© 2018 IJSRST | Volume 5 | Issue 3 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Mathematical Modelling and Automation of Real-time Queues

Srikrishna C.N
Department of Information Science and EngineeringPES Institute of Technology – South Campus,Bangalore, Karnataka, India

ABSTRACT

The traditional Queue System followed today is quite certainly failing in managing huge crowds, especially in urban populated areas. There are even instances where the customers (users) have to wait for hours to get simple tasks done. It is important to manage these systems effectively by the Organizations in order to improve customer experience. While the Token Queue System provides a fairly acceptable solution, it has many problems of its own. In a token system, waiting physically in the building premises is inevitable. The design discussed in this paper holds the prime idea of completely automating the Token System, where users' waiting time is almost negligible. Using the classical Queue Theory and Poisson's Probability Distribution, it is possible to predict the arrival of the users and the average service time, for a particular service. It uses minimal resources and is cost effective. This paper discusses a model that uses these ideas to build a system that automates the traditional queuing system. It aims not only to minimize the customer wait time, but also reduce the crowds at the service area.

Keywords: Queues, Queue theory, Token System, Automation, Smart phones.

I. INTRODUCTION

Arnold O. Allen, in his famous book - 'Probability, Statistics, and Queue Theory'[1], describes "Waiting in line for service is one of the most unpleasant experiences of life on this planet." It is indeed a dreadful experience which most of the people go through. For an organization, higher the unhappy customers, the more is their loss of reputation.

Ever since the rise of urbanization and population growth, managing huge crowds has become a serious challenge. This certainly raises an interesting question, as how to solve this problem using Automation. Although there are many models that have attempted to solve this issue, only a few of them have been proven effective [5][6][7]. This is the reason why traditional Token System seems to be a better alternative and is still popular. On the other hand, the growth of automation has changed the way the traditional things would otherwise work. Thus, the proposed model in this paper has been designed to promote automation in every aspect. The model assumes the queue dealt with, is a structured queue.

It is designed in such a way that it is cost effective as well as effective in managing the real time queue systems.

A brief walk through of the paper is given below,

- ✓ In section 2, (**Literature Review**), a brief review is discussed on the classic Queue Theory, Models of Queue and Probability Distributions that are used in the design.
- ✓ In section 3, (**Design and Implementation**), the proposal of the model as well as the design and implementation is discussed, in detail.
- ✓ In section 4, (**Results and Discussion**), test cases and results of the model are discussed. A brief discussion on emergency and error handling is done.

- ✓ In section 5 & 6, (**Conclusion and Future Work**) conclusion of the model and future work to improve rapid variations, are discussed.

II. LITERATURE REVIEW

A queue is a group of people waiting for a service one behind the other. The following gives a broad classification of queues

A. Types of Queues

1. Structured Queue: In this type of queue, people stand at a predictable position.
   E.g.: Queues at banks and supermarkets have a predictable number of people forming a queue.
2. Unstructured Queue: In this type of queues, people are considered to be located at unpredictable positions and timings
3. Virtual Queue: This is a new form of Queue system, where people use their personal devices such as smartphones and tablets to reserve their turn as well as be updated as their turn approaches.

B. Queue Theory

Queue theory is the mathematical study of behavior of queues. It was extensively researched by AgnerKrarupErlang[8], when he created models to describe the Copen-Hagen telephone exchange.
In the proposed model, the Queue Theory is used to compute various parameters affecting the users as well as the system as a whole.

C. Kendall's Notation

The Kendall notation, after David Kendall[9], is developed to describe queuing systems. The notation is of the form A/B/c/K/m/Z where,
A stands for the arrival time distribution,
B stands for the service time distribution,
c is the number of servers available,
K the total system capacity (maximum number of customers that could be accommodated in the system), m the total source population, and
Z - the type of queue discipline.

Usually, the shorter notation 'M/M/1': it is assumed that there is no limit to the length of the queue, the customer source is infinite, and the queue discipline is FCFS (First Come - First Serve).

D. The Poisson's Distribution

Poisson's Distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time. If a Queue System keeps track of the number of users waiting in the queue for a specific service, and if the event of arrival of one customer is independent of arrival of others, it follows a Poisson's Distribution. The discrete values of the random variable 'x' which denotes the event of arrival of 'n' customers at time 't'. [1]

$$P_t(\text{'x'}) = \left(e^{-\lambda} * \lambda^x\right) \div x!$$

Where P_t is the Probability function and it describes event of 'x' number of people arriving at time t ($0<P<1$).

E. The M/M/1 Model

The M/M/1 queue model assumes infinite system capacity (k) and infinite number of population source (m). The system follows Poisson's Distribution. Proofs of these formulas are provided in the reference [1]. We consider total system wait time as a sum of the wait time in the queue and wait time in the service.
L: Number of People
W: Waiting time
We use a M/M/1 queue model to design the model in the paper as it is an ideal model in real-time systems. The formulas derived for the M/M/1 model are listed as follows:
$\rho = \lambda / \mu$ (Traffic Intensity or Utilization parameter)

L_{sys} (Number of people in the System): $\lambda / (\mu-\lambda)$
W_{sys} (Total wait time) = $W_q + W_{service}$
$W_{sys} = 1/ (\mu - \lambda)$
Little's Law : $L = \lambda W$[10]
L_q(Number of people in Queue) = $L^*s - (\lambda / \mu)$

III. DESIGN AND IMPLEMENTATION

The design of the model is done assuming that the queue is a structured queue. As discussed before, the structured queues include the queues that are found in banks, retails and other similar public places.

The Company/Organization has to provide the list of services they provide on the user application.
The user shall then choose the required service.
The service chosen shall be entered in the database. Since the service time differs from one service to the other, there is a need to collect the service time data for all individual services. The data is collected so as to provide better prediction of P_t using formulae listed in the M/M/1 model (Section 2). The predicted time (P_t) is computed and is provided to the user. It shall estimate the time before which the user shall have to be physically present in the service premises. The user is given virtual P Sequence Token (P_{seq}) when he/she registers in the application along with the choice of their service.

The user is regularly reminded (notified) as his/her turn comes closer. If the user confirms their presence in/around premises, they shall be given a C Sequence Token (C_{seq}). Confirmation could be done using location APIs or a QR Scanning method. And the user shall be provided the service shortly after their arrival.

The prime reason behind providing the separate Sequence Tokens to the users is to ensure that the users are physically present right around the time of their turn. In case, the user fails to confirm his/her presence, it will to lead theTimeout (End of P_t). Also, they shall be reassigned the nearest available P_{seq} Token (In place of timeouts that happen for subsequent users or the last turn at that point of time), such that it shall be minimal at that point of time. This ensures that they can still be in the line without disturbing the system. The models of the user side application and the server are discussed below.

A. User Application

The user needs to have the application to register for their turn in the queue.

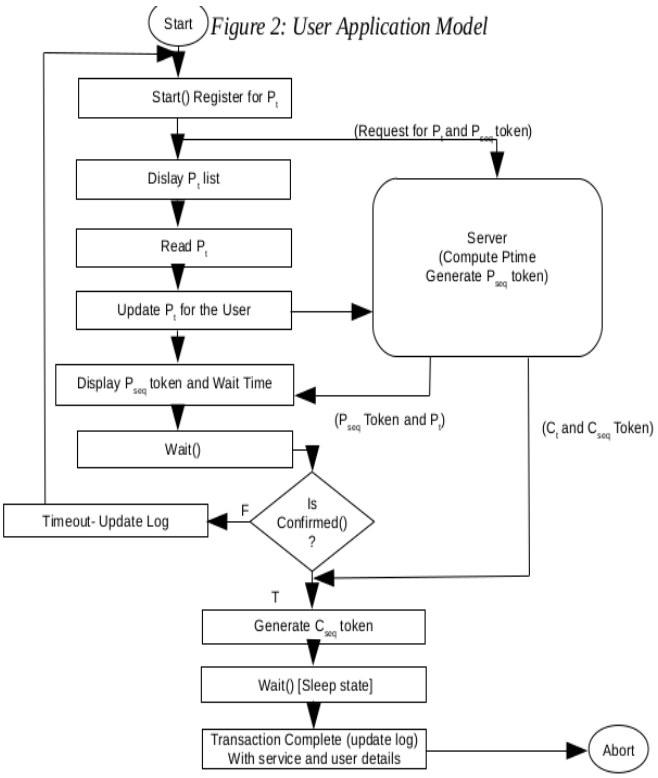


Figure 2: User Application Model

Figure 1

The application from which the user shall register has the following function

- Fetching the data from the server such as token, wait time, information of the position in the line.
- Transferring of data such as service time and other information related to emergencies.
- Update the real time queue information such as position and wait time.

1) Requesting the service:

As the user registers for a service, the request is sent to the server. This information shall contain id of the service which the user desires to receive. This in turn shall be stored in the database. The total wait time is calculated between the users' arrival and their finish of service. This information is useful in calculating the service rate for future predictions.

2) Calculate the predicted time:

The request sent by the user application shall trigger the server to generate a Token number along with its Predicted time (P_t). This data is then fetched by the user device.

3) Update the queue:

As the transactions progresses, when the customers finish their service, the server updates the time as well as the users' position in the queue.

The user application is depicted in Figure 2. This is a conceptual design of the system and it describes the work flow of the model. The notations used in the figures hold the same meaning as described earlier.

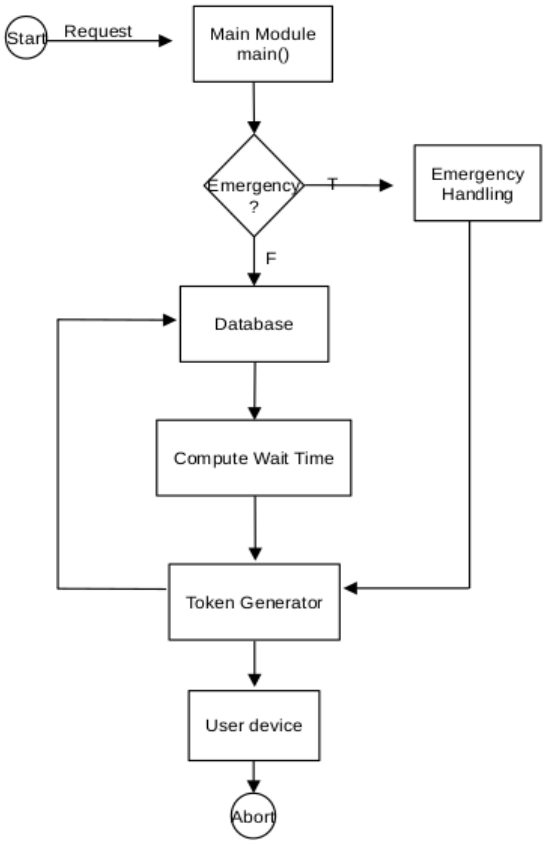
B. Server

The server on the other hand, has three major functions:

- Generate token and predict time.
- Update users as queue progresses
- Handling emergencies.

The computation of wait time depends on average service rate and arrival rate. As discussed in section 2, the P_t is calculated and transferred to the user device as depicted in Figure 3.

Figure 3: Server Model



Figuer 2

1) Database

Database holds the values of the arrival time of the user, along with the ID of the chosen service. Thus, the service rate could vary for two different services. A sample structure is depicted in the table below.

Table 1.sample database

User ID	Service ID	Arrival time	Service duration
01	S-02	8:45	4:23 min
02	S-01	8:50	2:11 min
03	S-02	9:01	3:53 min
04	S-03	9:04	1:40 min
05	S-01	9:09	2:35 min

2) Computing Wait time

It can be seen that the arrival rate of the customer varies from time to time. Thus, the arrival rate of the customer should be calculated for a specific time interval of the day. If suppose a bank records the of

arrival of customers from 10 a.m to 11 a.m, the mean arrival rate calculated will be different from that of the one which is calculated between 2 p.m and 3 p.m. The time estimation thus, should be done using the arrival rate related to a specific part of the day and not as the day as a whole. From the formula of arrival rate between time interval t_1 and t_2,

λ = number of customers between the interval/ $(t_2 - t_1)$

3) Generate Token

As the time is estimated, the token is generated and given to the user. The token Generator will have to manage the requests in order to provide the correct token (P_{seq} or C_{seq}) as per their turn.

I. RESULTS AND TEST CASES

The test cases of the model are tabulated below. The test cases show the conceptual working of the model of a structured M/M/1 Queue model. The values of time interval could vary from service to service, person to person who is handling the service, specific time of the day or a particular day of a week.

A. Test 1: General Case

The user registers for the service at 8:26 am, and she is given a P_{seq} token (P-003). As the turn approaches, she is notified for the confirmation of her presence. The user confirms at 9:12 am and accordingly a C_{seq} token (C-003) is given. The wait thereafter is about 3 minutes.

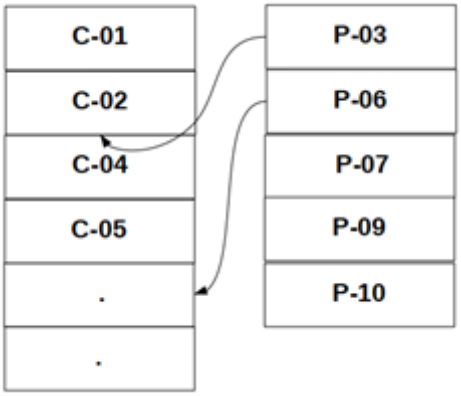


Illustration 1: General Case

Table 2.test case 1

Index	Description	Values
01	Service ID	01
02	Time of Registration	8:26
03	P_{seq} token	P-003
04	Estimated Wait Time (P_t)	9:12 min
05	Confirmed? C_{seq}	Yes (9:14)
06	C_{seq}	C-003
06	Service complete at	9:18

B. Test 2: Unconfirmed User – I

The user registers for the service at 3:47 pm, and he is given a P_{seq} token (P-002). As the turn approaches, he is notified for the confirmation of his presence. The user fails to confirm their presence and hence it will lead to Timeout. Since there is another Timeout after this user, user with P-001 token is given a position in place of that user. And further, after confirmation a C_{seq} token (C-005) is given.

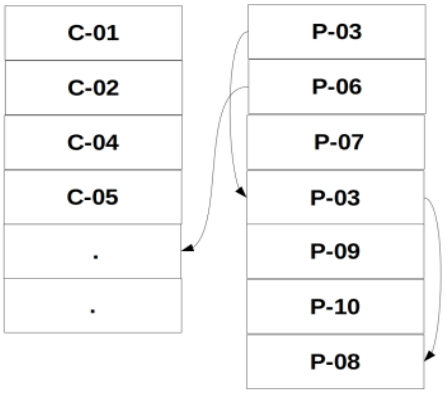


Illustration 2. Timeout Swap

Table 2. Test case 2

Index	Description	Values	Remarks
01	Service ID	02	
02	Time of Registration	3:47	
03	P_{seq} token	P-002	
04	Estimated Wait Time (P_t)	4:35 min	
05	Confirmed? C_{seq}	No	No C_{seq}
06	New P_{seq} Token (replaced in place of the nearest timeout user)	P-010	From the nearest timeout
07	New Estimated Wait time	4:55	
08	User confirmed for C_{seq} token	4:52	
09	C_{seq} token	C-010	
10	Service complete at	5:02	

C. Test 3 : Unconfirmed User – II

The user registers for the service at 2:05 pm, and he is given a P_{seq} token (P-004). As the turn approaches, he is notified for the confirmation of his presence. The user fails to confirm their presence and hence it will lead to Timeout. Since there are no other Timeout after this user, user with P-004 token is given a position in place of the nearest available turn.

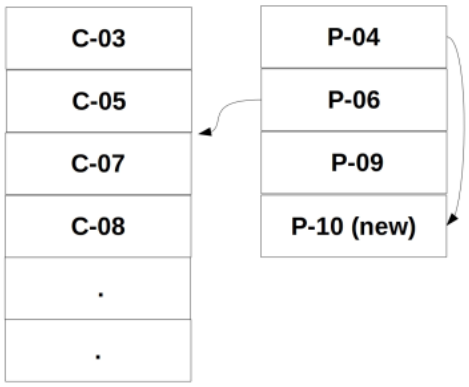


Illustration 2. Nearest Available Position

TABLE 3. TEST CASE 3

Index	Description	Values	Remarks
01	Service ID	02	
02	Time of Registration	2:05	
03	P_{seq} token	P-004	
04	Estimated Wait Time ($P_t$min)	3:37	
05	Confirmed? C_{seq}	No	No C_{seq}
06	New P_{seq} Token (Issue the nearest position)	P-010	New token
07	New Estimated Wait time	4:55	
08	User confirmed for C_{seq} token	4:53	
09	C_{seq} token	C-010	
10	Service complete at	5:03	

D. Test 4: Emergency Scenario

The user registers for the service at 2:33 p.m, and he is given a P_{seq} token (P-010). Since the user registered as an emergency case, the nearest Timeout after this user, user with P-007 token is swapped with P-010 user. The wait thereafter is about 4 minutes. After the service, the service time is recorded in the database.

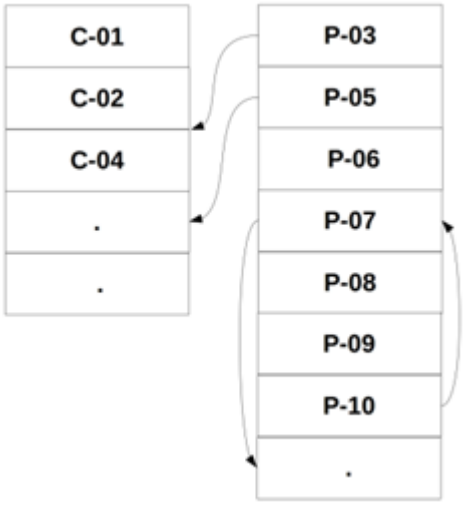


Illustration 3. Emergency Handling

TABLE 4. TEST CASE 4

Index	Description	Values	Remarks
01	Service ID	01	
02	Time of Registration	2:33	
03	P_{seq} token	P-010	Nearest

			timeout
04	Estimated Wait Time (P_t)	2:49	
05	Confirmed ?C_{seq}	Yes (2:49)	
06	C_{seq}	C-010	
06	Service complete at	2:51	

IV. CONCLUSION

The classical Queue Theory has been continuously studied for more than a century. Various models of the queues have been studied among which, the M/M/1 Queue Model is the most relevant in our model. As described before, the Model discussed in this paper is best suited for a structured queue. A structured queue with a single server follows Poisson's Distribution, and the probabilities of arrival could be computed by referring the data of arrival of customers. This takes us to the next topic of calculation of arrival and service rates. The observed values will serve as a reference to compute the average wait time for the subsequent users. It is also possible to estimate the expected number people arriving, which helps in improving the services and when there is a rapid growth.

V. FUTURE WORK

Having computed the estimated values of time and users on any particular part of the day, the prediction algorithms can be accordingly improved. The proposed model needs improvement in handling the emergencies and rapid variations in the queues. As this model directly interacts with a real world problem, there are chances that it might face a failure. Also emergencies are recommended to be handled directly by the Organization in order to make the decisions on the basis of their policies.

The system also requires an improvement in the prediction algorithms which could be enhanced through Data Analysis. The future work also involves the addition of an error correction factor to meet the needs of real-time systems.

VI. REFERENCES

[1]. Allen, Arnold A. (1990). Probability, Statistics, and Queueing Theory: With Computer Science Applications. Gulf Professional Publishing. p. 259. ISBN 0120510510.

[2]. Customers' Evaluations of Queues: Three Exploratory Studies A.Th.H. Pruyn, A. Smidts - European Advances in Consumer Research Volume 1

[3]. A Survey of Recent Developments in Queue Wait Time Forecasting Methods, Ron Davis Tamara Rogers , Yingping Huang

[4]. Automatic Queuing Model for Banking Applications Dr. Ahmed S. A. AL-Jumaily Dr. Huda K. T. AL-Jobori

[5]. Smart Queue Management System Using GSM Technology, Arun, R and Priyesh, P.P.

[6]. Smart Token Bank System Prof. Mr. Ganesh Attarde, Snehal P. Shahane, Prasad Mahajan, Vaibhav Yadnik

[7]. Erlang, Agner Krarup (1909). "The theory of probabilities and telephone conversations"

[8]. Kendall, D.G.:Stochastic processes occurring in the theory of queues and their analysis by the method of the embedded Markov chain, Ann. Math. Stat. 1953

[9]. Little, J. D. C. (1961). "A Proof for the Queuing Formula: L = λW"