Fuzzy Systems and Data Mining IX
A.J. Tallón-Ballesteros and R. Beltrán-Barba (Eds.)

doi:10.3233/FAIA231060

Numerical Comparison of the Three Poisson Arrival Queuing System Models

Lisha Piao[a, 1], Yingli Zheng[a] and Lizhen Wang[a]
[a] *Dianchi College of Yunnan University, Department of Mathematics, Kunming, Yunnan, China*

Abstract. The optimization of service systems is inseparable from the study of queuing models, involving the establishment of models and the comparative analysis between them. First, the general models of the queuing system are introduced. The queuing system is assumed to be *M/M/1* model and *M/M/c* model, in which the customer arrival is Poisson flow and the service time is negative exponential distribution. As any complex queuing model is derived from these two models, they are widely applied in service systems such as production and daily life. Second, in order to enhance the contrast between the two models, the two kinds of models are refined into three comparable cases, and average queuing time models are established separately. Finally, through theoretical derivation, software simulation, and theoretical testing methods, the comparison results of these three queuing models under the Poisson arrival process are obtained, which is of paramount importance for the practical application of queuing models.

Keywords. queuing theory, *M/M/1* model, *M/M/c* model, average queuing time

1. Introduction

Queuing is a common phenomenon in our daily life. For example, customers go to the store to buy things, patients go to the hospital to see a doctor, and when the sales staff and doctors cannot meet the needs of customers or patients in time, there will be queuing. With the rapid advancement in economic and social development, queuing theory has found extensive applications in fields like transportation systems, storage systems, communication systems, and production management systems [1]. As in the berth design scene, through the queuing system can be a reasonable number of parking spaces in the parking lot design, improve the utilization rate of the parking lot. Lam et al. [2] estimated the queuing time of vehicles and the probability of full parking lot by using the queuing theory. Gan et al. [3] used the single-server queuing system with Poisson distribution of customer arrival, namely *M/M/1* model, simplifies the exit lane of the parking lot, and estimates the queuing time of vehicles in the parking lot and the driving time of the evacuated vehicles. With the rapid development of new energy vehicles, the queuing theory has been widely used in charging station location and facility optimization scenarios, Han et al. [4], Yang et al. [5], Min et al. [6] analyzed the various factors affecting the planning of electric vehicle charging stations and proposed optimization

[1] Corresponding Author, Lisha Piao, Dianchi College of Yunnan University, Kunming 650228, Yunnan, China; E-mail: 22378551@qq.com. This work is supported by the scientific research project of Yunnan University Dianchi College (2022XZD03, 2022XZC12).

principles for station location and capacity allocation. Xie et al. [7] and Wang et al. [8] proposed the use of queuing theory models to optimize the planning of charging facilities under specific constraints, with the goal of minimizing overall costs. In the medical planning scene, Liu et al. [9] and Wu et al. [10] used *M/M/c* model to optimize the outpatient schedule and medical equipment configuration, providing a reference for improving the quality of medical service and work efficiency. In the inventory control scene, Zhang [11] uses *M/M/1* model to study the optimal inventory strategy of perishable goods. In the performance analysis scene, Wang et al. [12] used single and multiple queues to analyze the performance of customers with Poisson arrival and to derive the optimal scheduling method. Therefore, the study of queuing theory is of great significance to the optimal design of service system.

The study of queuing theory primarily involves the establishment of queuing models and the comparative analysis of quantitative indicators across models. The most critical step in model establishment is to derive the steady-state probability formula for the birth and death process [13], and then calculate quantitative indicators for model comparison. To achieve this, the *M/M/1* and *M/M/c* queuing models under the Poisson arrival process are studied, using queuing theory to establish the steady-state probabilities and average queuing times under different arrival rates and service rates. These two models serve as the foundation for theoretical analysis, with any complex queuing models being derived from them. Therefore, the modeling and comparison methods employed in these models can be applied to various application scenarios. To enhance the comparability of the models, the two types of models are divided into three cases. The first case is c standard *M/M/1* queuing systems with arrival and service rates that are $\frac{1}{c}$ of the arrival rate of a standard *M/M/c* queuing system, the second case is a single standard *M/M/1* queuing system, and the third case is a single standard *M/M/c* queuing system. Thus, Case 2 is a reference, Case 1 is a simple duplicate of Case 2, Case 3 is Case 2 when multiple service counters are involved. The quality of the model is primarily assessed by the average queuing time, with the comparison method mainly utilizing a combination of theoretical analysis and software calculation. Specifically, the average queuing time for each model is first derived by queuing theory, leading to the conclusion that the model in Case 2 outperforms that in Case 1. Subsequent to this, the average queuing time for the three cases is simulated using software, leading to the suspicion that the model in Case 3 may be superior. Finally, this hypothesis is verified through theoretical analysis. This method enables efficient comparison of queuing models, providing a reliable reference for the optimization design of service systems.

2. Queuing System Model

2.1. General model of queuing theory

The basic idea of queuing theory is to infer some parameters through the distribution of customer arrival time and service time, including average queuing time, average queue length, the number of service objects in the system, and the probability of system vacant. The distribution of customer arrival time and service time is generally assumed in advance. The common distributions in queuing systems include Poisson distribution, negative exponential distribution, Erlang distribution and so on. When using the Kendall notation *X/Y/Z* to represent a queuing system, *X* represents the time interval of customer arriving one after another, Y represents the distribution of service time, and Z represents the number of service counters. These three elements are the most important and influential characteristic elements of the queuing system. A complete queuing system model includes input process, queuing rules, service counters, and service rules. As shown in Figure 1, the part in the dotted line is the queuing system [13].

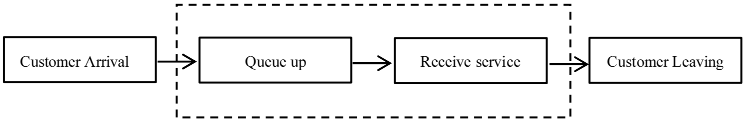


Figure 1. Main components of the queuing system model.

The input process describes how customers arrive at the queuing system according to certain rules. Generally, it can be described from three aspects: the customer population, the customer arrival pattern, and the probability distribution of customer flow. The customer population can be either people or items, and can be a finite or infinite set. The customer arrival pattern describes how customers arrive at the system, either individually or batch arrival. For example, patients visiting a hospital may arrive individually, while materials or products entering the warehouse may arrive in batches.

Customer queuing is divided into unlimited queuing and limited queuing. Unlimited queuing refers to a situation where the number of customers is unlimited and the queue can be of infinite length, also known as waiting-based queuing systems. When customers arrive, if the service counter is currently busy, they will join the waiting queue to wait for service. Limited queuing refers to situations where the number of customers in the system is limited, and is further divided into loss-based queuing systems, waiting-based queuing systems, and hybrid queuing systems.

Service counters can be divided into single service counter and multiple service counters. There are five forms in terms of their composition: a team and a counter, a team and multiple counters, multiple teams and multiple counters, multiple serial formations, and multiple mixed formations. In a waiting-based queuing system, the service counters often use four service models: first-in-first-out (FIFO), last-in-first-out (LIFO), random, and priority services.

2.2. Common models of queuing theory

In queuing theory, *M/M/1* and *M/M/c* are the most common queue systems, and are widely adopted in practical applications. Among them, *M/M/1* is the simplest and most basic standard queuing system model, with the following assumption conditions [13]:

- Input process: (1) The population of customers is infinite; (2) The number of customers entering the system is one at a time; (3) The time interval between customers follows Poisson distribution, with a parameter λ representing the number of customers arriving per unit time.
- Queuing rules: Waiting-based system with unlimited queue length
- Service counters and service rules: (1) single service counter; (2) Customers follow the principle of FIFO; (3) The service time follows a negative exponential distribution with parameter μ , representing the number of customers served per unit time.

The main difference between the standard *M/M/c* queuing system and the *M/M/1* system is that the former has multiple service counters that are independent of each other, with equal service rates. If customers arrive when all service counters are occupied, they will form a queue to wait. The remaining assumptions for the *M/M/c* system are the same as those for the *M/M/1* system.

Next, we will focus our discussion on the standard *M/M/1* and standard *M/M/c* queuing models under the above agreement. Model 1, Model 2, and Model 3 are defined by the juxtaposition of *c M/M/1* queuing systems, one *M/M/1* queuing system, and one *M/M/c* queuing system, respectively. According to the queuing theory, it is assumed that in Model 1, the number of customers arriving in unit time is λ, and the number of customers served in unit time is μ. In Model 2, the parameters are the number of customers arriving in unit time $c\lambda$ and the number of customers served in unit time $c\mu$. And in Model 3, the number of customers arriving in unit time is $c\lambda$, and the number of customers served in unit time is μ. This paper discusses the average queuing time of customers for three models, and compares which queuing model is superior based on this, to provides an important theoretical insights for the queuing systems in practical applications, thereby to better design the queuing system models.

3. Average Queuing Time for Three Queuing Models

3.1. Model 1: c M/M/1 queuing system

Let *c M/M/1* queuing systems be juxtaposed, with $\lambda_1 = \lambda,\ \mu_1 = \mu$ in each system, the queuing system model is shown in Figure 2.

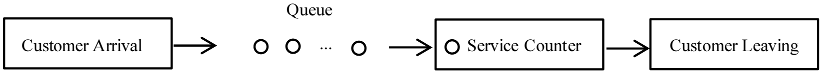
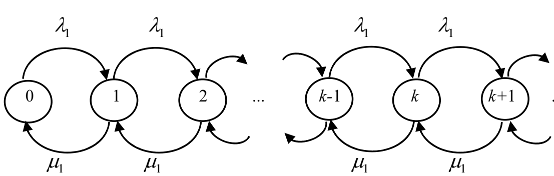


Figure 2. *M/M/1* queuing system of one team and one counter.

According to this queuing system, the state transition process in the infinite state is obtained, as shown in Figure 3.



Where k represents the number of customers in the queuing system. Based on the state transfer diagram, the equations of state balance is listed under steady state probability.

$$\begin{cases} -\lambda_1 p_0 + \mu_1 p_1 = 0 \\ \lambda_1 p_0 - (\lambda_1 + \mu_1) p_1 + \mu_1 p_2 = 0 \\ \quad \cdots \\ \lambda_1 p_{k-1} - (\lambda_1 + \mu_1) p_k + \mu_1 p_{k+1} = 0 \\ \quad \cdots \end{cases} \tag{1}$$

Solve p_1 from the first equation in Eq. (1), substitute it into the second equation to obtain p_2, and so on to obtain p_k. According to [13], $\rho = \frac{\lambda}{\mu}$ is used to represent service intensity, which refers to the ratio of the average service time to the average interval time between customers. It is a measure of system intensity, with a closer ratio indicating a higher service intensity of the system and a busier service organization. According to the finiteness of the flow [13], the steady-state probability of the system represented by $p_k = \rho^k(1-\rho)$. Additionally, the average queuing time of customers can be calculated by dividing the average length of the queue by the rate of customer arrivals per unit time. Therefore, the average queuing time in Model 1 is expressed as

$$W_1 = \frac{\sum_{k=1}^{+\infty}(k-1)\cdot p_k}{\lambda_1} = \frac{1}{\lambda}\left(\sum_{k=1}^{+\infty} kp_k - \sum_{k=1}^{+\infty} p_k\right) = \frac{1}{\lambda}\cdot\frac{\rho^2}{1-\rho} = \frac{\lambda}{\mu(\mu-\lambda)} \tag{2}$$

3.2. Model 2: 1 M/M/1 queuing system

The establishment process of Model 2 is similar to that of Model 1, but assumes that the arrival and departure speeds of customers are $\lambda_2 = c\lambda$ and $\mu_2 = c\mu$, respectively. This leads to a state transfor process that has the same structure as Figure 3, using a derivation method similar to Model 1, the average queuing time for Model 2 is calculated as

$$W_2 = \frac{\lambda}{c\mu(\mu-\lambda)} \tag{3}$$

3.3. Model 3: 1 M/M/c queuing system

The previous two models fall under the single-server scenario. Here, we establish a multi-server queuing model that more closely represents the real-world scenario, and derive the mathematical expression for the average queuing time. First, we make model assumptions, setting the number of servers in the system to *c* and assuming $\lambda_3 = c\lambda$ and $\mu_3 = \mu$. The queuing model is shown in Figure 4.

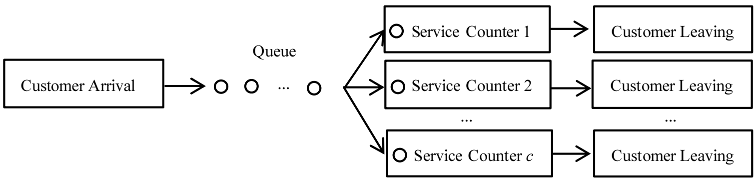
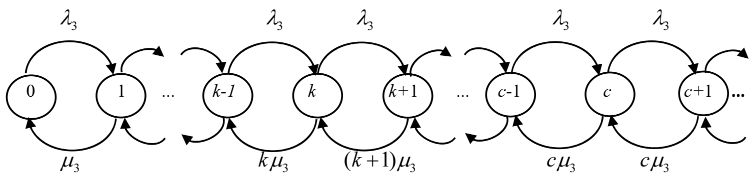


Figure 4. *M/M/c* queuing system of one team and multiple counters.

At this point, the state in the system is also infinite, and the state transition process is as illustrated in Figure 5.



Under the assumption of steady state probability, two situations are considered: when the state is $k = 0,1,...,c-1$, the steady-state probability of the system represented by

$$\begin{cases} -\lambda_3 p_0 + \mu_3 p_1 = 0 \\ \lambda_3 p_0 - (\lambda_3 + \mu_3) p_1 + 2\mu_3 p_2 = 0 \\ \cdots \\ \lambda_3 p_{k-1} - (\lambda_3 + k\mu_3) p_k + (k+1)\mu_3 p_{k+1} = 0 \end{cases} \tag{4}$$

and when the state is $k = c, c+1, ...$, the steady-state probability represented by

$$\lambda_3 p_{k-1} - (\lambda_3 + c\mu_3) p_k + c\mu_3 p_{k+1} = 0 \tag{5}$$

By substituting Eqs. (4) from top to bottom, if $k = 0,1,...,c-1$, then we have $p_k = \frac{1}{k!}\left(\frac{\lambda_3}{\mu_3}\right)^k p_0$. Similarly, if $k = c, c+1, ...$, we can also have $p_k = \frac{1}{c^{k-c} c!}\left(\frac{\lambda_3}{\mu_3}\right)^k p_0$.

Then, the average queuing time for Model 3 represented by

$$W_3 = \frac{\sum_{k=c+1}^{+\infty} (k-c)\cdot p_k}{\lambda_3} = \frac{c^{c-1}}{\lambda c!} \frac{\left(\frac{\lambda}{\mu}\right)^{c+1}}{\left(1-\frac{\lambda}{\mu}\right)^2} \left[\sum_{k=0}^{c-1} \frac{1}{k!}\left(\frac{c\lambda}{\mu}\right)^k + \frac{1}{c!}\frac{1}{1-\frac{\lambda}{\mu}}\left(\frac{c\lambda}{\mu}\right)^c \right]^{-1} \tag{6}$$

So far, the mathematical models of the average queuing time in three kinds of queuing models are derived. The developed average queuing time model can functionalize and quantify the performance metrics of the queuing system, laying a foundation for the comparative analysis of models in the next step. The form of the average queuing time in Model 3 is very complicated, and the numerical method is used for comparative analysis with the help of computer.

4. Numerical Comparisons and Further Proofs

The estimates of the average queuing time of the above three queuing models are compared under different parameters. Assuming that the arrival process of customers is a Poisson flow, the service time follows a negative exponential distribution, and if there are multiple service counters, they are juxtaposed, and a single FIFO queuing model, the customer inflow rate and the service rate are taken as λ and μ respectively in Model 1, $c\lambda$ and $c\mu$ in Model 2, and $c\lambda$ and μ in Model 3. For each model, given the number of service counters, the system service intensity ρ is changed by adjusting the arrival rate of the customers, where $0.1 \leq \rho \leq 0.9$. For each service intensity, the average queuing time of the three queuing models is simulated by *MATLAB*, and Table 1 shows the average queuing time of the three models with different service intensities when the number of service counters is 2, 5, and 10 respectively.

Table 1. Comparison of the average queuing times for the three models, where ρ is the system service intensity, ***c*** is the number of service counters, and W_1, W_2 and W_3 is the average queuing time for Model 1, Model 2 and Model 3 respectively.

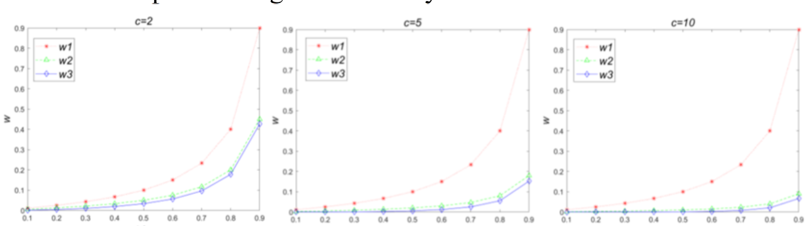
ρ	*c=2*			*c=5*			*c=10*		
	W_1	W_2	W_3	W_1	W_2	W_3	W_1	W_2	W_3
0.1	0.0111	0.0056	0.0010	0.0111	0.0022	3.9e-06	0.0111	0.0011	1.3e-09
0.2	0.0250	0.0125	0.0042	0.0250	0.0050	9.6e-05	0.0250	0.0025	6.0e-07
0.3	0.0429	0.0214	0.0099	0.0429	0.0086	0.0006	0.0429	0.0043	1.7e-05
0.4	0.0667	0.0333	0.0190	0.0667	0.0133	0.0020	0.0667	0.0667	1.5e-04
0.5	0.1000	0.0500	0.0333	0.1000	0.0200	0.0052	0.1000	0.0100	0.0007
0.6	0.1500	0.0750	0.0563	0.1500	0.0300	0.0118	0.1500	0.0150	0.0025
0.7	0.2333	0.1167	0.0961	0.2333	0.0467	0.0252	0.2333	0.0233	0.0074
0.8	0.4000	0.2000	0.1778	0.4000	0.0800	0.0554	0.4000	0.0400	0.0205
0.9	0.9000	0.4500	0.4263	0.9000	0.1800	0.1525	0.9000	0.0900	0.0669

By analyzing Table 1, the relationship of the average queuing time of the three kinds of queuing models satisfies the inequality:

$$W_3 < W_2 < W_1 \tag{7}$$

Where W_i represents the average queuing time of model i (i=1, 2, 3) .

Figure 6 shows the trend of the average queuing time of the three models with the service intensity when the number of servers is 2, 5 and 10. From the graph, the advantage of Model 3 over Model 1 is more significant when the service intensity increases, regardless of the number of service counters, and Model 3 is better than Model 2. The graph conclusion obtained from the experiment has some reference value, but it cannot be used as the final conclusion and needs more rigorous theoretical proof. So far, a comparative framework has been proposed for analyzing and comparing multiple queuing models, which involves theoretical derivation, software simulation, and theoretical verification. This method is rigorous and intuitive, and can effectively provide a reference for the optimal design of service systems.



(a) 2 service counters (b) 5 service counters (c) 10 service counters

Figure 6. The change trend of the average queuing time of the three models.

With the previous numerical experimental analysis, we can boldly guess the conclusion that Model 3 is superior to Model 1 and Model 2. Further theoretical proof is given below. Based on the previous discussion, it can be concluded that Model 2 is superior to Model 1, so the following proofs are mainly focused on the comparison of the average queuing time between Model 2 and Model 3. W_3 in Eq. (6) can be expressed in terms of W_2 as follows:

$$W_3 = \frac{c^{c-1}\rho^{c+1}}{\lambda c!(1-\rho)^2} p_0 = W_2 \cdot \frac{c^{c-1}\rho^{c-1}}{(c-1)!(1-\rho)} p_0 \tag{8}$$

Therefore, it is sufficient to show that $\frac{c^{c-1}\rho^{c-1}}{(c-1)!(1-\rho)} p_0 < 1$ must hold for $W_3 < W_2$.

Prove the inequality $\frac{c^{c-1}\rho^{c-1}}{(c-1)!(1-\rho)} p_0 < 1$ below.

$$\begin{aligned}
\frac{c^{c-1}\rho^{c-1}}{(c-1)!(1-\rho)} p_0 &= \frac{c^{c-1}\rho^{c-1}}{(c-1)!(1-\rho)\left[\sum_{k=0}^{c-1}\frac{1}{k!}(c\rho)^k + \frac{1}{c!}\frac{1}{1-\rho}(c\rho)^c\right]} \\
&= \frac{c^{c-1}\rho^{c-1}}{(c-1)!(1-\rho)\left[\sum_{k=0}^{c-2}\frac{1}{k!}(c\rho)^k + \frac{1}{(c-1)!}(c\rho)^{c-1} + \frac{1}{c!}\frac{1}{1-\rho}(c\rho)^c\right]} \\
&= \frac{(c\rho)^{c-1}}{(c-1)!(1-\rho)\sum_{k=0}^{c-2}\frac{1}{k!}(c\rho)^k + (c\rho)^{c-1}}
\end{aligned}$$

The denominator is always larger than the numerator, that is, $\frac{c^{c-1}\rho^{c-1}}{(c-1)!(1-\rho)} p_0 < 1$, since the number of service counters $c \geq 2$ and the system service intensity $\rho < 1$, the conclusion is proved.

5. Conclusions

Based on the Poisson arrival process, analyze system performance metrics, choose to construct three different queuing systems using the *M/M/1* and *M/M/c* models, and calculate the average queuing time respectively with the steady-state probabilities. For the three types of queuing systems, the parameter values of customer inflow rate and service acceptance rate are preset. A model comparison method is proposed, which involves theoretical derivation, software simulation, and ultimately, theoretical verification. It is found that under a Poisson arrival process, the average queuing time relationship among the three queuing models leads to the conclusion of Eq. (7). The comparison results suggest that “one team and multiple counters” structure is the optimal model of service systems, while a simple copy of "one team and single counter" structure is the most inefficient. The application of this method not only offers a unified framework for analyzing and comparing various queuing models, but also turns system performance indicators into functional and quantitative terms, making the optimal model more precise and intuitive. This facilitates efficient comparison of queuing models and provides a reliable reference for the optimization design of service systems.

References

[1] K. Zhang, Z. Tu, D. Chu, et al. A review of research related to service resource availability based on queuing theory[J]. Computer Science, 2021, 48(01): 26-33.
[2] W. Lam , Z. Li, H. Huang, et al. Modeling time-dependent travel choice problems in road networks with multiple user classes and multiple parking facilities[J]. Transportation Research Part B, 2006, 40(5): 368-395.
[3] J. Gan, X. Mao, J. Zhao. Optimal evacuation model for parking lot vehicles considering dynamic traffic flow[J]. Journal of Southwest Jiaotong University, 2012, 56(01): 123-130.
[4] Y. Han, R. Ren, M. Xu. Optimal allocation of electric vehicle charging facilities at destination charging stations[J]. Operations Research and Management, 2017, 26(08): 76-84.
[5] Y. Yang, P. Gao, C. Zhao. Modeling of temporary charging station siting and capacity setting for electric vehicles in cities[J]. Journal of Chongqing Normal University (Natural Science Edition), 2020, 37(06): 1-6.
[6] D. Min, K. Jiang, R. Liu, et al. A two-stage site selection and capacity determination model for pure electric truck charging station[J]. Journal of Chongqing Polytechnic University (Natural Science Edition), 2023, 37(01): 186-195.
[7] Y. Xie, L. Zhang, S. Deng, et al. Research on Optimizing Network Layout of Electric Vehicle Charging Facilities[J]. Mathematics in Practice and Theory, 2020, 50(10): 168-176.
[8] X. Wang, M. Shahidehpour, C. Jiang, et al. Coordinated planning strategy for electric vehicle charging stations and coupled traffic-electric networks[J]. IEEE Transactions on Power Systems, 2019, 34(1): 268-279.
[9] X. Guo, J. Zhang, W. Yang, et al. Optimization of vaccination clinic resource allocation based on queuing theory model[J]. Modern Preventive Medicine, 2021, 48(17): 3172-3175+3181.
[10] L. Wu, Q. Tan, S. Luo. Assessment and improvement of outpatient pharmacy service quality and efficiency based on queuing theory mathematical model[J]. Anhui Medicine, 2020, 24(08): 1666-1669.
[11] H. Zhang. Perishable M/M/1/N queuing Inventory control system based on (r,Q) strategy[J]. Modern Business Industry, 2019, 40(33): 205-208.
[12] S. Wang, X. Li, L. Chen. A review of random request performance analysis in cloud environments[J]. Journal of Computing, 2022, 45(06): 1241-1260.
[13] Y. Hu. A course in operations research[M]. 4th edition. Beijing: Tsinghua University Press, 2012.