Queueing Theory

- Queue is a line or a sequence of people, vehicles, etc awaiting their turn to be attended to or to proceed.
- Some examples of queueing theory
    - Banks/supermarkets – waiting for service
    - Ticket reservation
    - Public transport – waiting for a train or bus

Applications

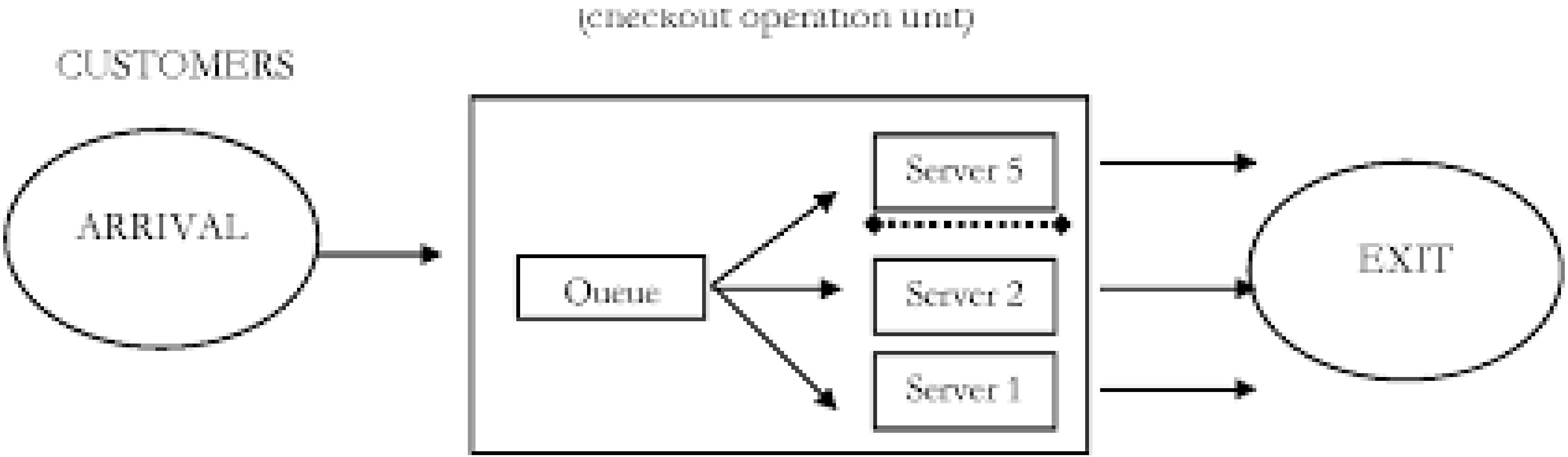
- Airplanes- waiting for their turn in order for them to fly or land in the airports.
- Traffic Management
- Jobs or products that are waiting and assembly line
- Determining sequence of computer operations
- Predicting computer performance

What is queuing theory?

Queuing theory is a mathematical study of the formation, function, and congestion of waiting lines, or queues.

- Someone or something that requests a service—customer, job, or request.
- Someone or something that completes or delivers the services—server.
- Analyze and optimize the system.

Model Queuing system



- ❑Use Queuing models to
  - ✓Describe the behavior of queuing systems
  - ✓Evaluate system performance

Characteristics of a queuing theory

- Arrival Process:
  - how customers arrive e.g. singly or in groups (batch or bulk arrivals)
  - Arrival pattern is very important in discussing the process
  - Customer population arrival is random and infinite. Thus the process follows Poisson or exponential distribution.

- Customer behaviour
  - Balking:

    Customer does not join the queue and leave the system without joining as queue is very long.
  - Reneging:

    Customer joins the queue for short time and then leave the system without service as the queue is very slow.

- Jockeying:

  Customer keep on changing queue in hope to get service faster.

- Priorities:

  Customer are served before others regardless of their order of arrival.

- Service Process/Pattern: Number of customers serviced follow a particular probability distribution. In most of these cases, it follows either Poisson or exponential distribution.

- Service/Queue Discipline: Queue discipline means the order by which customers are picked from the waiting line in order to provide service.

- FIFO : First in first out ,
- LIFO : Last in first out ,
- SIRO : Service in random order

- Capacity/Waiting Room : The number of customer allowed to wait in the queue based on the space availability i.e. capacity of queue.

Queuing theory uses the Kendall notation to classify the different types of queuing systems, or nodes. Queuing nodes are classified using the notation *A/S/c/K/N/D* where:

- ***A*** is the arrival process
- ***S*** is the mathematical distribution of the service time
- ***c*** is the number of servers
- ***K*** is the capacity of the queue, omitted if unlimited
- ***N*** is the number of possible customers or population size, omitted if unlimited
- ***D*** is the queuing discipline, assumed first-in-first-out if omitted

M/M/1 queue model

Assumptions - the Basic Queuing Process

- ✓ Jobs arrive according to a Poisson process with parameter λ for $t \geq 0$,
- ✓ The service time, s, has an exponential distribution with parameter μ , i.e., for $s \geq 0$
- ✓ There is a single server;
- ✓ The buffer is of infinite size; and
- ✓ The number of potential jobs is infinite.
- ✓ The queue discipline is FIFO

Formula Required for M/M/1 queueing model

- The ratio of arrival to service rate indicate the % time server is busy it is known as traffic intensity or utilization factor or average utilization or system utilization or channel efficiency and is denoted by ρ. The unit of the traffic intensity is “Erlang”.

$$\rho = \frac{Arrival\ rate}{Service\ rate} = \frac{\lambda}{\mu}$$

Where λ is the average number of arrivals per unit time i.e. mean arrival rate.

$\frac{1}{\lambda}$ is the mean arrival time.

μ is the average number of customers being served per unit time i.e. mean service time./service rate

$\frac{1}{\mu}$ is the average service time.

- Traffic intensity determines the degree to which the capacity of service station is utilized.
- Generally ρ<1 i.e. λ< μ implies system works.
- If $\rho \geq 1$, then the queue length increases and after certain period of time incoming population will not get serviced. In this condition we can’t find the solution as system ultimately fails.

- At any instant of time the probability that there are n arrivals(customers) in the queue(waiting lie) including those being served is given by $P_n = (1-\rho)\rho^n,\ n \geq 0$
- Probability that there is no one in the queue is $P_0 = (1-\rho)$
- The average or expected number of units(customers) in the system(waiting and being served) is given by $L_s = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$
- The average or expected number of units waiting in the queue is given by $L_q = L_s - \rho = \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)}$

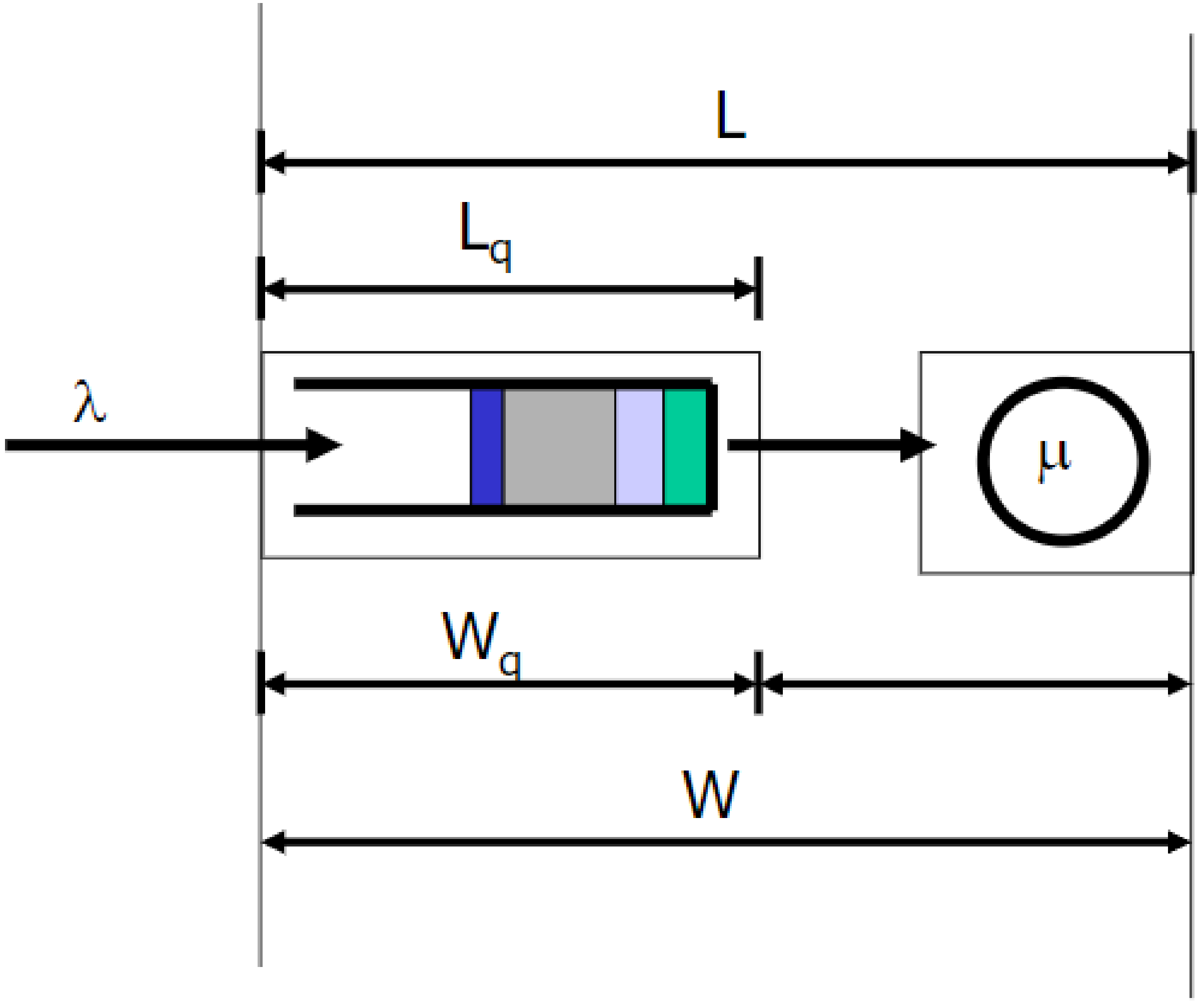
- The mean or expected waiting time of customers in the system (including service time) is $W_s = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu-\lambda}$
- The expected waiting time per unit in the queue i.e. expected waiting time of the customers in the queue excluding service time is $W_q = W_s - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)} = \frac{\lambda}{\mu(\mu-\lambda)}$
- Average length of a non empty queue i.e. expected number of customers in the queue when there is a queue, $L_n = \frac{1}{(1-\rho)} = \frac{\mu}{\mu-\lambda}$

- Average waiting time in a non empty queue, $W_n = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu-\lambda}$
- Probability of queue length being greater than or equal to n is

  equal to $P(queue\ length\ \geq n) = \rho^n = \left(\frac{\lambda}{\mu}\right)^n$
- Interrelation between L_s, L_q, W_s and W_q are

  $L_s = \lambda W_s$ and $L_q = \lambda W_q$

M/M/1 queue model



1. The arrival rate of customers at a banking counter has a mean of 45 per hour. The service rate of a counter clerk has a mean of 60 per hour.

(i) What is the probability of having

(a) No customers in the system.

(b) Having 5 customers in the system.

(ii) Find L_s, L_q, W_s and W_q.

2. In a 24hours service station vehicles arrive at the rate of 30 per day on the average. The average service time for a vehicle is 36 minutes. Find

(i) The mean queue size

(ii) the probability that the queue size exceeds 9?

Solution: $\text{Arrival rate}(\lambda) = 30 \text{ per day} = \dfrac{30}{1440} = 0.0208 \text{ minutes}$

given avg service time=36 minutes

$$\text{service rate}(\mu) = \frac{1}{36} \text{ per minutes}$$

$$\textit{traffic} \text{ intensity}(\rho) = \frac{\lambda}{\mu} = \frac{0.0208}{1/36} = 0.75$$

$$(\mathrm{i})\ \text{mean queue size}\left(\mathrm{L_q}\right) = \frac{\rho^2}{1-\rho} = \frac{(0.75)^2}{0.25} = 2.25$$

$$(ii)\,\mathrm{P}\left(\text{queue exceeds } 9\right) = P\left(\text{queue size} > 9\right) = P\left(\text{queue size } \geq 10\right) = \rho^{10} = (0.75)^{10}$$

3. The mean duration of a telephone conversation is estimated to be 3 minutes. If no more than a 3 minutes wait for the phone may be tolerated. Find the largest amount of incoming traffic that can be supported.

$\textit{Solution}: \text{ Avg service time } = 3 \text{ minutes}$

$\text{service rate}(\mu) = \frac{1}{3} \textit{call per} \min$

$\text{Arrival rate}(\lambda) = ?$

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = 3$$

$$\frac{\lambda}{\left((1/3)^2 - \lambda/3\right)} = 3 \Rightarrow \lambda = 3\left(\frac{1}{9} - \frac{\lambda}{3}\right) = \frac{1}{3} - \lambda \Rightarrow 2\lambda = \frac{1}{3} \Rightarrow \lambda = \frac{1}{6}$$

4. The arrivals at a telephone booth are considered to be following Poisson law of distribution with an average time of 10 minutes between one arrival and the next. Length of the phone call is assumed to be distributed exponentially with a mean of 3 minutes.

(a) What is the probability that a person arriving at the booth will have to wait?

(b) What is the average length of queue that forms from time to time?

(c) The telephone department will install a second booth when convinced that an arrival would expect to wait at least three minutes for the phone. By how much must the flow of arrivals be increased in order to justify a second booth?

Solution: given avg arrival time=10min $\Rightarrow$ Arrival rate$\left(\lambda\right)=\dfrac{1}{10}\,per\min$

Avg service time=3 min $\Rightarrow$ service rate$\left(\mu\right)=\dfrac{1}{3}\,per\ min$

$\left(i\right)$ P$\left(\text{arrival has to wait}\right)=P\left(phone\text{ is busy}\right)=1-P\left(phone\,not\,busy\right)=1-P\left(phone\ is\ idle\right)$

$$=1-P_0=1-\left(1-\rho\right)=\rho=\frac{\lambda}{\mu}=\frac{3}{10}=0.3$$

$\left(ii\right)$ avg queue length$\left(L_q\right)=\dfrac{\rho^2}{1\text{-}\rho}=\dfrac{\left(0.3\right)^3}{0.7}=0.129$

$\left(iii\right)$ Given $\mathrm{W_q}\geq 3\min utes$

$Let\ \lambda_1$ to be the new arrival rate such that $\mathrm{W_q}\geq 3$

$$W_q \geq 3$$

$$\Rightarrow \frac{\lambda_1}{\mu(\mu-\lambda_1)} \geq 3$$

$$\Rightarrow \frac{\lambda_1}{(1/3)((1/3)-\lambda_1)} \geq 3$$

$$\Rightarrow \lambda_1 \geq \frac{1}{6}$$

$$\textit{Increase} \text{ in arrival rate} = \frac{1}{6} - \frac{1}{10} = 0.06 \text{ arrivals per minute}$$

6. A T.V. Repairman finds that the time spent on his jobs have an exponential distribution with mean of 30 minutes. If he repairs sets in the order in which they come in, and if the arrival of sets is approximately Poisson with an average rate of 10 per 8 hour per day, what is repairman's expected idle time each day? How many jobs are ahead of the average set just brought in?

Solution: Avg service time=30 min$=\dfrac{30}{60}=\dfrac{1}{2}\,hrs$

$$\text{service rate}\left(\mu\right)=\frac{1}{30}\ per\ \text{min}=2\ per\ hr$$

Arrival rate $(\lambda) = 10$ per 8 hrs per day= $\frac{10}{8}$ $per\, hr\ \ per\, day$

(a) Expected idle time of repairman each day

= Number of hours for which the repairman remains busy in an 8-hour day (traffic intensity) is given by

$$= 8 * \frac{\lambda}{\mu} = \frac{10}{2} = 5\, hours$$

Hence, the idle time for a repairman in an 8-hour day will be: (8-5)=3 hours

(b) Expected (or average) number of TV sets in the system

$$E(L_s) = \frac{\lambda}{\mu - \lambda} = \frac{5/4}{2 - 5/4} = \frac{5}{3} = 2(Approx)$$

7. The rate of arrivals of planes at an international airport is 20 per hour and the airport can land 30 planes per hour on an average. When there is congestion the planes are forced to fly over the field in the stack awaiting the landing of other planes that arrived earlier.

(i) How many planes would be flying over the field on an average?

(ii) How long a plane would be on the stack and in the process of landing?

Solution: Arrival rate $(\lambda) = 20\ per\ hr,$

$$Service\ rate(\mu) = 30\ per\ hr$$

$$Traffic\ \text{int}ensity(\rho) = \frac{\lambda}{\mu} = \frac{20}{30} = 0.66$$

$$(i)\ \mathrm{L_q} = \frac{\rho^2}{1-\rho} = \frac{(0.66)^2}{1-0.66} = 1.33$$

$$(ii)\ \mathrm{W_s} = \frac{1}{\mu(1-\rho)} = \frac{1}{30(1-0.66)} = 0.098 = 0.1hr$$

8. Customers arrive at the first class ticket counter of a railway station at the rate of 12 per hour. There is a clerk serving the customers at the rate of 30 per hour.

(i) What is the probability that there is no customer in the counter?

(ii) What is the probability that there are more than 2 customers in the queue?

(iii) What is the probability that a customer is being served and no one is being served?

Solution: Arrival rate$(\lambda) = 12 \ per\, hr$,

service rate$(\mu) = 30 \ per \ hr$

traffic intensity$(\rho) = 0.4$

$$(\text{i})\, P(no\, customer) = P_0 = 1 - \rho = 0.6$$

$$(ii)\, P(queue\, length > 2) = P(queue\, length \geq 3) = \rho^3 = (0.4)^3 = 0.064$$

$$(iii) P(customer\, being\, served\, and\, no\, one\, is\, being\, served\, in\, system)$$

$$= P_1 = (1 - \rho)\rho^1 = (0.6)(0.4) = 0.24$$

8. In a departmental store one cashier is there to serve the customers and the customers pick up their needs by themselves. The arrival rate is 9 customers for every 10 minutes and the cashier can serve10 customers in 5 minutes. Assuming Poisson arrival rate and exponential distribution for service rate, find:

(a) Average number of customers in the system.

(b) Average number of customers in the queue or average queue length.

(c) Average time a customer spends in the system.

(d) Average time a customer waits before being served.

$$Solution: \text{ Arrival rate}(\lambda) = 9\ per\ 10 \min utes = \frac{9}{10} = 0.9\ per \min$$

$$Service\ rate(\mu) = 10\ per\ 5 \min utes = \frac{10}{5} = 2\ per \min$$

$$\text{Traffic intesity}(\rho) = \frac{\lambda}{\mu} = 0.45$$

(a) The average number of customers in the system is given by

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{0.9}{2 - 0.9} = 0.81\ per\ \min$$

(b) Average number of customers in the queue or average queue length is given by

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{(0.9)^2}{2(2 - 0.9)} = 0.36\ per\ \min$$

(c) Average time a customer spends in the system is given by

$$w_s = \frac{1}{\mu - \lambda} = \frac{1}{2 - 0.9} = 0.909 \ \ per \ \min$$

(d) Average time a customer waits before being served is given by

$$w_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{0.9}{2(2 - 0.9)} = 0.409 \ \ per \ \min$$

9. In a railway marshaling yard, goods train arrives at the rate of 30 trains/day. Assuming that the inter-arrival time follows an exponential distribution and the service time(time taken to hump a train) distribution is also exponential with an average 36 min. Calculate the following :

(i) Average number of trains in the yard.

(ii) The probability that the queue size exceeds 9.

(iii) Average number of trains in the queue.

(iv) Estimate the fraction of a day that the marshalling yard will be busy.

Solution: Arrival rate $(\lambda) = 30$ *per day*

avg service time=36 minutes

$$\text{Service rate}(\mu) = \frac{60}{36}\text{ per hour} = \frac{60}{36}\times 24\text{ per day} = 40\text{ per day}$$

(i) Average number of trains in the yard is given by

$$L_s = \frac{\lambda}{\mu-\lambda} = \frac{30}{40-30} = 3\text{ trains}$$

(ii) P(queue size exceeds 9) = $P(queue \geq 10) = \rho^{10} = \left(\frac{\lambda}{\mu}\right)^{10} = 0.056$

(iii) Average number of trains in the queue is given by

$$L_n = \frac{\mu}{\mu - \lambda} = \frac{40}{40 - 30} = 4 \text{ trains}$$

(iv) The fraction of a day that the marshalling yard will be busy is given by

$$\text{traffic intensity}(\rho) = \frac{\lambda}{\mu} = 0.75$$

At Bharat petrol pump, customers arrive according to a Poisson process with an average time of 5 minutes between arrivals. The service time is exponentially distributed with mean time equal to 2 minutes. On the basis of this information, find out:

(i) What would be the average number of customers in the queuing system?

(ii) What is the average time spent by a car in the petrol pump?

(iii) What is the average waiting time of a car before receiving petrol?

Solution: Given Interarrival time = 5 minutes

$$\Rightarrow \text{Arrival rate}(\lambda) = \frac{1}{5} \text{ per minute}$$

$$\text{Service time} = 2 \text{ minutes}$$

$$\Rightarrow \text{Service rate}(\mu) = \frac{1}{2} \text{ per minute}$$

Average number of customers in the queuing system:

$$L_s = \frac{\lambda}{\mu - \lambda} = 0.666 \sim 1$$

Average time spent by a car in the petrol pump:

$$W_s = \frac{1}{\mu - \lambda} = 3.33 \min$$

Average waiting time of a car before receiving petrol:

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = 1.33 \min$$

Consider a switch which has an infinite buffer and an infinite number of users generating messages according to a Poisson process with average inter-arrival time of 800 milliseconds. The switch serves requests with a service time that is exponentially distributed with an average service time of 500 milliseconds. What is the average waiting time? Suppose that the switch is upgraded to reduce average service time to 400 milliseconds. How would that affect the average waiting time?

Solution: Given Interarrival time = 800 milliseconds

$$\Rightarrow \text{Arrival rate}\left(\lambda\right) = \frac{1}{800} \text{ per milliseonds}$$

$$\text{Service time} = 500 \text{ milliseconds}$$

$$\Rightarrow \text{Service rate}(\mu) = \frac{1}{500} \text{ per milliseonds}$$

Average waiting time : $W_s = \dfrac{1}{\mu - \lambda} = 1333.33$

switch is upgraded to reduce average service time to 400 milliseconds.

$$\Rightarrow \text{Service rate}(\mu_1) = \frac{1}{400} \text{ per milliseonds}$$

Average waiting time : $W_s = \dfrac{1}{\mu - \lambda} = 800$

Average waiting time is affected by is

$= 1333.33 - 800 = 533.33$

Arrivals of machinists at a tool crib are considered to be Poisson distributed at an average rate 6 per hour. The length of time the machinists must remain at the tool crib is exponentially distributed with average time of 0.05 hours.

a) What is the probability that a machinist arriving at the tool crib will have to wait?

b) What is the average number of machinists at the tool crib?

c) The company will install a second tool crib when convinced that a machinist would have to spend 6 minutes in waiting and being served at the tool crib. At what rate should the arrival of machinists to the tool crib increase to justify the addition of a second crib?

Chhabra Saree Emporium has a single cashier. During the rush hours, customers arrive at the rate of 10 per hour. The average number of customers that can be processed by the cashier is 12 per hour. On the basis of this information, find the following:

(i) Probability that the cashier is idle

(ii) Average time a customer spends in the queue

(iii) Average number of customers in the queue

On an average 96 patients per 24-hour day require the service of an emergency clinic. Also on an average, a patient requires 10 minutes of active attention. Assume that the facility can handle only one emergency at a time. Suppose that it costs the clinic Rs 100 per patient treated to obtain an average servicing time of 10 minutes, and that each minutes of decrease in this average time would cost Rs. 10 per patient treated. How much would have to be budgeted by the clinic to decrease the average size of the queue from one and one-third patients to half patients.