Dean&Francis

ISSN 2959-6157

Benefits Brought by Basic Models of Queuing Theory in Terms of Transportation

Che Huang

International Department of Jinling High School Hexi Campus, Nanjing, 210000, China
Corresponding author: hclz090622@gmail.com

Abstract:

This essay is dedicated to emphasizing the importance of basic queuing models, by analyzing with specific examples in the transported situations. The most significant part is improving the intensity of service and reduce the total cost, helping decision makers make the optimism choices both considers the allocation of resources and service methods. This essay focus on two basic queuing models, M/M/1 and M/M/c queuing models. In the specific example of two queuing models, motorway charging is an example. By calculating the probability of vacancy for both ETC and MTC channels, the total cost was estimated. After the calculation for some specific value of the rate that vehicles having ETC, the percentages of ETC channels can be decided correspond to different rate of possessing ETC. In the example of M/M/c queuing model, the parking gate machine is exemplified. The author compares the total cost for two situations with $c = 2$ and $c = 3$. The result of comparison shows the better options. However, those applications are for the preliminary step. The decision makers should improve their own models based on the realistic conditions to be more accurate and closer to the reality, thereby improving the feasibility.

Keywords: Queuing theory; Transportation; M/M/1 queuing models; M/M/c queuing models.

1. Introduction

Due to the limit of the resources in different field such as the parking, which means not every vehicle can be served at the same time, the phenomenon of queuing can be seen in many situations. There are several schemes to minimize the issue of queuing. However, it is difficult to select an optimism option. Queuing models help to estimate a rough value to help server to make a plan to reduce the queuing time, thereby maximizing the profit for the system. The estimation of basic queuing model maybe far away from the real value, but it gives a preliminary step to calculate the information about the service system. User can vary the data further depends on the actual situations. For example, increase in a certain percentage in number of cars at peak hours. This improves the accuracy of the data.

Many professors have already done many research in their essay. Ruiyin Xu developed a queue length variation model under traffic incidents [1]. Doctor Hu proposed a grid-based road network division method using spectral clustering and applied queuing theory to determine the optimal number of traffic

police required per grid, reducing police allocation and response time [2]. Doctor Zhang used queuing theory with a multi-server limited-capacity model to determine the optimal number of car elevators in underground cylindrical garages [3]. Doctor Pan developed M/D/C and M/G/K queuing models with engineering approximation formulas to identify real-time congestion levels [4]. Doctor Liu built an M/M/c queuing model for a gas station, improving efficiency by 11.45% [5]. Doctor Li applied an M/M/c/N queuing model to analyze charging station congestion, increasing the number of charging piles from 5 to 7 significantly reduces waiting time [6].
This essay aims to show two types of queuing model. In section 2, assumptions about the queuing models are made, and it also introduces two queuing models. Formulas of queuing models will be illustrated, and some limitations and corresponding solutions will be stated. The section 3 focuses the application of two models to make it useful in reality. The comparison will be made and the best option will be selected after making estimations, which proofs that queuing models are useful in minimizing the costs. There are some innovative points in this essay. Firstly, due to the introduction of both M/M/1 and M/M/c queuing model, it demonstrates the adaptability and decision capabilities of the same theoretical framework in different practical problems. Secondly, this essay points out the gap between the reality and the model assumptions such as Poisson's arrival, infinite waiting space, and independence, and proposes simple and practical correction strategies. Lastly, the total costs are classified to waiting cost, operational cost and vacant cost, which is comprehensively calculate the cost brought by queuing and not.

2. Queuing model

What queuing theory research do is to predict the time of queuing and the number of the waiting vehicles, so that servers can make the choice to maximize the utility and profits with the fixed time. A whole system of the service contains four processes: arrival, waiting, being served and departure. To evaluate the system, the rate of the utility and the average waiting time can be used to judge it.

2.1 Aassumptions of the model

At the very beginning of the construction model, some meaning of parameters should be stated. Let λ be the average arrival rate, μ be the average service rate. Both λ and μ have the unit vehicles per hour and several assumptions should be made:
Assumption I: The probability distribution of the number of arriving vehicles is constant within any time interval. The probability distribution of the service time does not change with the time.
Assumption II: The number of vehicles which wait for serving, the capacity of the waiting line and the serving resources are unlimited.
Assumption III: The service times and arrival times of vehicles are independent between vehicles.

2.2 M/M/1 queuing model

M/M/1 queuing model is used to describe the system with single service station, which is the meaning of the '1'. For the two 'M', it indicates that the time intervals between the arrivals of vehicles obey the Poisson process and the service time obey the negative exponential distribution. Furthermore, the capacity and the service targets are unlimited and the model's discipline follows by first-come, first-served, that is, vehicles are served by the order of their arrival time. Several formulas will be provided below.
The intensity of the service, denoted by ρ, is calculated as the ratio of the average arrival rate λ to the average service rate μ, i.e., $\rho = \lambda / \mu$. The probability that no vehicles are present in the service station, P_0, is given by $P_0 = 1 - \rho$. The average number of vehicles waiting in the queuing line, L_q, can be expressed as

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\rho^2}{(1-\rho)} \tag{1}$$

where λ is the average arrival rate, μ is the average service rate and $\rho = \lambda / \mu$, which is the intensity of the service. Finally, the average waiting time of vehicles, W_q, is computed as $W_q = \frac{\lambda}{\mu(\mu - \lambda)}$ [7].

2.3 M/M/c queuing model

M/M/c queuing model is based on M/M/1 queuing model. The difference between 'c' and '1' is that M/M/c queuing model is used to estimate the system with multiple service station, which its number is c. c can be any positive integer not less than two. But still follows by first-come, first served. The conditions of M/M/1 queuing model apart from single service station are still possessed by M/M/c queuing model. There are two forms of system. One is single queuing line, the other is multiple queuing line. The former means all vehicles stay in one line to wait for many services station. The latter refer to every service station has one corresponding queuing line. Vehicles cannot switch between lines while waiting. That can be equivalent to the parallel combination of multiple M/M/1 queuing model.

ISSN 2959-6157

For the M/M/c queuing model with multiple waiting line, the utility of the system, which also called the intensity of the service system, can be calculated by the formula $\rho = \lambda / c\mu$. The average waiting time:

$$W_q = \frac{\lambda / c}{\mu(\mu - \lambda / c)} \quad (2)$$

and the average length of queuing line should multiply by c, which is:

$$L_q = \frac{\lambda^2 / c}{\mu(\mu - \lambda / c)} \quad (3)$$

For the model with single waiting line, the formulas are much more complicated. The probability of the system which is vacant can be calculated by:

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)} \right]^{-1} \quad (4)$$

This value is necessary in other calculations of the prediction. The formula of the average waiting time of vehicles is:

$$W_q = \frac{P_0 \cdot (c\rho)^c \cdot \rho}{c!(1-\rho)^2 \lambda} \quad (5)$$

and that of the average length of the queuing line is:

$$L_q = \frac{P_0 \cdot (c\rho)^c \cdot \rho}{c!(1-\rho)^2} = W_q \cdot \lambda \quad (6)$$

which is known as Little's Law [8]. Compare to the multiple waiting line, this form is better due to the smaller average length of the queuing line and the average waiting time. Because multiple waiting lines cannot be in the same length every second. Some of the waiting line can be shorter, but vehicles cannot switch to that line. This illustrates the unfairness. So, the probability of service stations to be vacant reduces, being most efficient and flexible. However, in the reality, the form with multiple waiting lines is more feasible, such as the form in highway toll station.

2.4 Limitations of the queuing model and the solutions

For the assumption I, hardly can that be satisfied in the reality. Generally, in the peak hour, the number of vehicles who need to be served is much more than other times. So, the number of vehicles may vary with time. Also, the service time varies with different vehicles, drivers and different service station. Some sudden situation will affect the unstable state. The transient behaviors need further analysis.

As for the assumption II, hardly can that be achieved. As the space is limited such as the car park, not every vehicle has a place to wait. Moreover, when drivers find the waiting line is too long to wait, balking and reneging, which means leave when just reach to the system and waiting in the line respectively, occurs between vehicles preparing to wait or waiting in the line. Those behaviors contradict the estimation from the model. As a result, values calculated by the formula may not be accurate.

In terms of the assumption III, not every behavior from vehicles can affect other vehicles. Take the motorcade as an example. Many vehicles travel in a group and arrive to the service system at the same time. Consequently, the Poisson process no longer holds.

To resist the limitations, several variations can be made. Dividing a day into many time periods according to the vehicles flow to calculate can solve the problem of unstable state. The author also can predict whether the place will be filled and causing inaccurate data. If not, the model can still be used. To correspond the issue of groups of vehicles, people can increase the estimate arrival rate by a certain percentage on the basis of the proportion of motorcade. For example, increase the data 5% if approximately 10% vehicles of total are in groups. The prediction can be more accurate.

3. Application of the queuing model

By using the appropriate models, one can estimate the approximate cost, which aim to maximize the profit. In the situations relative to vehicles, the total cost of charging station can be classified in three forms, operational costs, queuing costs and vacancy costs.

3.1 Application in M/M/1 queuing model

In Doctor Liu's essay, the application of M/M/1 queuing model in motorway charging was researched. Assume that cars with ETC all go by ETC channels, and cars without ETC choose the MTC entrances. As cars are physically separated, which are forced to enter the corresponding channel, this situation can be simplified in M/M/1 queuing model [9]. E is for ETC and M is for MTC, by the formula (1), one can calculate the probability of being vacant for two types of channels $P_E = (1-\beta)(1-\rho_E)$ and $P_M = \beta(1-\rho_M)$, where β represents the probability for a car possessing ETC. The losing profit caused by vacancy Y_1 (dollar per hour) can be calculated by:

$$Y_1 = P_E \bullet C_{FE} \bullet C_E + P_M \bullet C_{fM} \bullet C_M \quad (7)$$

where C_{fi} represent the cost of one vacant entrance and C_i represents the number of entrances, $i \in (AM)$, both have the unit dollar per hour. By the formula (2), the average number of cars waiting in one MTC entrance N_M is:

$$N_M = (1-\beta)^2 \lambda^2 / C_M \mu_M^2 \cdot \left(1 - \frac{(1-\beta)\lambda}{C_M \cdot \mu_M}\right)^{-1} \quad (8)$$

and that in one ETC entrance N_E is:

$$N_A = \beta^2 \lambda^2 / C_E \mu_E^2 \cdot \left(1 - \frac{\beta\lambda}{C_E \cdot \mu_E}\right)^{-1} \quad (9)$$

The losing profit caused by queuing Y_2 (dollar per hour) can be calculated by:

$$Y_2 = N_E \cdot C_{WE} \cdot C_E + N_M \cdot C_{WM} \cdot C_M \quad (10)$$

where C_{wi} represents the waiting cost of one entrance. Eventually, the operational cost, can be calculated by:

$$Y_3 = a_E C_E + a_M C_M \quad (11)$$

where a_i refers to the operational cost in one entrance.

Combining the value in the above equations, the total cost in charging station on motorway $Y_{total} = Y_1 + Y_2 + Y_3$. To maximize the profit, the total cost Y should be minimized. After using the model to estimate, it is much easier to determine how to minimize the cost by controlling the number of MTC and ETC channels. Which is the most affected factor. In the essay, the author estimated the trend between β and the proportion of ETC channel by setting particular values\[\]β. The graph which proportion of ETC channel against with β was plotted. The graph shows that the proportion of ETC channels increases with the increase of β. And when β equals to 0.6, the total cost is minimized when 40% are ETC channels. To cope with different value of β, some channel can be changeable, which can switch between ETC channel and MTC channel. As a result, the total costs can be simplified by using the M/M/1 queuing model, making the costs more measurable [10].

Table 1. Illustration of parameters and their values.

parameters	values
λ	180 cars/h
μ	120 cars/h
c	2,3 gate machines
C_f (the cost of vacancy)	6 dollar per hour
C_w (the cost of waiting)	20 dollar per hour
C_o (operational cost)	10 dollar per hour

3.2 Application in M/M/c queuing model

In this part, the author will set values to parameters and estimate the costs in parking lot by using M/M/c queuing model, see Table 1. It adopts the queuing method of single queuing line.

There are two conditions: one is two gate machines and another is three gate machines. Next, people can calculate the basic information about the parking lot gate machines by M/M/c model. And estimate the total costs for both conditions. Eventually, one can compare between 2 cases to determine the one with minimized cost.

The intensity ρ of the service station is given by Eq. (3). When $c=2$, it is calculated $\rho = \frac{\lambda}{c\mu} = \frac{180}{2\times120} = \frac{180}{240} = 0.75$, while when $c=3$, it is calculated $\rho = \frac{\lambda}{c\mu} = \frac{180}{3\times120} = \frac{180}{360} = \frac{1}{2} = 0.5$.

The probability P_0 can be calculated by using Eq (4). When $c=2$, it is calculated

$$P_{02} = \left[(1+1.5) + \frac{2.25}{2\times0.25}\right]^{-1} = \frac{1}{7} = 0.1429.$$

While when $c=3$, it is calculated

$$P_{03} = \left[(1+1.5+1.125) + \frac{3.375}{6\times0.5}\right]^{-1} = \frac{1}{4.75} = 0.2105.$$

The average waiting time W_q is governed by Eq. (5). When $c=2$, it is calculated

$$W_{q2} = \frac{0.1429\times2.25\times0.75}{2\times0.0625\times180} = \frac{0.2411}{22.5} = 0.01072h = 38.59s.$$

In contrast, when $c=3$, it is calculated

$$W_{q3} = \frac{0.2105\times3.375\times0.5}{6\times0.25\times180} = \frac{0.3552}{270} = 0.001316h = 4.74s.$$

The average number of cars L_q in queuing line is given by Eq. (6). When $c=2$, it is calculated $L_{q2} = 180\times0.01072 = 1.93$ cars. However, when $c=3$, it is calculated $L_{q3} = 180\times0.001316 = 0.24$ cars.

ISSN 2959-6157

For the cost of vacancy Y_a, it is calculated that $Y_{a2} = c \times P_0 \times C_f = 2 \times \frac{1}{7} \times 6 = \frac{12}{7} \approx 1.71 dollar/h$ and $Y_{a3} = c \times P_0 \times C_f = 3 \times \frac{4}{19} \times 6 \approx 3.79 dollar/\text{h}$ for $c = 2$ and 3, respectively.

For the cost of waiting Y_b, it is calculated that $Y_{b2} = \lambda \times W_q \times C_w = 180 \times \frac{3}{280} \times 20 = \frac{270}{7} \approx 38.57 dollar/h$ and $Y_{b3} = \lambda \times W_q \times C_w = 180 \times \frac{1}{760} \times 20 = \frac{90}{19} \approx 4.73 dollar/h$ for $c = 2$ and 3, respectively. In addition, for the operational cost Y_c, when $c = 2$, it is calculated $Y_{c2} = c \times C_o = 2 \times 10 = 20 dollar/h$. When $c = 3$, it is calculated $Y_{c3} = c \times C_o = 3 \times 10 = 30$ dollar/h.

By combining these equations, one can estimate the total cost in this situation. Specifically, when $c = 2$, it is calculated

$$Y_{total2} = Y_{a2} + Y_{b2} + Y_{c2} = 20 + \frac{270}{7} + \frac{12}{7} = \frac{422}{7} \approx 60.3 dollar/h \quad (12)$$

In contrast, when $c = 3$, it is calculated

$$Y_{total3} = Y_{a2} + Y_{b2} + Y_{c3} = 30 + \frac{90}{19} + \frac{72}{19} = \frac{732}{19} \approx 38.5 dollar/h \quad (13)$$

As a result, the cost of case with $c = 3$ is 21.8 dollar/h lower than that in case with $c = 2$. Therefore, in the long run, $c = 3$ is better option compared to $c = 2$. By using M/M/1 and M/M/c queuing models, the total cost can be estimated and seek out which option is optimum. Queuing models do help in people's reality.

4. Conclusion

To summarize, this essay illustrates the formulas, assumptions, and limitations of two basic queuing models-M/M/1 and M/M/c-within transportation systems. Key assumptions include Poisson arrivals, exponential service times, unlimited capacity, and vehicle independence. To address real-world limitations such as time-varying demand, limited space, and platoon arrivals, solutions like time-period division and percentage adjustments are proposed. Using motorway toll collection (M/M/1) and parking gate machines (M/M/c) as examples, the essay demonstrates how total costs, including operational, vacancy, and queuing costs, can be estimated and minimized. In the parking case, increasing service stations from two to three reduces total cost by approximately 21.8 dollars per hour, proving that queuing models support data-driven decisions. These models provide a preliminary yet valuable step for transportation estimations. With future internet advancements, real-time data will enable more accurate algorithms. As car ownership continues to rise, queuing theory will play an increasingly important role in optimizing resource allocation and reducing congestion. Decision-makers should refine these basic models based on local conditions to achieve more feasible and cost-effective solutions.

However, limitations are exposed by this study. First of all, all numerical calculations are based on assumption rather than the data in the reality. The given value to the parameters is hypothetical. So, the adaptability of the model in specific real location is limited. Secondly, the analysis only considers stable state conditions. There are inadequate considerations of transient behaviors such as sudden surge in arrival rates or temporary service interruptions. Lastly, the study only examines two basic queuing models, M/M/1 and M/M/c models. But other variants such as models with finite queue capacity are not elaborated. Future research should address these limitations by using empirical data, conducting sensitivity analyses, and extending the models to more complex traffic situations.

References

[1] Xu R. Analysis method of urban road traffic capacity based on queuing theory. China Science and Technology Information, 2019(2): 93-95.

[2] Hu Z, Zhou J, Guo X, et al. Research on the allocation of road traffic police resources based on queuing theory. Chinese Journal of Safety Science, 2024, 34(7): 178-185.

[3] Zhang M, Zhang Z, Du X. Research on the number of car elevators in underground cylindrical garages based on queuing theory. China Journal of Highway and Transport, 2017, 30(4): 133-139.

[4] Pan H, Chen L, Chen W, et al. A real-time traffic condition recognition method for toll stations based on queuing theory. Transport Manager World, 2022(1): 62-64.

[5] Liu G, Cui Y, Wang Z. Research on peak hour operation optimization of gas stations based on queuing theory. Economist, 2021(1): 283-284.

[6] Li J, Zhang S, Zhang R, et al. Research on the scale of new energy electric vehicle charging stations based on queuing theory. Storage and Transportation in China, 2025(7): 158-159.

[7] Jiang Y. Output process of two-dimensional M/M/1 queuing model. Changsha: Central South University, 2023.

[8] Sun J, Ding R, Chen Y. Comparative simulation of M/M/c type and M/M/1 type queuing systems. Journal of Beijing University of Technology, 2016, 42(9): 1324-1331.

[9] Wang K, Liu X, Wang S, et al. Dynamic network load queuing model for the entire area of expressway ramp toll stations. Journal of Beijing Jiaotong University, 2024, 48(1): 167-175.

[10] Liu C. Research on lane configuration of expressway toll collection system based on queuing theory. Shandong Communications Science and Technology, 2020(2): 115-117.