

Queue Model Simulation of M/M/1 with Retention of Reneged Customers and Balking

Michelin Wara J¹, Dodi Devianto², Ferra Yanuar³

¹ Post Graduate of Mathematics, Faculty of Mathematics and Natural Sciences,
Andalas University, West Sumatera, Indonesia

^{2,3} Mathematics Department, Faculty of Mathematics and Natural Sciences,
Andalas University, West Sumatera, Indonesia



Abstract - Impatience on customers towards a line can be solved with a queue theory. By this matter, this article gives the queue model simulation of M/M/1 with the retention of reneged customers and balking. The retention of reneged customers is a customer who is impatient (reneged) but staying in a line to wait to be served (retention), while balking is the behavior of customers who are reluctant to enter the queue when they arrive. A simulation is conducted by generating data which meets the queue model of M/M/1 with retention of reneged customers and balking, and then takes a number of variations in the value of system capacity (N). The purpose of the simulation is to see the effect of system capacity on various measures of system effectiveness. The result shows that, in this queue model, the greater the capacity system, the smaller the measure of system effectiveness.

Key words - *Queuing, Reneging, Retention, Balking, Steady State Solutions, System Capacity.*

I. INTRODUCTION

Queuing is often time found on a daily routine, in hospitals, banks and other public services. The queue process relates to the arrival of the customer in a service spot, waiting in line if they have not received the service yet, being served, then leave the queue after getting the service. However, customers often find themselves impatient when lining up, then decide to leave the line before getting the service. If this keeps happening, it will cause a loss for a service provider because they make a customer decrease. Therefore, a solution to solve this problem is a queuing theory model.

In the year of 2011, Liao [10] has done a research about the lining pattern to estimate the loss using a balking index and reneging rate. Sugito et al. [12] have used a queue theory to analyze the model of lining system in the highway of Banyumanik. They obtained some queuing models for the

regular highway and an electronic lane (automatic lane). Then, Jhala and Bhatawala [8] also conducted a study about the theory implementation of the banking sector.

In the queuing theory, there are some queue models, one of the few is M/M/1 model. According to Bronson and Naadhimutu [2], the M/M/1 model is a queue model where the customers's arriving in a Poisson distribution (exponentially distributed inter-arrival time), the departure of the customer is Poisson distribution (exponentially distributed service time) and has one server (service facility).

The queue model is built to refer to the situation. Often time in building a queue model, the situation is assumed to be ideal where human behavior which tends to wait impatiently is being ignored, while in reality, it is not all the customers are patient to wait to be served, unless it is a product. According to Gross and Harris [6], human behavior in the queue system

if acting as a customer as the following descriptions. Firstly, reneging, describing a situation where a customer enters the queue then decides to leave the line before getting the service. This reneging customer can do some efforts to stay in the line by hope to get the service; this is what is called the retention of reneged customers [9]. Secondly, balking is the reluctance of the customer to enter the queue when they arrive. Thirdly, jockeying, describing the situation if there are two or more lines in the system, the customer can switch the lines from one to another.

By considering a human behavior as a customer, the queue data obtained is getting closer to the actual situation, therefore we will discuss the queue model simulation of M/M/1 with the retention of reneged customers and balking, where the purpose of the simulation is to see the effect of capacity system on various measures of system effectiveness in the queue model of M/M/1 with retention of reneged customers and balking.

II. Literature Review

2.1 Poisson Distribution and Exponential Distribution

Poisson distribution is often used as a description of the input process in the queueing system [3]. It means, this distribution is used to describe the distribution of the number of customers that arrive during a certain time interval, meanwhile the exponential distribution, is widely used in problems of reliability such as to show the solution of queue with service time as the sum of exponential random variables [4]. This means, the exponential distribution is used to describe the distribution of service time in a service facility.

2.2 M/M/1 queue model

According to Bronson and Naadhimutu [2], the M/M/1 model is a queue model where the customers's coming in a Poisson distribution (exponentially distributed inter-arrival time), the departure of the customer is Poisson distribution (exponentially distributed service time) and has one server (service facility).

2.3 Kendall Notation

According to Taha [13], the standard notation for modeling the queue system was first put forward by D.G. Kendall in form of $a/b/c$ which was then known as Kendall notation. Then A.M. Lee added the symbols d and e so that they become $a/b/c/d/e$ called Kendall-Lee notation, then Taha adds the symbol f so that the characteristics of a queue can be denoted in the following standard format:

$$(a/b/c):(d/ef)$$

where,

- a : arrival distribution (inter-arrival time),
- b : departure distribution (service time),
- c : number of servers/service facilities,
- d : service discipline, eg. first come first served (FCFS), last come first served (LCFS), priority service (PS), service in random order (SIRO),
- e : the maximum number allowed in the system (system capacity), can be limited or unlimited,
- f : the size of the caller's source, can be limited or unlimited.

The standard notation of a and b (the arrival and departures distribution) can be replaced with the following code.

- M : arrivals and departures are Poisson or markov distribution, or equivalently, the inter-arrival time and service time are distributed exponentially.
- D : inter-arrival time or service time is deterministic.
- E_k : inter-arrival time or service time is erlang distribution.
- GI : general independent distribution of arrivals or inter-arrival time.
- G : general distribution of departure or service time.

2.4 Birth-Death Processes

The process of arrival and departure in a queue system can be illustrated by the process of birth and death. A birth occurs when a customer enters the queue system meanwhile a death happens if a customer leaves it.

The process of birth and death is a stochastic process in form of a summing process in a system where the state of the system always produces n positive integer which states the number of customers [14].

The state of the system when t is defined as the difference between the number of births and deaths at the t time. It means that the state of the system when t in a queue system is the difference between the number of arrivals and departures at the t time. For instance, the number of arrivals is denoted by $X(t)$, the number of departures is denoted by $Y(t)$, the number of customers when t is denoted by $N(t)$, then $N(t)=X(t)-Y(t)$. The probability of n customer in the queue system when t is denoted by $P(N(t)=n)$ or $P_n(t)$.

There are several models of birth and death processes, one of a few being used are the process of pure birth and death. The pure birth process is the birth without death, meaning the arrival without the departure, while the process of pure death means the death without birth, meaning the departure without the arrival.

2.5 Steady State Condition

A very important assumption in queuing theory is whether the system reaches a state of equilibrium or is called a steady state. This means that it is assumed that the characteristic of the operation after the queue length and the average waiting time will have a constant value after the system runs for a period of time. This means that the probability of n customers at time t does not depend on time. Steady state condition occur when $\frac{dP_n(t)}{dt} = 0$, so $\lim_{t \rightarrow \infty} P_n(t) = P_n$. As a result of this, $P_n(t) = P_n$ for all t .

2.6 Measures of system effectiveness

To measure the effectiveness of queue performance system, it uses:

1. The expected value of queue system size (L_s).

According to Ecker [5], L_s can be stated as follow:

$$L_s = \sum_{n=0}^{\infty} n P_n \quad (1)$$

2. The expected value of customer waiting time in queue system (W_s).

Little's Law says that, under steady state condition, the average number L of customers in a queuing system equals the arrival rate λ multiplied by the average time W that a customer spends in system [11]. Expressed algebraically, the law is

$$L = \lambda W$$

This relationship is remarkably simple and general. So, based on the equation of little law above, if W_s is customer waiting time in queue system, then the relation between L_s and W_s can be written below:

$$L_s = \lambda W_s \quad (2)$$

3. The expected value of queue length (L_q)

As stated by Hillier [7], L_q can be formulated below:

$$L_q = \sum_{n=0}^{\infty} (n - c) P_n \quad (3)$$

The symbol c states the number of active servers.

4. The expected value of customer waiting time in line (W_q).

By the equation of Little Laws [11], the relation between L_q and W_q can be written below:

$$L_q = \lambda W_q \quad (4)$$

III. MATERIALS AND METHOD

3.1 Data Sources and Variable Research

The data used for this research are data generated using Minitab software that meets the queue model of M/M/1 with retention of reneged customers and balking.

The generated data is the data of the customer's arrival time, on the service time, after service, and on renegeing time. The generated data is 200. This data will be used for simulation in the queue model of M/M/1 with the retention of reneged customers and balking, where the simulation will see the effect of variations in system capacity values (N) on various measures of system effectiveness.

3.2 Method of Analysis

Steps to be taken in this study are as follows.

1. Determine the form of the M/M/1 queue model with retention of reneged customers and calculate the steady state solution and various measures of its effectiveness.
2. Perform the queue model simulation of M/M/1 with the retention of reneged customers and balking with the following steps:
 - a. Generating data that meets the queue model of M/M/1 with the retention of reneged customers and balking. The generated data is inter-arrival time, service time, and renegeing time.
 - b. Calculating the characteristics of the queue model from the data; arrival rate, service rate, renegeing rate, renegeing probability, retention probability and the probability of zero customers.
 - c. Simulating the data by taking a number of variations in capacity system (N) values to see how they affect the various measures of system effectiveness.
 - d. Draw conclusions from simulation results.

IV. RESULT AND DISCUSS

4.1 Assumptions on the Queuing Model Of M/M/1 with Retention of Reneged Customers and Balking

There are several assumptions used in the queue model of M/M/1 with the retention of reneged customers and balking, such as:

1. Arrival occurs according to the Poisson distribution one by one with the arrival rate parameters λ and the inter-arrival time exponentially distributed.
2. There is only one server and the service time is exponentially distributed with the service rate parameter μ .

3. The queue discipline is FCFS.
4. Limited capacity system, N .
5. Every customer who arrives in the queue will wait for a certain time (reneging time) to start being served. If the customer has not yet begun to be served, he will be impatient and leaves the line without getting service with probability p and may remain in line to wait for the service with probability $q=1-p$. The reneging time follows the exponential distribution with the reneging rate parameter ξ .
6. Customers who come will be balking (reluctant to enter the queue) with probability $\frac{n}{N}$ when customer is in front, where n is the number of customers in the system and N is the maximum number of customers allowed by the system (system capacity) and vice versa. It means that they will not balk or decide to enter the queue with probability $1 - \frac{n}{N}$.

4.2 Queuing Model of M/M/1 with Retention of Reneged Customers and Balking

To get the formula of M/M/1 with retention of reneged customers and balking, it has to make sure the possibilities that can occur so that there are n customers at time $t+\Delta t$ as determined as follows.

1. The system is in the state of n at time of t and during the time interval Δt , there is no arrival and departure and no reneging occurs.
2. The system is in the state of n at time of t where an arrival and a departure both occur during the time interval Δt .
3. The system is in the state of n at time of t and during the time interval Δt , reneged the customer with probability $q=1-p$, does not leave the queue, in other words, it remains in the system waiting to be served.
4. The system is in the state of $n-1$ at time of t and there is an arrival during the time interval of Δt .
5. The system is in the state of $n+1$ at time Δt where there is one departure and no arrival during the time interval t .
6. The system is in the state of $n+1$ at the time of Δt and during the time interval Δt , the impatient customer with the probability p , leaves the queue.

Because the possibilities events that caused the presence of n customers at time $t+\Delta t$ are independent, then:

$$\begin{aligned}
 P_n(t + \Delta t) &= P_n(t)(1 - n\xi\Delta t + \xi\Delta t - \mu\Delta t - \lambda\Delta t + \frac{n}{N}\lambda\Delta t) \\
 &+ o(\Delta t) + P_n(t)(qn\xi\Delta t - q\xi\Delta t) + P_{n-1}(t)(\lambda\Delta t - \frac{n-1}{N}\lambda\Delta t) \\
 &+ P_{n+1}(t)(\mu\Delta t) + P_{n+1}(t)(n\xi p\Delta t); 1 \leq n \leq N-1
 \end{aligned} \quad (5)$$

The symbols of λ , μ and ξ sequentially stated the arrival rate, service rate and reneging rate.

The initial condition given to this model is:

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t) \quad (6)$$

Based on this theorem limit:

$$\lim_{\Delta t \rightarrow 0} \frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = \frac{dP_n(t)}{dt}$$

so, the equation of (5) can be resolved by reducing the both equation by $P_n(t)$ then it is divided by Δt , then we can take the limit of Δt through a zero number, then it is obtained:

$$\begin{aligned}
 \frac{dP_n(t)}{dt} &= -((1 - \frac{n}{N})\lambda + \mu + (n-1)\xi p)P_n(t) \\
 &+ ((1 - \frac{n-1}{N})\lambda)P_{n-1}(t) + (\mu + n\xi p)P_{n+1}(t); 1 \leq n \leq N-1
 \end{aligned} \quad (7)$$

for $n=N$, it is obtained:

$$\frac{dP_N(t)}{dt} = ((1 - \frac{N-1}{N})\lambda)P_{N-1}(t) - (\mu + (N-1)\xi p)P_N(t) \quad (8)$$

4.3 Steady State Solution on the Queue Model of M/M/1 with the Retention of Reneged Customers and Balking

In a steady state condition, $\frac{dP_n(t)}{dt} = 0$, so $\lim_{t \rightarrow \infty} P_n(t) = P_n$. As a result of this, $P_n(t) = P_n$ for all t . Therefore, the equation of (6), (7) and (8) in a steady state are as follows.

$$0 = -\lambda P_0 + \mu P_1 \quad (9)$$

$$\begin{aligned}
 0 &= -((1 - \frac{n}{N})\lambda + \mu + (n-1)\xi p)P_n \\
 &+ ((1 - \frac{n-1}{N})\lambda)P_{n-1} + (\mu + n\xi p)P_{n+1}; 1 \leq n \leq N-1
 \end{aligned} \quad (10)$$

$$0 = ((1 - \frac{N-1}{N})\lambda)P_{N-1} - (\mu + (N-1)\xi p)P_N; n = N \quad (11)$$

Furthermore, by completing the equation of (9) - (11), then it is obtained a steady state solution as follows.

$$P_n = P_0 \prod_{m=1}^n \frac{N - (m-1)}{N} \frac{\lambda}{\mu + (m-1)\xi p}; 1 \leq n \leq N-1 \quad (12)$$

$$P_N = P_0 \prod_{m=1}^N \frac{N - (m-1)}{N} \frac{\lambda}{\mu + (m-1)\xi p}; n = N \quad (13)$$

Because $\sum_{n=0}^N P_n = 1$, so it is obtained:

$$P_0 = \frac{1}{(1 + \sum_{n=1}^N \prod_{m=1}^n \frac{N - (m-1)}{N} \frac{\lambda}{\mu + (m-1)\xi p})} \quad (14)$$

4.4 Measures of effectiveness of M/M/1 queue model with Retention of Reneged Customer and Balking

For the queuing model of M/M/1 with the reneged retention customer and balking, the measure of effectiveness is as follow.

1. The expected value of queue system size (L_s)

Based on the equation of (1) and (12), it is obtained:

$$L_s = \sum_{n=0}^N n \prod_{m=1}^n \frac{N - (m-1)}{N} \frac{\lambda}{\mu + (m-1)\xi p} P_0 \quad (15)$$

2. The expected value of the customer waiting time in the queue system (W_s)

Based on the equation of (2) and (15), it is obtained:

$$W_s = \frac{1}{\lambda} \sum_{n=1}^N n \prod_{m=1}^n \frac{N - (m-1)}{N} \frac{\lambda}{\mu + (m-1)\xi p} P_0 \quad (16)$$

3. The expected value of value of queue length (L_q)

Based on the equation of (3) and (15), it is obtained:

$$L_q = \sum_{n=1}^N n \prod_{m=1}^n \frac{N - (m-1)}{N} \frac{\lambda}{\mu + (m-1)\xi p} P_0 - \frac{\lambda}{\mu} \quad (17)$$

4. The expected value of customer waiting time in line (W_q)

Based on the equation of (4) dan (15), it is obtained:

$$W_q = \left(\frac{1}{\lambda} \sum_{n=1}^N n \prod_{m=1}^n \frac{N - (m-1)}{N} \frac{\lambda}{\mu + (m-1)\xi p} P_0 \right) - \frac{1}{\mu} \quad (18)$$

4.5 Queue Model Simulation of M/M/1 with Retention of Reneged Customers and Balking

In this simulation, the generated data is 200. It means that the number of customers or $n=200$. The generated data

consists of inter-arrival time, service time and reneging time of the customer.

The generated data is in form of random numbers that have met the queue model of M/M/1 with the retention of reneged customers and balking, where the arrival of the customer is Poisson distribution (exponential inter-arrival time), customer departure is Poisson distribution (service time is distributed exponential), and reneging time is exponentially distributed.

Furthermore, the random numbers are converted to the desired time unit. In this article, a second unit is selected, so each random number is multiplied by 60 (assuming that at first, the random number is in unit of minutes).

For data of inter-arrival time, it means that the generated data is the difference in time of arrival between the first customer and the second customer, and others. It can be determined at what time the first, second and other customers arrive at the queue time according to the data that has been researched.

For the service time data, it means that the generated data is the length of time of each customer served, so that it can be determined at what time the first, second and other customers start and finish the service. Whereas for the reneging time data, it means that the generated data is the length of time the customer waits until the customer finally cancels the queue (reneging), so that it can be set at what time a customer do the reneging according to the data that has been research. In this article, 33 customers reneging are then taken.

After obtaining data that meets the model, the data is processed to obtain some characteristic of the queue as follow.

1. Arrival rate (λ)

Arrival rate is the number of arrivals per time unit. Based on inter-arrival time data, there are 200 arrivals during 25.106 seconds ≈ 7 hours, so that arrival rate (λ) is $\frac{200}{7} = 28,57$ arrivals per hour.

2. Departure rate (μ)

Departure rate is the number of customers who have been served per time unit. Based on service time data, there are 167 customer who have been served during 5,56 hours, so that departure rate (μ) is $\frac{167}{5,56} = 30,03$ departures per hour.

3. Reneging rate (ξ)

Reneging rate is the number of reneging customer per time unit. Based on reneging time data, there are 33 reneging customers during 5,01 hours, so that reneging rate (ξ) is $\frac{33}{5,01} = 6,58$ reneging per hour.

4. Reneging probability (p)

Based on generated data, there was 33 reneging customers from the total number of the customers, 200. So that, reneging probability (p) is $\frac{33}{200}=0,165$.

5. Retention probability (q)

According to assumption 5 that is $q=1-p$, then $q=1-0,165=0,835$. This means that the probability of retention customer (q) is 0,835.

These five characteristic value of the queue are summarized in the following table 1.

Table 1. Value for Some Characteristics of the queue

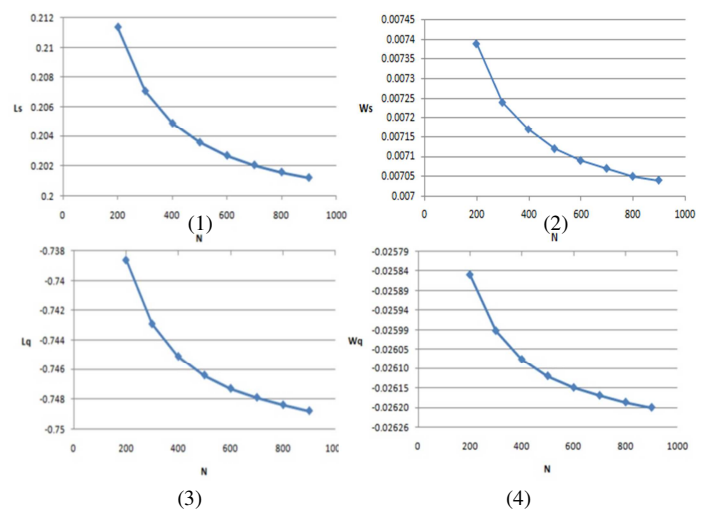
Queueing Characteristics	Value
Arrival Rate (λ)	28,57
Departure Rate (μ)	30,03
Reneging Rate (ξ)	6,58
Reneging Probability (p)	0,165
Retention Probability (q)	0,835

Then, we must determine the P_0 value by substituting the queue characteristic value in table 1 into the equation (14). After that, a simulation of the variation value of capacity system (N) against various measures of effectiveness system is carried out. Because the capacity system (N) is assumed to be limited, so, $N=200, 300, 400, 500, 600, 700, 800$ and 900 are taken. The result of the simulation and P_0 value are summarized in the following table 2.

Table 2. The result of the simulation and P_0 values

N	P_0	L_s	W_s	L_q	W_q
200	0,1618733507	0,2113848468	0,00739	-0,7386	-0,025852
300	0,1593116585	0,2070700465	0,00724	-0,7429	-0,026002
400	0,1580072652	0,2048840480	0,00717	-0,7451	-0,026079
500	0,1572168114	0,2035630183	0,00712	-0,7464	-0,026125
600	0,1566865160	0,2026783345	0,00709	-0,7473	-0,026156
700	0,1563060820	0,2020444374	0,00707	-0,7479	-0,026177
800	0,1560198446	0,2015679232	0,00705	-0,7484	-0,026195
900	0,1557966712	0,2011966503	0,00704	-0,7488	-0,026209

To see the relationship between variation value of capacity system (N) against various measures of effectiveness system, the result of the simulation in table 2 can be presented in the following graphs.



The result of the simulation in table 2 and graphs can be explained as follows.

1. Variation in the value of capacity system (N) to the expected value of system size (L_s).

The L_s value in table 2 is obtained by substituting several variations of N value of $200 \leq N \leq 900$, the queue characteristic value (table 1) and P_0 values to equation 15.

The simulation result in table 2 and the graph (1) show that the greater the value of capacity system (N), the expected value of the system size (L_s) becomes smaller. That is, by increasing the capacity of the system, the expected value of the queue system size becomes small.

2. Variation in the value of capacity system (N) to the expected value of customer waiting time in the queue system (W_s).

The W_s value in table 2 is obtained by substituting some N values of $200 \leq N \leq 900$, queueing characteristic value (table 1) and P_0 values to equation 16.

The simulation result in table 2 and the graph (2) show that the greater the value of capacity system (N), the expected value of customer waiting time in the system (W_s) becomes smaller. That is, by increasing the system capacity, then the customer does not have to wait long in the system.

3. Variation in the value of N to the expected value of the queue length (L_q).

The L_q value in table 2 is obtained by substituting some N values of $200 \leq N \leq 900$, queue characteristic value (table 1) and P_0 values to equation (17).

The simulation result in table 2 and graph (3) show that the greater the value of the capacity system (N), then the expected value of the queue length (Lq) becomes smaller. The negative sign indicates that the value of Lq is very small which means there is no queue (for a value of $200 \leq N \leq 900$).

4. Variation in the value of N to the expected value of customer waiting time in line (Wq).

The value of Wq in table 2 is obtained by substituting some N values of $200 \leq N \leq 900$, queuing characteristic value (table 1) and P_0 values to equation (18).

The simulation result in table 2 and graph (4) show that, the greater the value of the capacity system (N), then the expected value of the customer waiting time in line (Wq) becomes smaller. The negative sign indicates that the value of Wq is very small, which means the customer does not need to wait in line (for a value of $200 \leq N \leq 900$). This relates to the value of Lq obtained previously where the value of Lq is very small which means there is no queue. As there is no queue, the customer does not need to wait in line.

V. CONCLUSION

Based on the research conducted about the queue model simulation of M/M/1 with a retention of reneged customers and balking, where the simulation of system capacity value to various measures of system effectiveness have been done, it can be concluded that system capacity affects the size of system effectiveness. By increasing the capacity system, the measures of the system effectiveness becomes small. If the measures of system effectiveness are small, it means that the customer does not need to linger in line or be in the system, so that the customers who tend to be impatient can remain in line (retention), as a result, the service provider will not lose a lot of customers due to impatience. So the losses can be minimized.

REFERENCES

- [1] Bain, L.J. and Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistic*. United States of America: Brooks/Cole.
- [2] Bronson, R and Naadimuthu, G. (1992). *Schaums Outline of Theory and Problems of Operations Research*. New York: McGrawHill-company.
- [3] Cooper, R.B. (1981). *Introduction to Queuing Theory (second ed.)*. New York: Elseveir North Holland,inc.
- [4] Devianto, D., Maiyastri, Oktasari, L. and Anas, M. (2015). "Convolution of Generated Random Variable from Exponential Distribution with Stabilizer Constant". *Applied Mathematical Sciences*. 9 (96): 4781-4789.
- [5] Ecker, J and Kupferschmid, M. (1988). *Introduction to Operation Research*. New York: Jhon Wiley and Son.
- [6] Gross, D and Harris, C.M. (1974). *Fundamental of Queuing Theory*. New York: Jhon Wiley and Sons.
- [7] Hillier, F.S. and Lieberman, G.J. (2011). *Introduction of Operations Research. Ninth Edition*. New York: McGrawHill-company.
- [8] Jhala, N and Bhatawala, P. (2016). "Aplication of Queuing Theory in Banking Sector". *IOSR Journal of Mathematic*. 12(2): 73-75.
- [9] Kumar, R and Sharma, S.K. (2012). "An M/M/1/N Queueing Model with Retention of Reneged Customers and Balking". *American Journal of Operational Research*. 2(1): 1-5.
- [10] Liao, P. (2011). "A Queuing Model with Balking Index and Reneging Rate". *International Journal of Service and Operations Management*, 10(1): 1-12.
- [11] Little, J.D.C and Graves, S.C. (2008), *Little's Law. In: chhajed, D. and Lowe, T.J. (eds.) Building Intuition: Insight from Basic Operation Management Model and Principles*, Springer Science+Business Media.
- [12] Sugito, Rahmawati, R. and Wardhani, J.K.. (2018). "Queue Theory for triangular and Weibull Arrival Distribution Model (Case Study of Banyumanik Toll)". *Journal of Physic. Conference series*. 1025: 1-8.
- [13] Taha, H.A. (2007). *Operation Research an Introduction*. New Jersey: Pearson Education.
- [14] Winston, W.L. (1994). *Operation Research*. California: Duxbury Press.