

## Applications of Queueing Systems

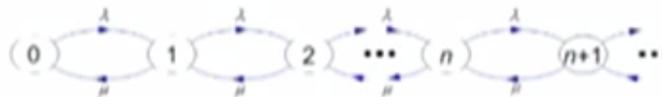
- Analyzing Network delays.
- Telephone conversations.
- Aircraft landing problems.
- Barber's shop



There are many applications of queueing system. We are going to discuss the abstract queueing system in the further lecture.

## M/M/1 Queueing Model

- Arrival process: Poisson Process with rate  $\lambda$ .
- Service times: Exponential with parameter  $\mu$
- Service times and inter-arrival times are independent
- Single server
- Infinite capacity in a system
- $N(t)$ : Number of customers in a system at time  $t$  (state)



State transition diagram



The easiest or the simplest queuing model that is a Markovian queuing model that is MM1 queuing model. Later we are going to correlate with the birth-death process also. In the MM 1 queuing model the inter-arrival time is exponentially distributed. As I discuss the Poisson process in the previous lecture whenever you have arrival follows a Poisson process then the inter-arrival time follows exponential distribution and are independent also. So here the first information is a arrival process follows a Poisson process with the intensity or rate  $\lambda$ . That means the inter-arrival times are independent and each one is exponentially distributed with the parameter  $\lambda$ .

The second information that is service time. Service times are exponentially distributed with the parameter  $\mu$  and the service times are independent for each customers and that's also independent with the arrival process. That means there is no dependency over the arrival pattern with the service pattern. Service times and the inter-arrival times are independent.

Then the third information only one server in the system. That's a queuing system in which only one server and the fourth information is missing that means it is a default it's infinite capacity order. Infinite capacity model. Now our interest is to find out the behavior of a queuing system or the behavior of number of customers in the system at any time  $t$ . Therefore you can define a random variable  $N$  of  $t$  that is nothing but the number of customers in the system at time  $t$ . Therefore this is going to follow- form a stochastic process over the  $t$ . Since the inter-arrival time is exponentially distributed and the service times are exponentially distributed, the memoryless property is going to be satisfied throughout all the time. Therefore this stochastic process there is a discrete state continuous time stochastic process satisfying the Markov property. Therefore this is a Markov process. Since our inter-arrival time is exponentially distributed and the service time is exponentially distributed and both are independent and the service time is also independent for each customers. Therefore, this stochastic process satisfies the memoryless property at all time points. Therefore, this discrete state because the possible values of  $N$  of  $t$  since it is a number of customers the possibilities are 0, 1, 2, and so on countably infinite. Therefore it's a discrete state and you are observing the system over that time therefore it is a continuous string. Therefore this stochastic process is a discrete state continuous time stochastic process satisfying the Markov property based on these assumptions. Therefore  $N$  of  $t$  is a Markov process.


Since the state space is a discrete therefore this is a Markov chain. Therefore this is a continuous time Markov chain. Therefore  $N$  of  $t$  is CTMC. So one can write the state transition diagram for this CTMC. That means the possible states are 0, 1, 2 and so on so this will form a notes and you try to find out what is the rate in which the system is moving from one state to other state. Since it is a MM1 queue model, queuing model therefore whenever the system is in the state 0 by the inter-arrival time which is exponentially distributed the number of customers in the system will be incremented by 1. Therefore that rate will be  $\lambda$  or the system moving from the state 0 to 1 it spends exponentially distributed amount of time here before moving into the state 1. Once the system come to the state 1 either one more arrival is possible or the customer who is under service that service could have been finished. Therefore, the service time is exponentially distributed with the parameter  $\mu$  therefore the system goes from the state 1 to 0 with the parameter  $\mu$ .

Similarly from 1 to 2 because of the inter-arrival time is exponentially distributed with the parameter  $\lambda$  therefore this is  $\lambda$ . Since the arrival follows a Poisson process in a very small interval of time only one customer is possible with the probability  $\lambda \Delta t$  and

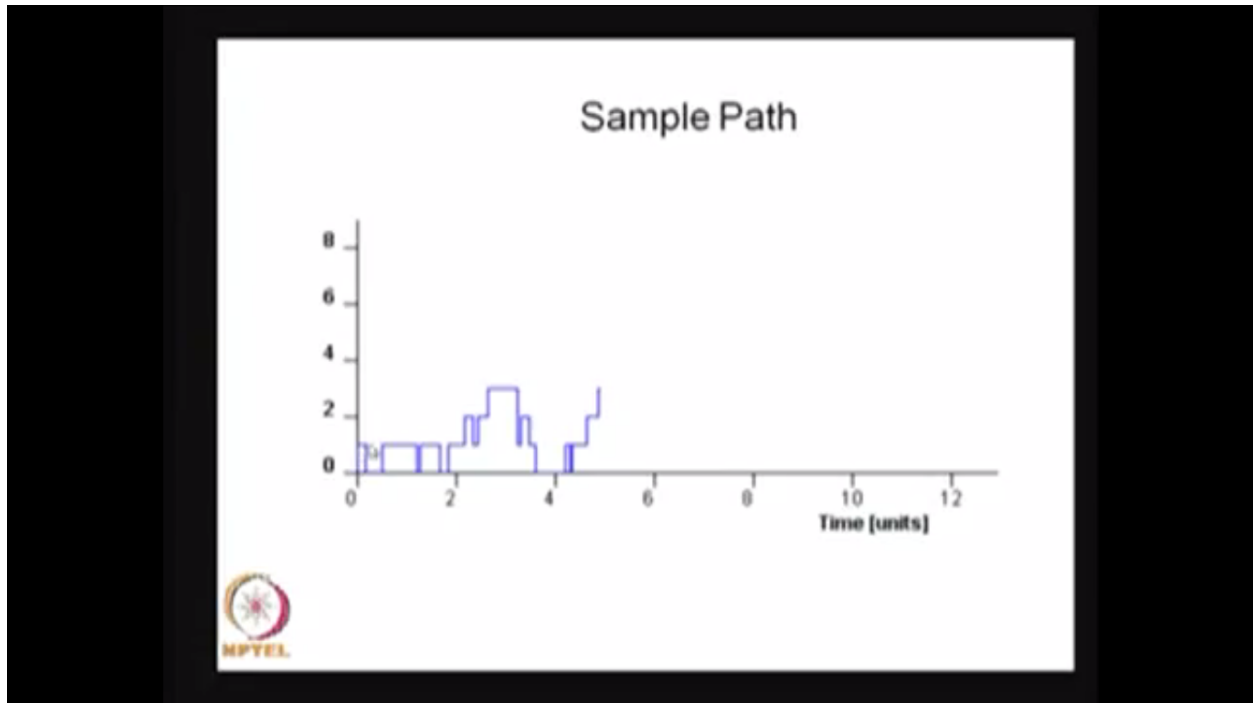
so on. Therefore there is no way the system goes from one state to jump into more than one state. That is not possible forward. So only one step forward is possible because of the arrival process follows a Poisson process and since we have only one server in the system this system also decremented by only one level below. Therefore, this is going to form but death-process. The reason for this CTMC going to be a birth-death process because of the arrival process follows a Poisson process. So whatever the assumptions we have it for the Poisson process that is going to be satisfied and since we have only one server in the system and he does the service for only one customer at a time after finishing that server – after finishing the customer service then it move into the next service immediately and so on if the customers are available in the queue. Therefore, the system goes to the one state below by only one only. It won't move from 2 to 0 or 3 to 1 and so on. Therefore this CTMC is a birth-death process. Therefore, I'm connecting the CTMC with the MM1 queue in particular in the CTMC the birth-death process because of the transitions due to arrival or departure of a customer and only nearest neighbors transitions are allowed because of the assumptions which we have made.

### CTMC Formulation

- Transitions due to arrival or departure of customers
- Only nearest neighbors transitions are allowed.
- State of the process at time  $t$ :  $N(t) = i$  ( $i \geq 0$ ).
- $\{N(t); t \geq 0\}$  is a continuous-time Markov chain with

$$\begin{aligned}
 q_{i,i+1} &= \lambda \\
 q_{i,i-1} &= \mu \\
 q_i &= -(\lambda + \mu) \\
 q_{i,j} &= 0 \quad \text{for } |i - j| > 1
 \end{aligned}$$


Therefore this is going to a continuous time Markov chain with the rate in which the system moves so from the state  $i$  to  $i$  plus 1 that rate is lambda and the system moves from the state  $i$  to  $i$  minus 1 that rate is mu and all other rates are going to be 0 other than the diagonal element and these rates also constraint, not state dependent rates. Therefore this is a birth-death process with the birth rates lambda under death rates mu.



So this is a sample path. Suppose at time 0 one customer in the system then it services over. Then the second customer enter into the system. Now the number of customers in the system is one and so on so that means that this duration is the service time for the first customer and from this point to this point that is a inter-arrival time of the second customer entering into the system and from this time point to this time point that is the service time for the second customer which is independent of the service time for the first customer and this is the time point second customer enter and this is a time point in which the third customer enters. Therefore the inter-arrival time is from this point to this point and so on. So this is the dynamics of a number of customers in the system over the time. Therefore, this stochastic process is a discrete state continuous time stochastic process [Indiscernible] [00:09:19] the Markov property therefore this is a continuous time Markov chain. So later I am going to simulate the MM1 queuing model using some simulation technique.

So the conclusion is the underlying stochastic process for the MM1 queuing model is a birth-death process. The  $N$  of  $t$  is a stochastic process. So this stochastic process is a birth-death process. Therefore, now we are going to discuss the stationary distribution, time dependent probabilities and so on. So how to find the stationary distribution. Solve  $\pi Q = 0$ .  $\pi$  is the vector consists of a  $\pi_i$ 's where  $\pi_i$ 's are nothing but what is the probability that  $N$  customers in the system what is the probability that  $i$  customers in the system in a long run. So that long-run is defined in this way the  $N$  of  $t$  is a stochastic process as  $t$  tends to infinity the number of customers in the system in a long run that is going to be the  $N \pi_i$  is nothing but a probability that  $N_i$  customers in the system in a long run.

## Stationary Distribution

$$\pi = (\pi_0, \pi_1, \dots) ; \pi_i \geq 0 ; \sum_i \pi_i = 1$$

$$\pi Q = 0$$

$$0 = -\lambda \pi_0 + \mu \pi_1$$

$$0 = \lambda \pi_{i-1} - (\lambda + \mu) \pi_i + \mu \pi_{i+1}, \quad i \geq 1$$

$$\pi_1 = \frac{\lambda}{\mu} \pi_0$$

$$\pi_{i-1} = \frac{\lambda}{\mu} \pi_i = \frac{\lambda^{i-1}}{\mu^{i-1}} \pi_0 ; i = 1, 2, \dots$$

$$\pi_i = P[N=i]$$

$$N = \sum_{k=0}^{\infty} n(k)$$



So now we are going to solve  $\pi Q = 0$  with the normalized equation summation of  $\pi_i$  is equal to 1. So once you frame the equation you will get a  $\pi_1$  in terms of  $\pi_0$  and the  $\pi_1$  minus 1 in terms of first  $\pi_i$  then substitute recursively you will get in terms of  $\pi_0$ . So since it is a homogeneous equation we will get all  $\pi_i$ 's in terms of  $\pi_0$ . So use the normalizing equation summation of  $\pi_i$  is equal to 1 you will get  $\pi_0$ . So the  $\pi_0$  is equal to  $1 - \rho$  where  $\rho$  is  $\lambda / \mu$  and since I am relating this stochastic process with the birth-death process with the infinite capacity if you recall the stationary distribution exists as long as the denominator of  $\pi_0$  that series converges. So that will converge only if  $\lambda / \mu$  is less than 1. If  $\lambda / \mu$  being is greater or equal to 1 then that denominator diverges accordingly you won't get the stationary distribution. So to have a stationary distribution you need a  $\rho$  has to be less than 1. That also you can intuitively say whenever the system is stable that is corresponding to  $\rho$  is less than 1 in that way you have a stationary distribution that means in a longer run this is a proportion of the time the system will be empty and the  $\pi_n$  is nothing but the  $n$  customers in the system in a longer run that is  $1 - \rho$  times  $\rho^n$  where  $\rho$  is less than 1. This  $\rho$  can be visualized as the offered load also because the  $\rho$  is nothing but the mean arrival rate and the  $\mu$  is a mean service rate and this ratio will give the offered load and  $1 - \pi_0$  that is the probability that the system is non-empty and that is nothing but the server utilization. Server utilization is nothing but what is the probability that the server is busy. The server will be busy as long as the system is not empty. So the  $\rho$  is the server utilization that can be obtained in the – from this one and in the longer run the server utilization is  $\rho$ .