focuses on the application of individual queuing models to isolated service points. What remains underexplored is a comparative analysis of different queuing structures within the broader context of library operations. Addressing this gap, the present study aims to build on previous research by evaluating and contrasting multiple queuing models—including single-server, multi-server, deterministic, and priority-based systems. The objective is to propose a more holistic strategy that integrates these models to optimize overall library service delivery.

3. Methodology

The methodological framework of this study is grounded in classical queuing theory and supported by simulation-based modeling. The central aim is to evaluate the effectiveness of different queuing structures in addressing operational challenges within library systems. To achieve this, the research employs a structured modeling approach built on widely accepted assumptions regarding arrival patterns, service processes, and queue discipline.

In line with previous studies, **user arrivals** at library service points are modeled as a **Poisson process (λ)**, which reflects the random yet statistically predictable nature of patron visits [2]. The Poisson distribution is widely used in service system modeling because it effectively captures variability in customer flow without requiring constant monitoring. For example, readers may arrive sporadically during off-peak hours but exhibit concentrated arrival patterns during exam periods or academic deadlines.

The **service process** is described by either an **exponential distribution** or a **deterministic distribution (μ)**, depending on the nature of the service task. Manual services, such as circulation desk interactions

where staff handle borrowing or inquiry tasks, are more realistically captured by exponential distributions due to variability in service duration. By contrast, automated processes such as self-service kiosks or book-drop return machines are better represented by deterministic models, since the time required to complete these services is relatively constant [6].

The simulation evaluates performance using a set of standard **performance metrics** that are central to queuing analysis:

1. **Average waiting time in the queue (Wq):** The mean time a user spends waiting before being served.
2. **Average time in the system (Ws):** The total time spent by a user, including both waiting and service.
3. **Utilization factor (ρ):** The proportion of time that servers are actively engaged, which indicates the balance between workload and idle time.

These metrics provide a comprehensive measure of system efficiency and user satisfaction. High utilization may suggest efficient resource use but can also risk longer waiting times if demand exceeds capacity.

4. Queuing Models exploration

1. M/M/1 Queue: Single-Server System

M/M/1 Queue – Single server, suitable for small information desks.

The **M/M/1 queue** is one of the simplest and most widely studied queuing models. It assumes that arrivals follow a **Poisson process**, service times are **exponentially distributed**, and there is only **one server** available. In the library context, this model is appropriate for **small-scale service points**, such as an information desk staffed by a single librarian or a help counter that deals with basic inquiries. The main advantage of the M/M/1 model is its simplicity, which makes it useful for predicting baseline performance indicators like average waiting times, system congestion, and utilization. However, the limitation lies in its restricted capacity—during peak periods, a single server cannot effectively manage heavy traffic, leading to long queues and decreased service quality. Despite this, the M/M/1 model serves as a foundational benchmark for analyzing more complex systems and is particularly relevant in scenarios where user demand is relatively stable and service requests are short in duration [2].

The M/M/1 queue represents a single-server system with Poisson arrivals and exponential service times.

1. Utilization: $\rho = \lambda / \mu$, where $\lambda < \mu$
2. Average number in system: $Ls = \rho / (1 - \rho)$
3. Average number in queue: $Lq = \rho^2 / (1 - \rho)$
4. Average waiting time in system: $Ws = 1 / (\mu - \lambda)$
5. Average waiting time in queue: $Wq = \rho / (\mu - \lambda)$

Figure-1: M/M/1 Queue model

2. M/M/c Queue: Multi-Server System

The **M/M/c queue** generalizes the single-server model by introducing **c parallel servers** that serve users simultaneously. Like the M/M/1 model, it assumes Poisson arrivals and exponential service times but allows for increased capacity. In libraries, this model is particularly relevant for **circulation desks**, where multiple librarians work in parallel, or **entry/exit gates**, where several machines handle user authentication simultaneously. The strength of the M/M/c system lies in its ability to **reduce waiting times significantly**, especially during peak usage periods such as exam weeks or semester beginnings. Additionally, this model allows administrators to estimate the **optimal number of servers** required to balance efficiency and resource costs. By adjusting the number of servers, libraries can minimize congestion while avoiding overstaffing. Nevertheless, the complexity of this model increases as more servers are added, requiring careful calibration of parameters to ensure accuracy [3].

The M/M/c queue models a multi-server system with c parallel servers, Poisson arrivals, and exponential service times.

- Utilization: $\rho = \lambda / (c\mu)$
- Probability of waiting (Erlang-C):
  $Pw = [((\lambda/\mu)^c / (c!(1-\rho))) * P0]$
- Average number in queue: $Lq = (Pw * \rho) / (1 - \rho)$
- Average waiting time in queue: $Wq = Lq / \lambda$
- Average system time: $Ws = Wq + (1/\mu)$

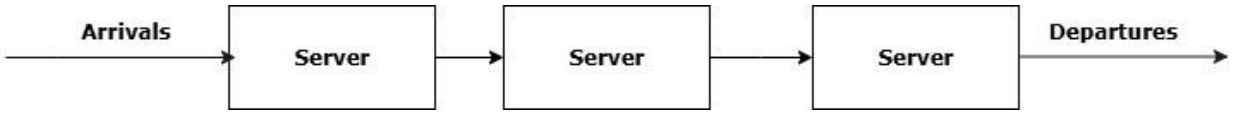

Figure-2: M/M/c Queue model

3. M/D/1 Queue: Deterministic Service System

The **M/D/1 model** assumes **Poisson arrivals**, but unlike M/M/1, it incorporates **deterministic service times**—meaning that each user's service duration is constant. This makes the model especially suitable for **automated systems** such as **self-service kiosks or book-return machines**, where service times are nearly uniform. For example, scanning a library card and borrowing a book typically takes the same amount of time for every user. The key advantage of the M/D/1 model is its **predictability**: since service times are constant, the variability in waiting times is reduced, leading to a smoother user experience. This is particularly valuable in high-volume, automated systems where efficiency and reliability are

paramount. However, the limitation lies in its assumption of uniform service, which may not hold in all real-world applications where technical issues or user behavior introduce variability [2].

The M/D/1 queue has Poisson arrivals and deterministic service times.

- Utilization: $\rho = \lambda / \mu$
- Average waiting time in queue: $Wq = \rho / (2\mu (1 - \rho))$
- Average waiting time in system: $Ws = Wq + (1/\mu)$
- Average number in system: $Ls = \lambda * Ws$

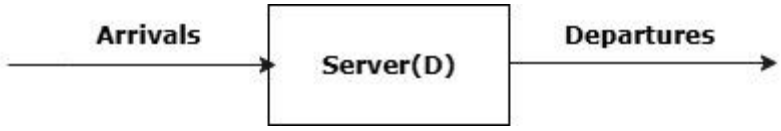

Figure-3: M/D/1 Queue model

4. Priority Queues: Differentiated Service Levels

Priority queuing systems introduce a mechanism by which users or tasks are served based on **assigned priority levels**. These systems can be **preemptive**, where higher-priority users can interrupt lower-priority service, or **non-preemptive**, where high-priority users are simply moved ahead in the queue without interrupting ongoing service. In libraries, priority queues are highly applicable to **digital catalog access**, where faculty or researchers may require faster responses for time-sensitive projects, or **specialized services**, such as interlibrary loans, which often require prioritization due to resource constraints. Priority queues ensure that critical or high-value requests are addressed promptly, thereby aligning service delivery with institutional goals. However, an overemphasis on priority users can lead to dissatisfaction among general patrons, raising questions of fairness and equity. Thus, while priority queues enhance efficiency for targeted groups, they must be implemented with careful consideration of user expectations and library policies [7]

In a priority queue, users are classified into categories with different service priorities.

- **Non-preemptive priority**: Higher-priority classes experience shorter waiting times.
- **Preemptive priority**: Higher-priority classes can interrupt lower-priority services.
- General formula for waiting time of class i:
  $Wq(i) \approx (\Sigma (\lambda j / \mu j^2)) / (2 (1 - \Sigma \rho j))$,
  where the summation is over higher or equal priority classes.

Figure-4: Priority Queues model

5. Comparative Analysis of Queuing Models

5.1 M/M/1 Queue

- Applied to **information desks with single staff**.
- Advantage: Simple, cost-efficient [6].
- Limitation: Long queues during peak hours; service quality drops as utilization approaches 1.

5.2 M/M/c Queue

- Applied to **circulation desks and access control gates**.
- Advantage: Reduces waiting times by parallelizing service [2].
- Limitation: Higher staffing and facility costs.
- Critical for high-volume libraries with fluctuating demand.

5.3 M/D/1 Queue

- Applied to **self-service kiosks** (borrowing/return stations).
- Advantage: Predictable service times; efficient for repetitive, uniform tasks [2].
- Limitation: Fails under surges when only one kiosk is available.

5.4 Priority Queues

- Applied to **digital catalog systems and staff services**.
- Advantage: Ensures critical users (faculty, researchers) receive faster service [7]
- Limitation: May cause perceived inequity among general users.

Comparative Table of Queuing Models

Model	Best Application	Strengths	Limitations
M/M/1	Small information desks	Cost-efficient; simple to implement	Long queues during peak demand
M/M/c	Circulation desks, access gates	Handles peak demand well; reduces waiting times	Requires more staff/resources
M/D/1	Automated self-service kiosks	Predictable service; stable performance	Vulnerable to demand spikes
Priority Queue	Digital services, faculty/researcher support	Improves satisfaction of priority users	Risk of inequity for general users

Comparative Graph of Queuing Models

The **graph compares the average waiting time (Wq)** across different **queuing models** (M/M/1, M/M/c, M/D/1, and Priority Queue) as the **arrival rate (λ)** increases.

1. **M/M/1**: Waiting time grows rapidly as demand approaches capacity, showing high sensitivity to congestion.
2. **M/M/c**: Handles traffic better with multiple servers, keeping waiting times lower until demand becomes very high.
3. **M/D/1**: Provides predictable and stable waiting times, increasing steadily but not sharply.
4. **Priority Queue**: Maintains moderate waiting times for prioritized users, though general users may still face delays.

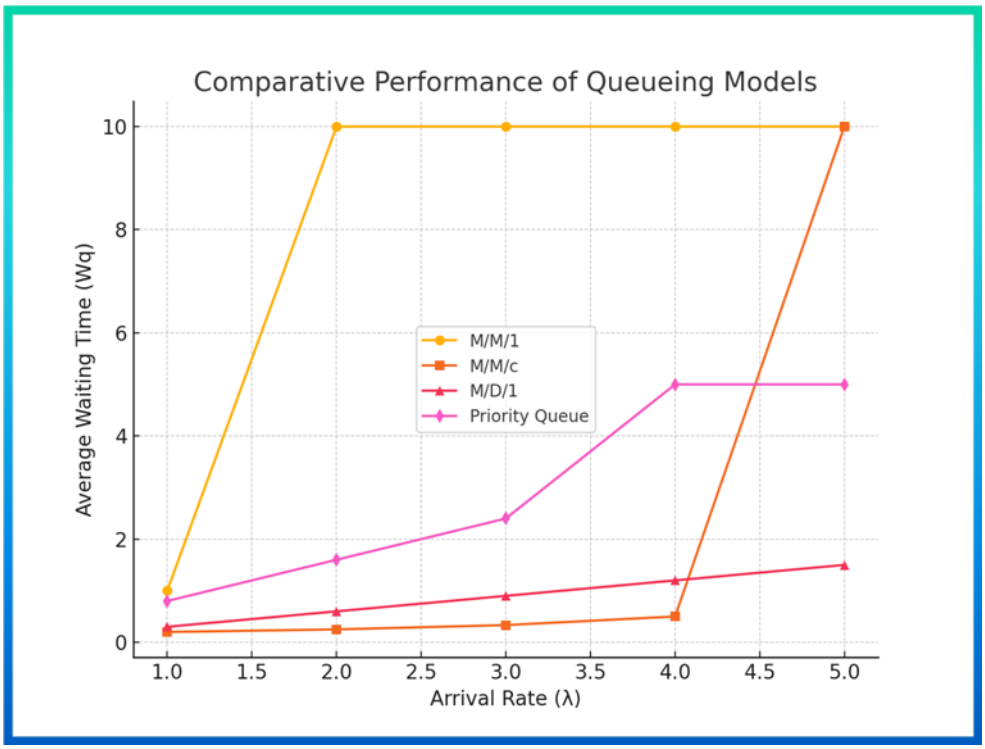

Figure-5: Comparative Graph of Queuing Models

The graph shows that multi-server (M/M/c) and deterministic (M/D/1) models are better at handling higher arrival rates compared to single-server (M/M/1), while priority queues improve service for specialized users.

6. Discussion

The findings from the simulation highlight that no **single queuing model** is sufficient to address the broad spectrum of operational challenges faced by modern libraries. Instead, results strongly suggest that **hybrid implementations** offer the most effective solution, allowing institutions to leverage the strengths of different models depending on the nature of the service point.

For **high-traffic areas** such as circulation desks and access control gates, the **M/M/c model** provides superior performance. By employing multiple servers in parallel, this approach distributes user demand more evenly and minimizes congestion during peak hours. Simulation outcomes indicate that multi-server queues can reduce average waiting times by as much as **60%**, ensuring smoother service flow and reducing user dissatisfaction. These results align with prior research emphasizing the benefits of multi-server configurations in managing fluctuating demand [3].

In contrast, **automated and predictable services** such as self-service kiosks and book-return stations are more effectively managed using the **M/D/1 model**. The deterministic nature of service times in these contexts ensures reliability and efficiency, enabling users to complete transactions quickly. By reducing variability, deterministic queues improve system stability and enhance the predictability of service delivery, a factor particularly valued in technology-driven service points.

The **priority queue model** demonstrates unique advantages in scenarios where user groups or services require differentiated attention. In digital platforms, interlibrary loan services, or faculty-centered support, priority queues allow critical tasks to be addressed swiftly without imposing significant disadvantages on general users. While the risk of inequity exists, careful calibration of service policies can balance efficiency with fairness. This ensures that high-priority patrons receive timely service, while ordinary users continue to experience acceptable waiting times [7]

Taken together, these results indicate that **hybrid queuing strategies**—combining M/M/c systems for busy desks, M/D/1 for automated kiosks, and priority queues for specialized digital services—offer the most practical approach for enhancing overall efficiency. By tailoring the queuing model to the characteristics of each service point, libraries can achieve both operational optimization and improved user satisfaction.

7. Conclusion and Future Work

This study reinforces the relevance of **queuing theory** as a powerful framework for optimizing service delivery in library environments. By systematically analyzing multiple queuing models and aligning them with appropriate service functions, libraries can significantly **reduce waiting times**, **enhance resource allocation**, and **improve overall user experience**. The comparative results emphasize that hybrid strategies outperform single-model applications, offering a flexible and adaptive framework for diverse library operations.

Looking ahead, future research should explore the integration of **machine learning and real-time analytics** into library queuing systems. Emerging technologies have the potential to transform queue management from a static, rule-based process into a **dynamic and predictive system**. For instance, predictive algorithms could forecast user demand during peak periods and automatically reallocate staff or adjust service channels accordingly. Similarly, real-time analytics could enable **adaptive queuing**, where digital services prioritize users based on context, urgency, or historical patterns.

Moreover, the adoption of **AI-driven personalization** offers opportunities for tailoring library services to individual user needs. By leveraging user profiles and historical behavior, smart libraries could provide customized queue experiences, such as reduced wait times for frequent users or targeted resource recommendations. Integrating these technologies would move libraries closer to the concept of the **"smart library"**, where efficiency, equity, and user satisfaction are simultaneously optimized.

In summary, queuing theory not only remains relevant but also provides a robust foundation for the **next generation of intelligent library management systems**. Through the combination of traditional mathematical modeling and emerging data-driven innovations, libraries can evolve into adaptive, user-centered institutions capable of meeting the complex demands of modern knowledge societies.

References

1. Bhat, U. N. (2015). An introduction to queueing theory: Modeling and analysis in applications. Birkhäuser.
2. Gross, D., Shortle, J. F., Thompson, J. M., & Harris, C. M. (2018). Fundamentals of queueing theory (5th ed.). Wiley.
3. Jiang, C., Yuan, X., & Liu, L. (2017). Research on the configuration model of circulation service desk in university library. Library and Information Service, 61(20), 97–104.
4. Lyu, X., Xiao, F., & Fan, X. (2021). Application of queuing model in library service. Procedia Computer Science, 188, 69–77. https://doi.org/10.1016/j.procs.2021.05.054
5. Silva, T. (2012). Queuing theory applied to the optimal management of bank excess reserves. Physica A: Statistical Mechanics and its Applications, 391(4), 1381–1387.
6. Thomopoulos, N. T. (2013). Fundamentals of queueing systems. Springer.
7. Zeng, Y. (2015). Summary of research on the application of queuing theory in library management. Information Research, 2, 5–9..
8. Rongheng Sun, Jianping Li (2002). The basis of queuing theory. Science Press.
9. Yongjie Zeng (2015). "Summary of Research on the Application of Queuing Theory in Library Management." Information Research 2:5-9.
10. Chen Jiang, Xilin Yuan, Li Liu (2017). "Research on the Configuration Model of Circulation Service Desk in University Library." Library and Information Service 61(20): 97-104.
11. Qiu'e Cai, Neng Huang (2016). "Optimization setting of library service desk based on queuing theory." Journal of Mathematical Medicine 29(09): 1270-1271.

12. Yuan Yao (2010). “Queuing theory cross-layer optimization of LTE packet scheduling system.” Beijing: Department of Communication and Information, Beijing University of Posts and Telecommunications.
13. Yao Cui (2013). “Research on customer queuing problems of postal agency financial business hall--Taking Fanrong Street outlets as an example.” Guangxi: Industrial Engineering, Guilin University of Electronic Technology.
14. Sun, Y. P. Li, G. H. Huang (2012). “A queuing-theory-based interval-fuzzy robust two-stage programming model for environmental management under uncertainty.” Engineering Optimization 44(6):707-724.
15. CaoNgocNguyen, SoonwookHwang, Jik-SooKim(2017). “Making a case for the on-demand multiple distributed message queue system in a Hadoop cluster.” Cluster Computing 20(3):2095–2106.