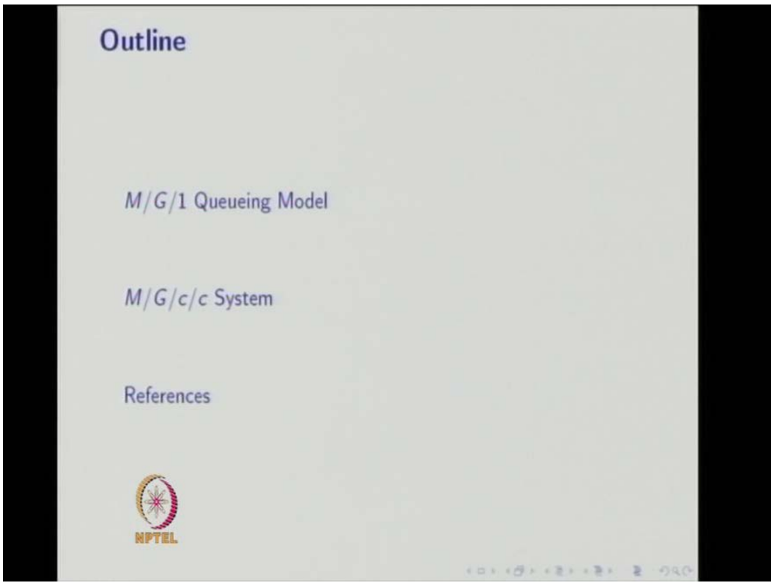
Stochastic Processes
Prof. Dr. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Module - 8
Renewal Processes
Lecture - 4
Non Markovian Queues

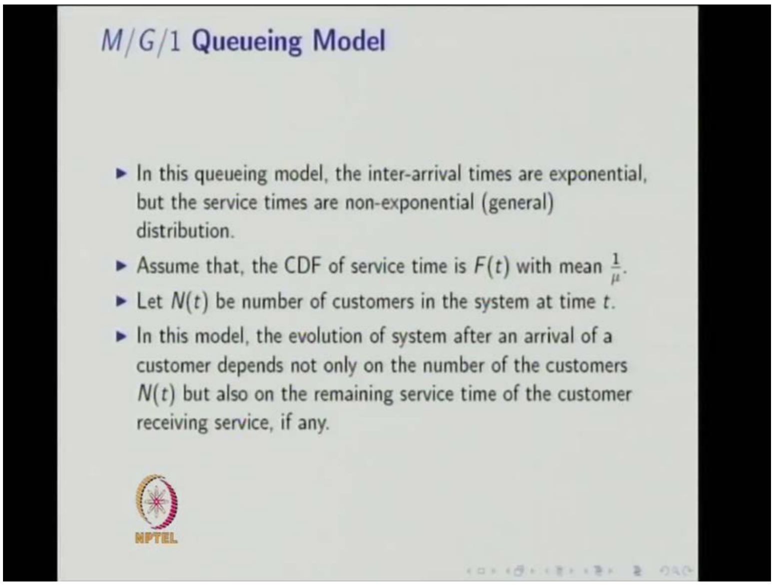
This is stochastic processes, module 8 renewal processes. In the first 3 lectures, we have discussed the renewal processes and its properties. Then later, we have discussed the important limiting theorem on renewal processes. Then, we have discussed the Markov renewal process and Markov region rating process.

(Refer Slide Time: 01:05)



In between we have discussed the reward renewal process also. In this lecture we are going to cover the non Markovian queues with respect to the service time distribution. In particular, we are going to discuss the M G 1 queue and M G c c system.

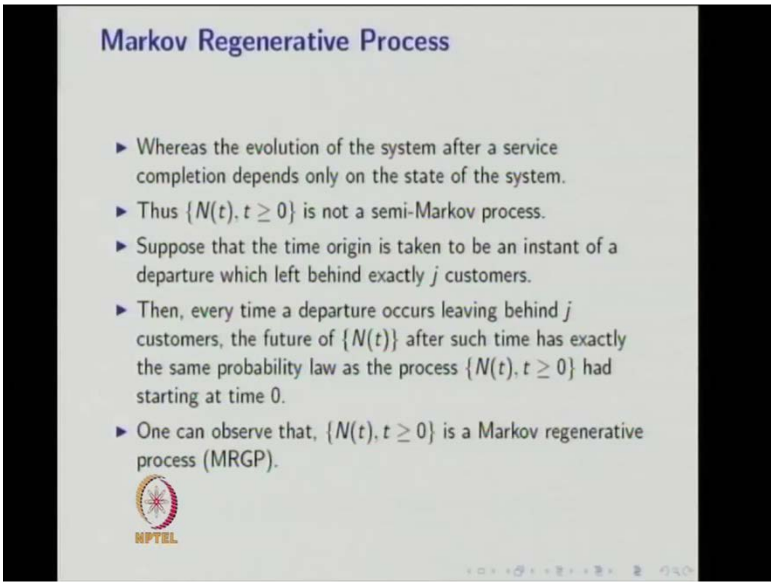
(Refer Slide Time: 01:30)



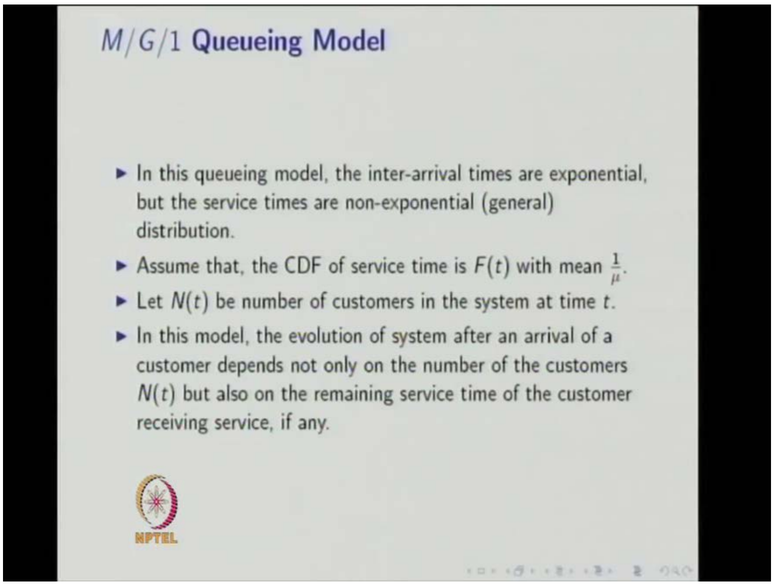
What is M G 1 queue? In this queuing model the inter arrival times are exponential, but the service times are non exponential, nothing but general distribution. Assume that the CDF of the service time is F of t with the mean 1 by mu. So, suppose you assume that service time is exponential distribution with the mean 1 by mu, then it is a M M 1 queuing model. Let N (t) be the number of customers in the system at time t. In this model, the evaluation of the system after the arrival of customer depends not only on the number of customers N of t, but also the remaining service time of the customers receiving service.

The remaining service is nothing but this is the service times are random variable, the remaining service time is elapsed service or remaining or residual service time. Whenever the service times are exponential distribution, the remaining time or elapsed service time or residual service time all are exponential distributions. Whenever the service times are not a exponential distribution, then the remaining service time will be some other distribution. The evaluation of the system after the service completion depends only on the system of the state. In this queuing model the system changes the state based on two types of transition epochs; one is arrival epochs and other one is service epochs.

(Refer Slide Time: 03:00)

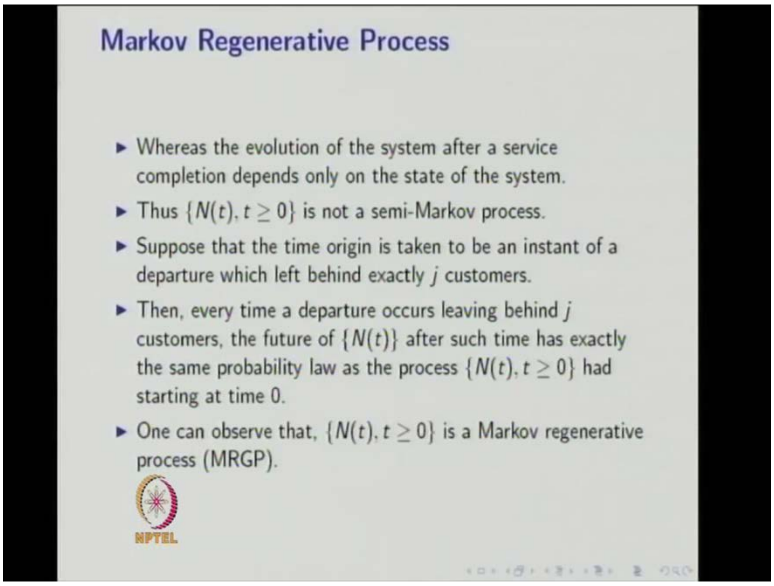


(Refer Slide Time: 03:22)



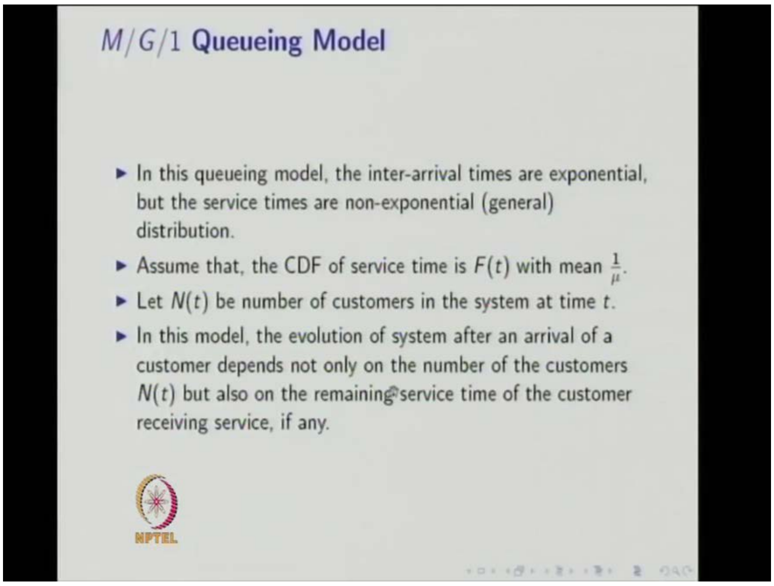
First, we are discussing the arrival scenario the evaluation of the system after the arrival of a customer depends not only the number of customers N of t, but also the remaining service time.

(Refer Slide Time: 03:37)



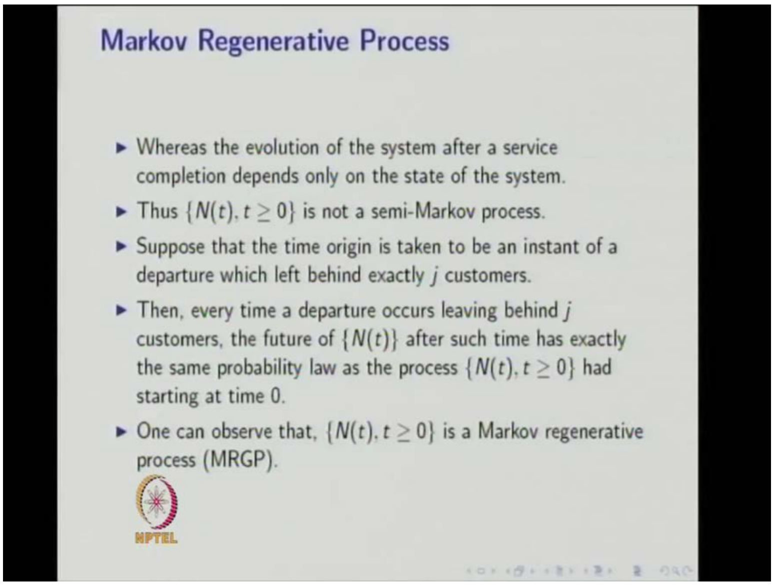
Whereas, the evaluation of the system after service completion depends only on the state of the system. Therefore, N of t is not a semi Markov process because it depends on the remaining service time.

(Refer Slide Time: 04:06)



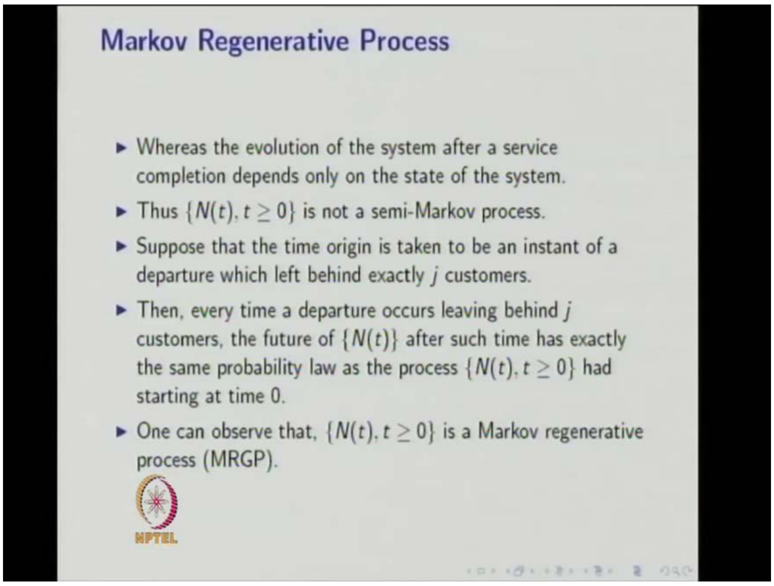
Since the, since the after the arrival of a customer the remaining service time also play role.

(Refer Slide Time: 04:12)



Whereas, after the service completion, it depends only on the state. If you recall the definition of semi Markov process, the memory less property will be satisfied at all time transition instance. The memory less property will be satisfied at all time transition instance. Then only the stochastic process is called as semi Markov process.

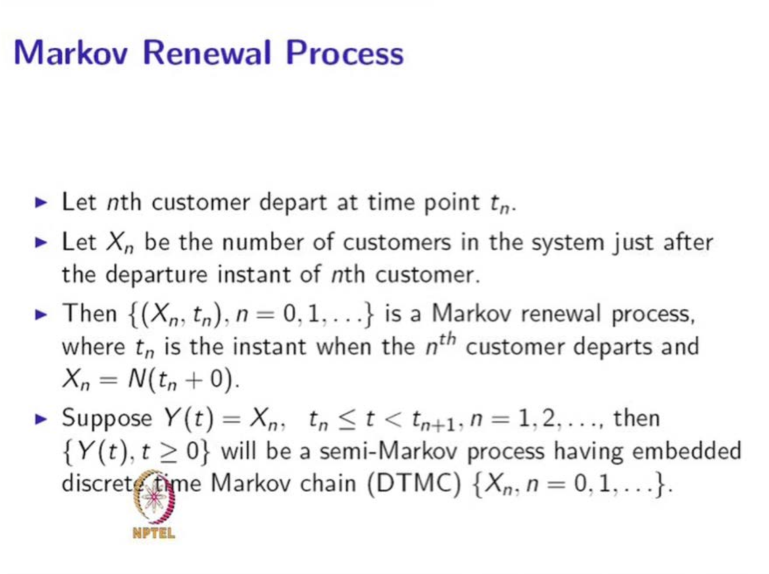
(Refer Slide Time: 04:44)



Therefore, here the N of t is not a semi Markov process. Suppose, the time of time origin is taken to be an instant of departure, which left behind exactly j customers, then every time the departure occurs leaving behind j customers, the future of N of t after such time has exactly the same probability law as the process had started starting at time 0. So, this is so called the probabilistic replica with the Markov property as well as time homogeneity.

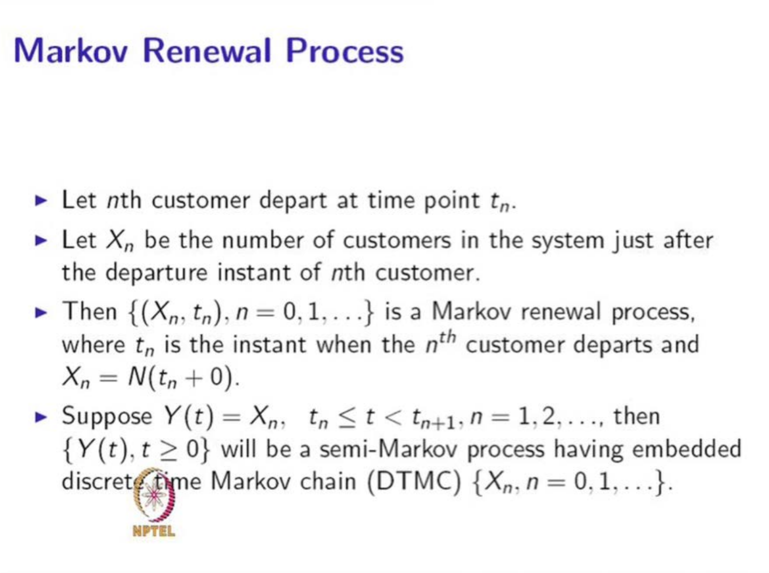
If you recall the definition of a Markov region rating process, a stochastic process has the property of probabilistic replica with the few time instance it has the Markov property with the time homogeneity with few states at few states. Therefore, N of t is a Markov region rating process. Please refer the previous lecture for understanding the definition of a Markov region rating process. So, by using the definition of Markov region rating process the N of t will be the Markov region rating process.

(Refer Slide Time: 06:12)



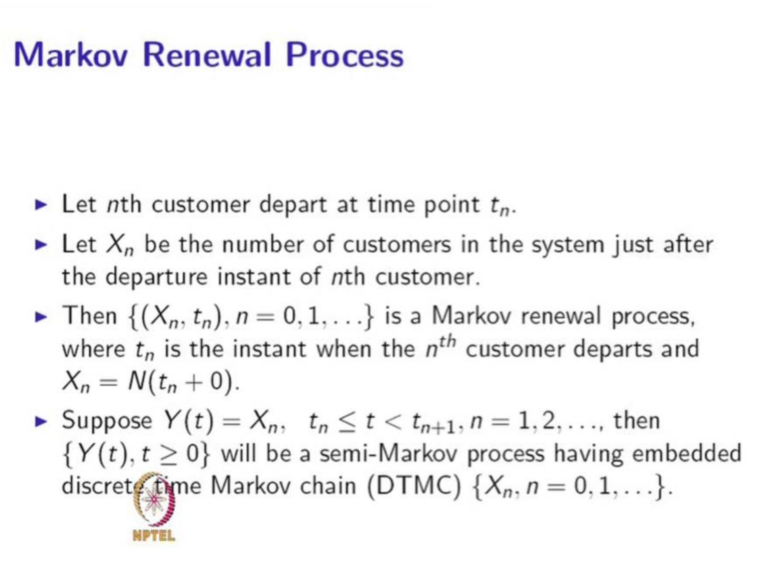
If you consider the time epochs are departure time epochs. Let N nth customer departure at time instant t N. Let X n be the number of customer in the system just after the departure instant of nth customer. That means X n is nothing but the N of t n plus 0 just after the departure instant of nth customer that will be treated as the random variable X n. Then X n comma t n will form a Markov renewal process. The t n are the time instant at the departure epochs not the arrival epochs; only the departure epochs that collection of time points along with x n form a Markov renewal process, where X n is n of t n plus 0.

(Refer Slide Time: 07:30)

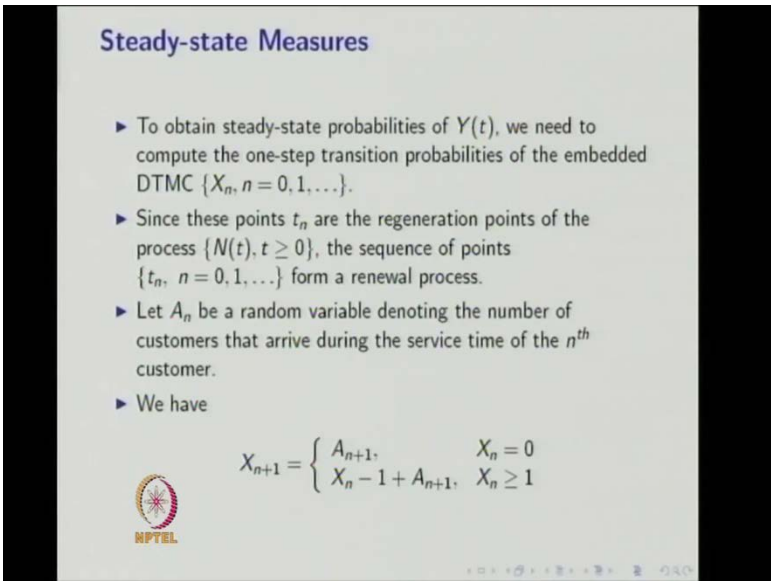


That in words it is the number of customers in the system just after the departure instant of nth customer leaves. We can create a semi Markov process by Y of t is equal to X n. If that value will be between t n to t n plus 1, if it is a same, then it will form a semi Markov process.

(Refer Slide Time: 08:11)



(Refer Slide Time: 08:20)

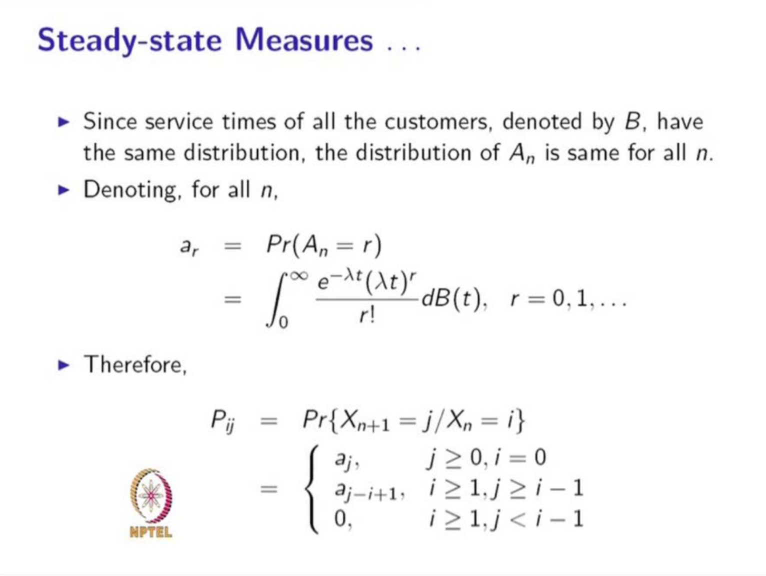


That means the system is not moving any any other states in between the time duration, then the Y of t will be a semi Markov process. As such N of t is not a semi Markov process. Whereas, N X n comma t n will form a Markov renewal process and the N of t will be a Markov region rating process.

With embedded the Y of t will be a semi Markov process, if I make Y of t is equal to X n, where t is lies between t n to t n plus 1 and having the embedded DTMC X n. Our interest is to find out the steady state measures. To obtain the steady state probability of Y t, we need the one step transition probabilities of embedded DTMC. Suppose, these time points t n are the regeneration points, then the t n's will form a renewal process. Let an be the random variable denoting the number of customers, that arrive during the service time of nth customer. That means X n plus 1 will be A n plus 1, if X n was 0. If X n was greater than or equal to 1.

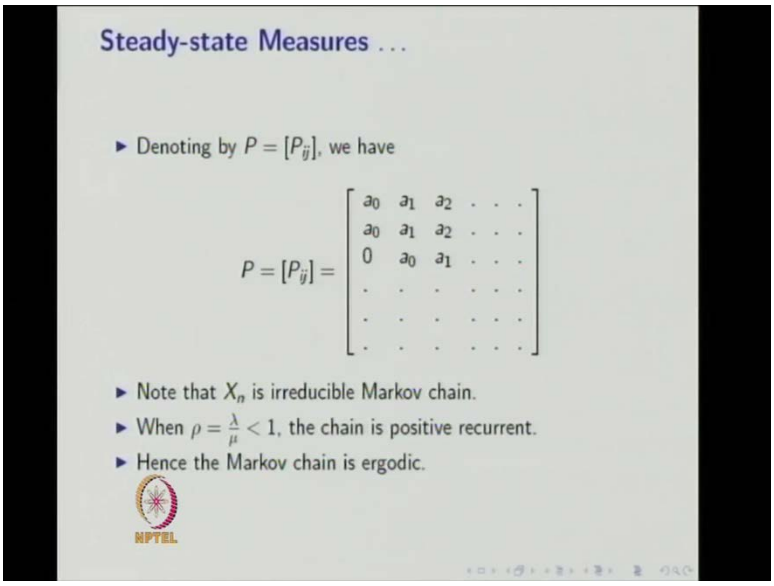
Then X n plus 1 will be X n minus 1 plus A n plus 1. That is nothing but how many customers in the system when the n plus 1 eth customer leaves, that is same as when the nth person leaves how many customers in the system minus 1 is for the n plus 1 eth customer himself. How many customers enter into the system during his service, that is A n plus 1, whenever the X n was greater than or equal to 1. If X n was 0, then the number of customers will be in the system then n plus 1 eth customers leaves will be same.

(Refer Slide Time: 10:08)



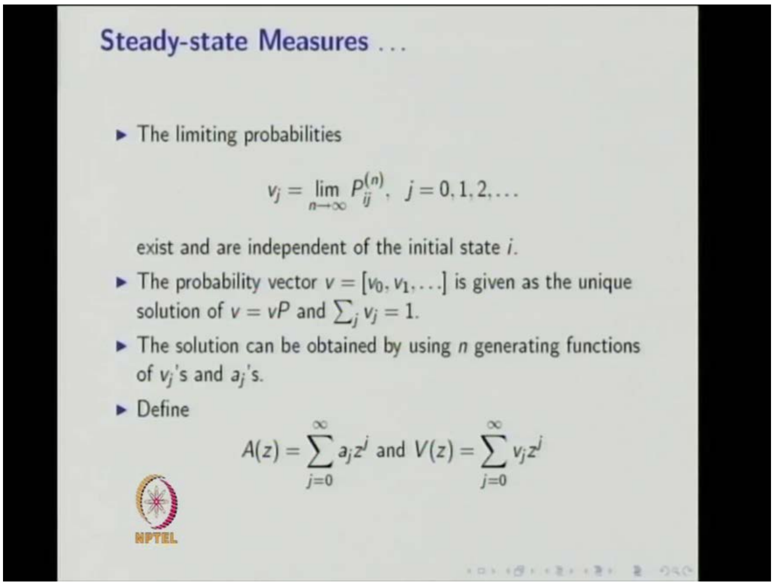
As number of customers who enter into the system during the during his service time. Since, the service times of all all the customers have the same distribution the distribution of AS n is same for all n. Therefore, you can go for, for all n, we can go for a r is the r customers enter into the system during the any customer. So, that is probability mass function of a A n. Now, you can find out the transition probability of system moving from the state i to j with respect to the stochastic process X n. X n is embedded Markov chain in the N of t. So, that will be a j, if j is greater than or equal to 0 and i equal to 0, similarly a j minus i plus 1 or 0, according to the values of i and j.

(Refer Slide Time: 11:20)



Now, you can make one step transition probability matrix, that is nothing but the P, the rho sums are going to be 1. Note that X n is irreducible Markov chain because each state is reachable by every other states. If we make rho, that is nothing but lambda by mu is less than 1, lambda is the parameter for the inter arrival time which is exponential distribution. So, whenever rho is less than 1, the chain is a positive recurrent. Rho is less than 1 is require because only then the main recurrence time will be finite. Hence, the Markov chain will be a positive recurrent.

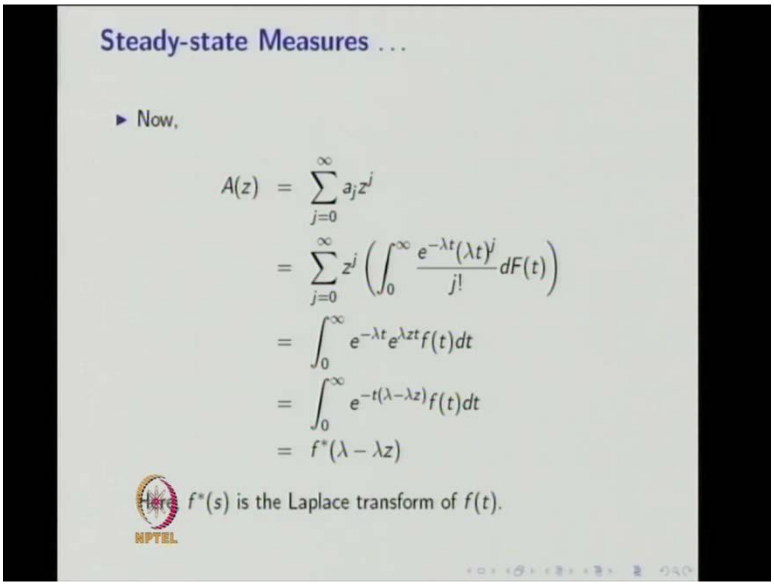
(Refer Slide Time: 12:22)



Therefore, it is ergodic. Irreducible positive recurrent gives the Markov chain is ergodic. Therefore, the Markov chain is the embedded Markov chain is ergodic.

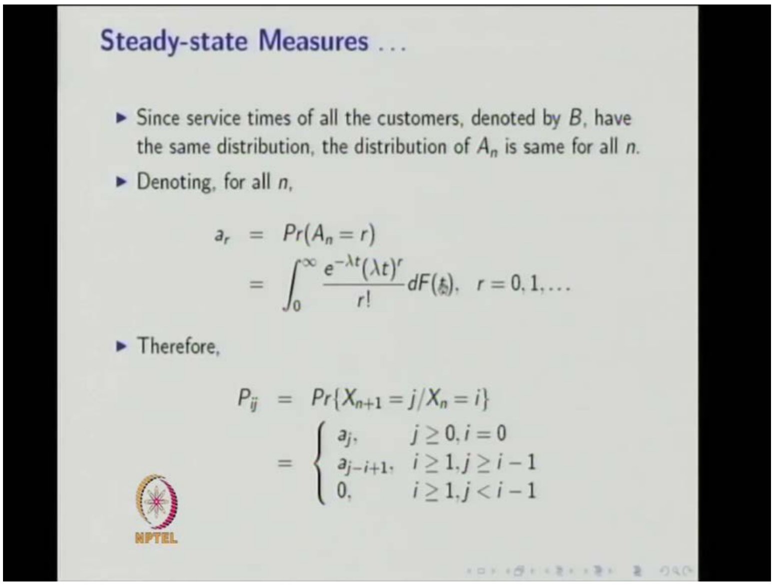
So, we can find out the limiting probabilities that will exist as it is independent of initial state i. You can find out the limiting probabilities by solving v is equal to v p and summation of v j is equal to 1. The solution can be obtained by using the probability generating function of v's and a's v's are probability vector steady state probability vector. a's are the number of customers arrived during the service of any customer.

(Refer Slide Time: 13:25)



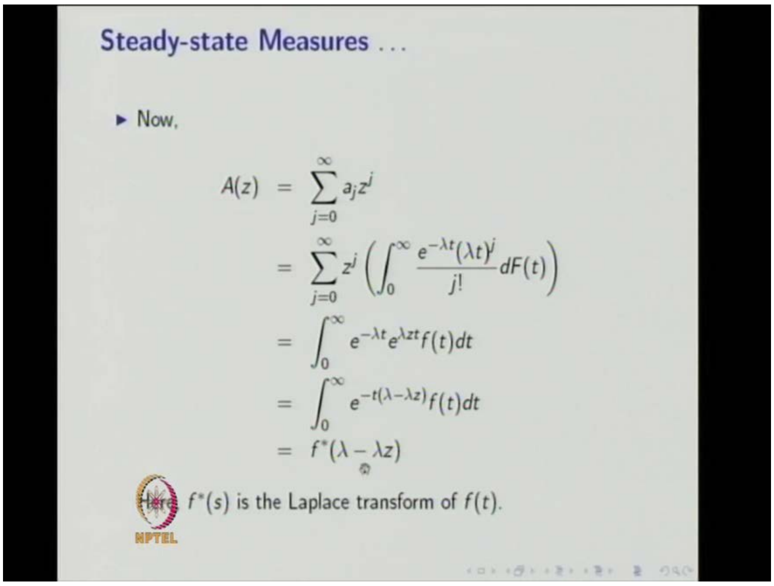
So, A z is nothing but the probability generating function for a j's and v j's v j v of z is nothing but probability generating function for the random variable for the probabilities v j's. Now, you can find out a of z in terms of the Laplace transform of the function F of t. Because A of z is nothing but the summation of a j into z power j and you can replace a j by the integration. Now, integration and summation can be interchanged and F of t is nothing but F of t is nothing but the the distribution of service time.

(Refer Slide Time: 14:00)



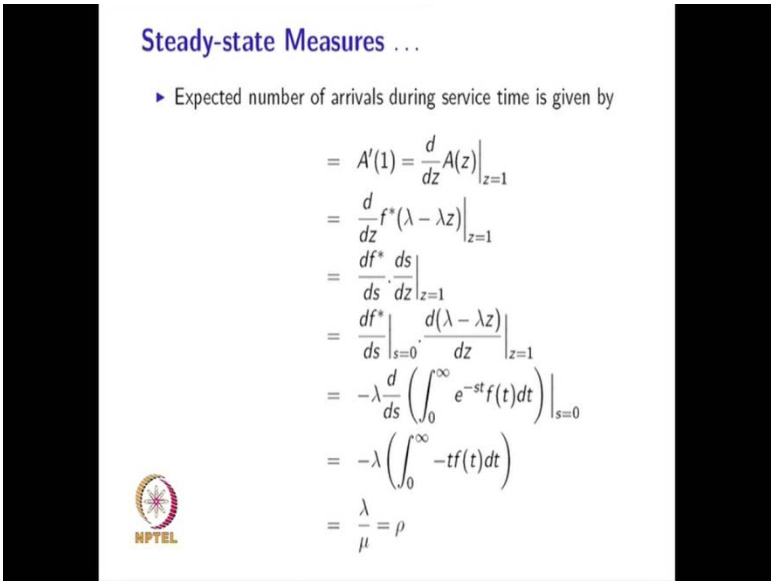
F of t is nothing but the distribution of service time, therefore the small f of t is a probability density function of the service time. So, the f star is nothing but the Laplace transform of F of t.

(Refer Slide Time: 14:20)



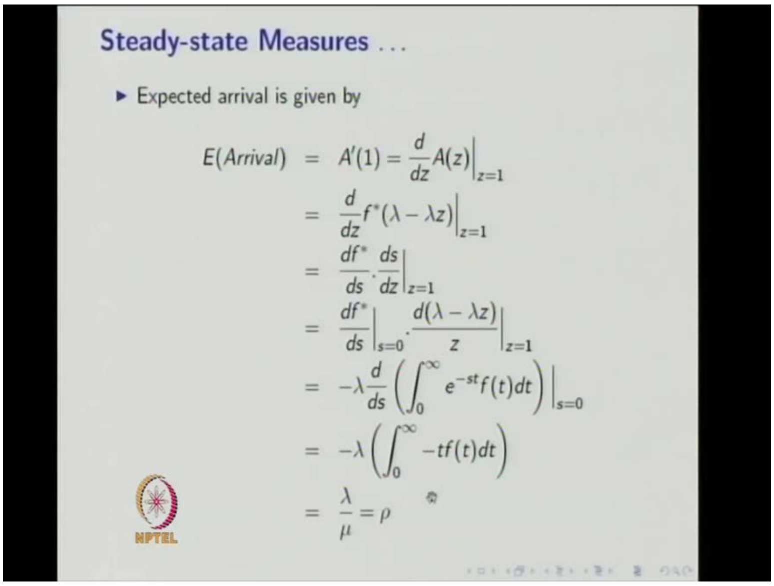
So, the probability generating function of a of j's is nothing but the Laplace transform of probability density function of a service time distribution.

(Refer Slide Time: 14:40)



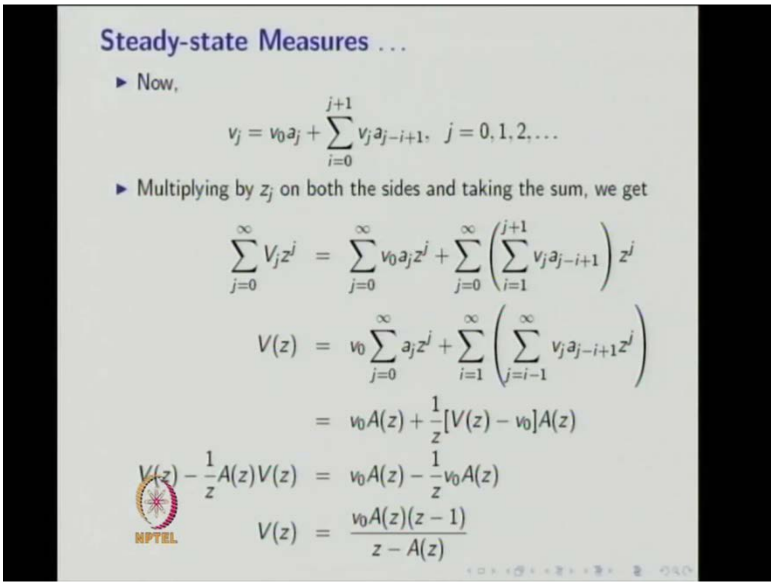
Now, we can find the expected number of arrivals during the service time that is nothing but since we got the probability generating function differentiate with respect to z and substitute z is equal to 1 will be the average number of arrivals during the any customers service time. So, you can simplify that by using the Chain rule.

(Refer Slide Time: 15:14)



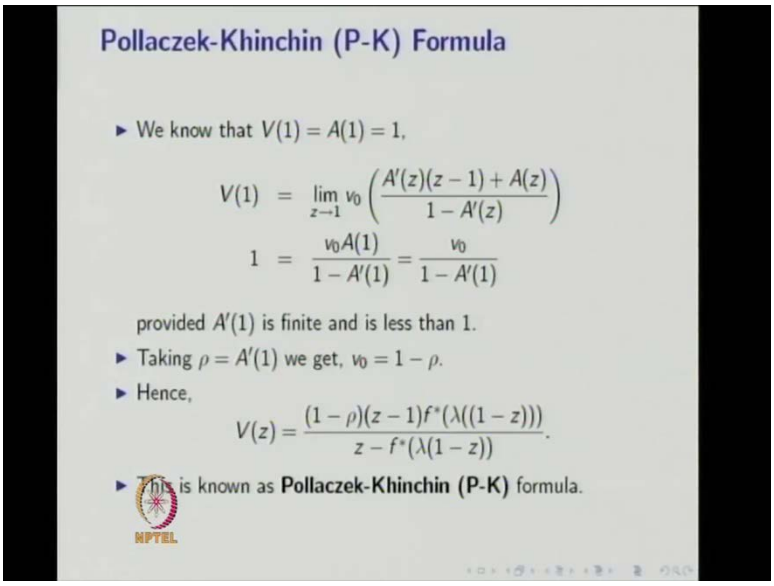
So, that will be nothing but lambda by mu where lambda by mu is nothing but the rho. So, the expected number of arrivals is rho. Rho is nothing but lambda by mu.

(Refer Slide Time: 15:25)



Our interest is to find out the steady state probability probabilities V j's. So, the V j's you can relate in terms of by expanding V is equal to V p you will get V j is equal to V naught a j's plus summation form. So, multiply z z power j in both sides and taking the summation, you will get after doing, after simplification we will get V z in terms of A z with V naught. V naught is no, V naught is the steady state probability of no customer.

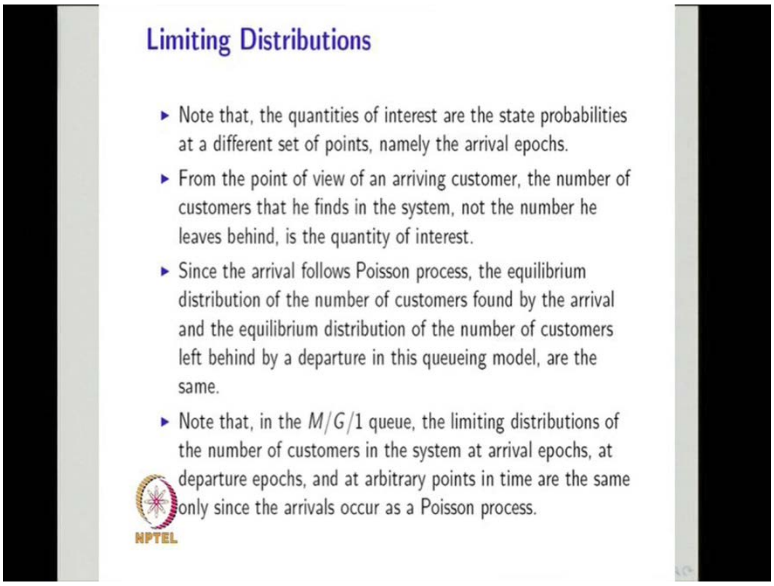
(Refer Slide Time: 16:15)



Sorry V z in terms of A z and V naught. You know that since V of z and A of z's are probability generating functions. So, v of 1 a of 12 are equal to be 1. So, using that relation, you can find out what is a value of V naught. So, V naught is nothing but 1 minus A dash of 1 and A dash of 1 is nothing but the expected arrivals. Just now we got expected arrival is equal to rho, therefore V naught is equal to 1 minus rho.

So, here we got V z in terms of A of z with V naught and already we we have the relation A of z is a Laplace transform of probability density function of service time distribution. Just now we got V naught in terms of 1 minus, V naught is equal to 1 minus rho. Therefore, the probability generating function of the steady state probabilities will be in terms of Laplace transform of probability density function of service time distribution with rho. This equation is known as Pollaczek Khinchin or PK formula. So, this is the formula used to find out the steady state probabilities because if you get the co efficient of z that is steady state probabilities.

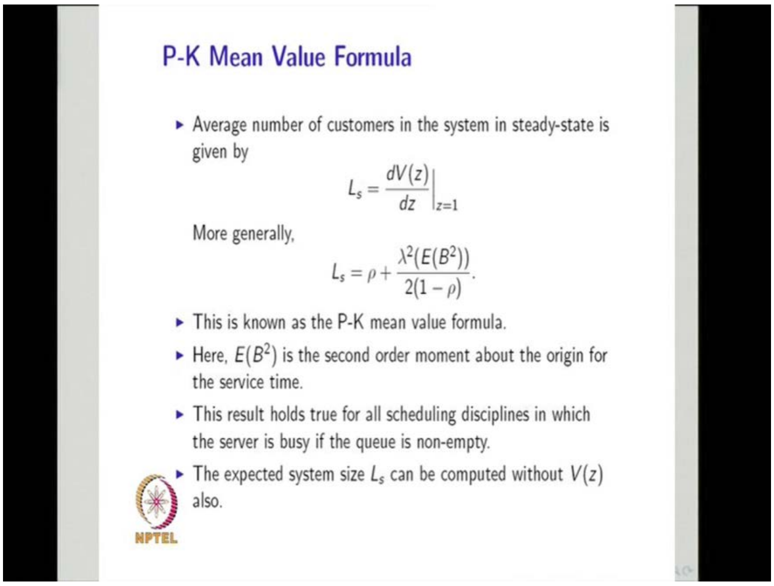
(Refer Slide Time: 17:45)



Note that the quantities of interest are the state probabilities at a different set of points, namely the arrival epochs from the past of view of arriving customer. The number of customers that he finds in the system not the number, he leaves behind is the quantity of interest. Since, the arrival follows Poisson process, the equilibrium distribution of the number of customers find found by the arrival. The equilibrium distribution of the number of customers left behind departure in this queuing model are the same.

Note that the the M G 1 the limiting distribution of the number of customers in the system at arrival epochs and the departure epochs and at the arbitrary time points are the same only, since the arrival occurs as a Poisson process. So, we found the limiting probabilities at the departure epochs, but since the arrival follows the Poisson process, the limiting distribution of number of customers in the system at the arrival epochs. The departure epochs and at the arbitrary time points all are same.

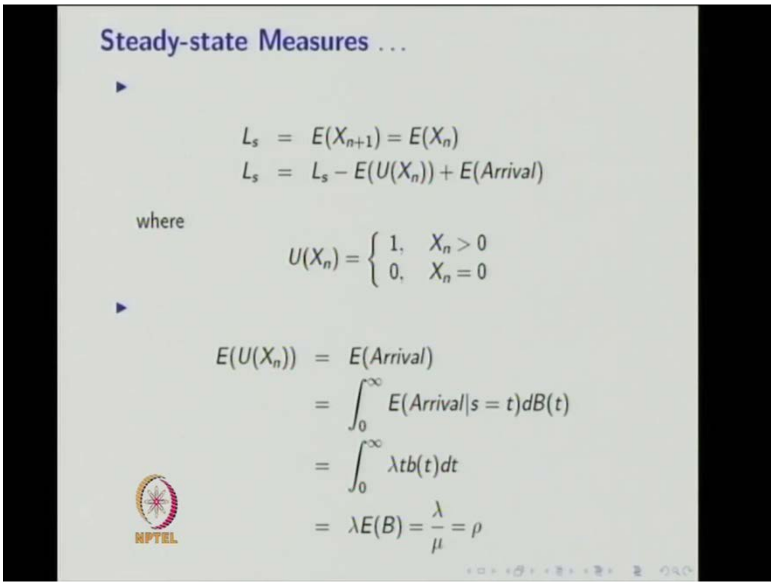
(Refer Slide Time: 19:07)



This is the standard result, which we shall be using without proof. Now, you can find out the average measures average number of customers in the system in steady state is given by differentiate the probability generating function. Substitute z is equal to 1 will be the average number of customers. If you do the simplification, you will get rho times rho plus lambda square expectation of the service time distribution whole square expectation of B square divided by 2 times 1 minus rho.

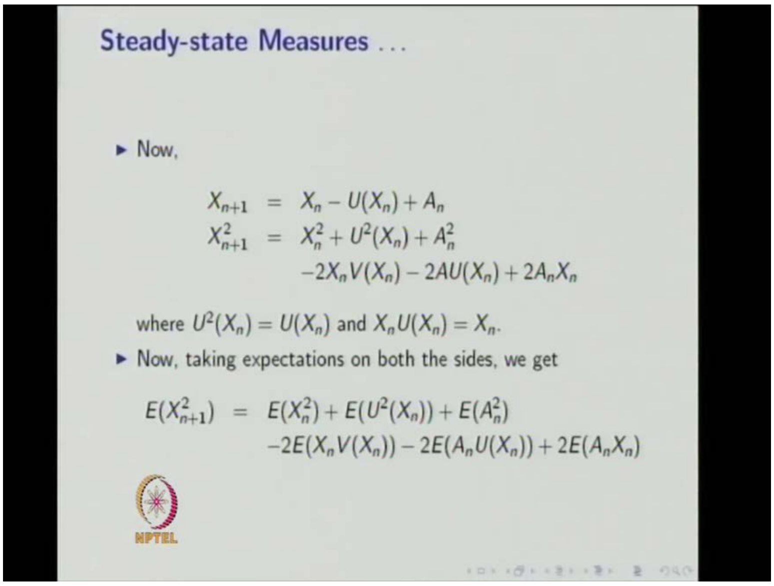
This equation is known as Pk mean formula. Previous one was the PK formula because that gives the steady state probabilities. Whereas, this gives the average measures, therefore this is called the PK mean formula. Here expectation of B square is the second order moment about the origin for service time. This result holds true for all schedule link discipline in which the server is busy if the queue is non empty.

(Refer Slide Time: 20:35)



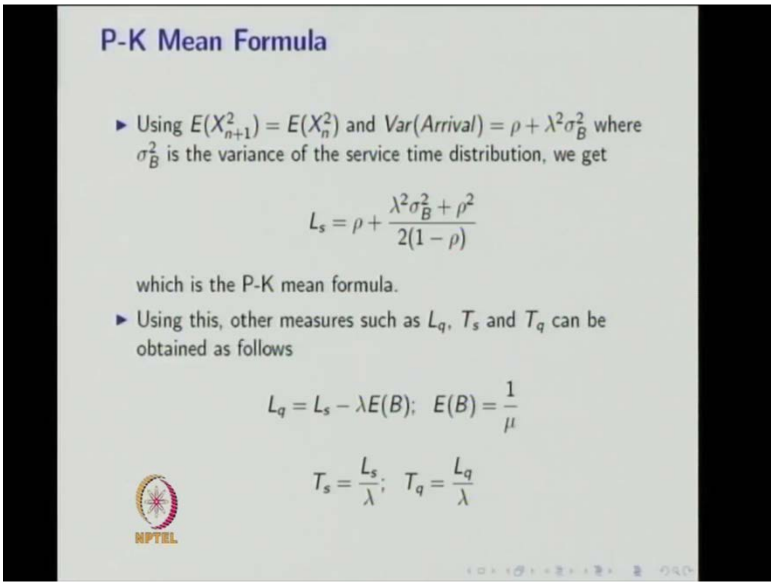
The expected system size L s can be completed without V z also. Because if you know if you know rho as well as if you know the expectation of B square you can find out the expected system size. Now, we are deriving the L s in a different way not via the PK formula. So, the derivation is as follows. So, the L s can be written in terms of expectation of X n plus 1 and you can find out expectation of U of X n using this you will get is equal to rho.

(Refer Slide Time: 21:05)



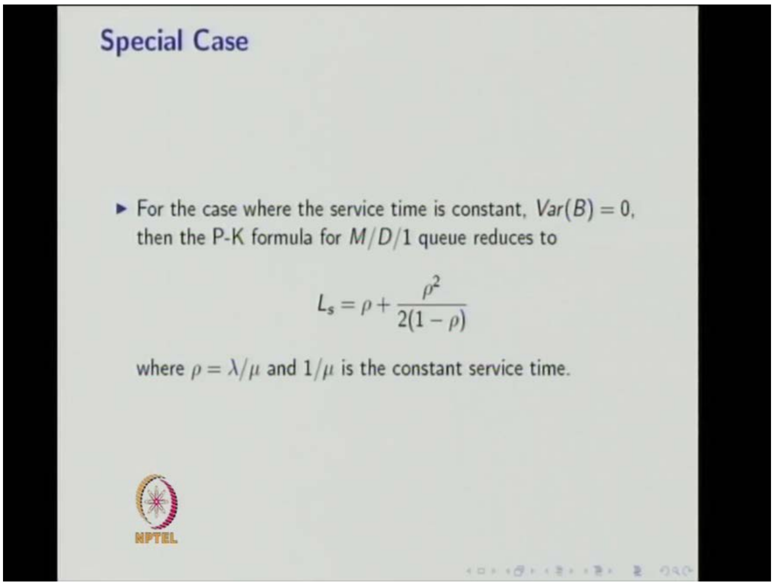
After you do the simplification, you can get the expectation of X n plus 1 square also in terms of expectation of X n square as well as expectation of A n square.

(Refer Slide Time: 21:19)



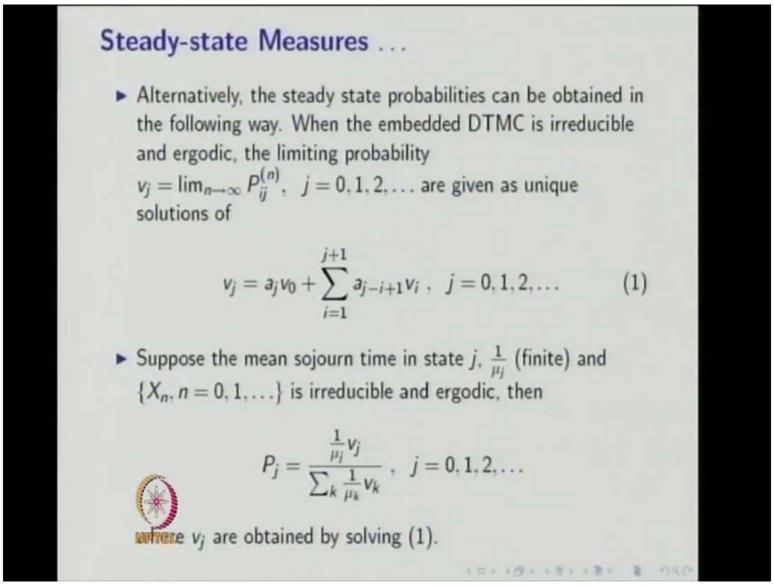
So, once you know the expectation of X n plus 1 square and variance of arrival where variance of arrival means the variance of number of arrivals during a service time. You can get the expected number of customers in the system in steady state. So, this is called the PK mean formula without using the PK formula. Once you know the L s that is average number of customers in the system, you can find out the L q, that is average number of customers in the queue. T s is nothing but total time spent in the system and the T q is, nothing but the sorry, T s is nothing but the average time spent in the system and the T q is average time spent in the queue. Using Little's formula, you can find the all other measures.

(Refer Slide Time: 22:17)



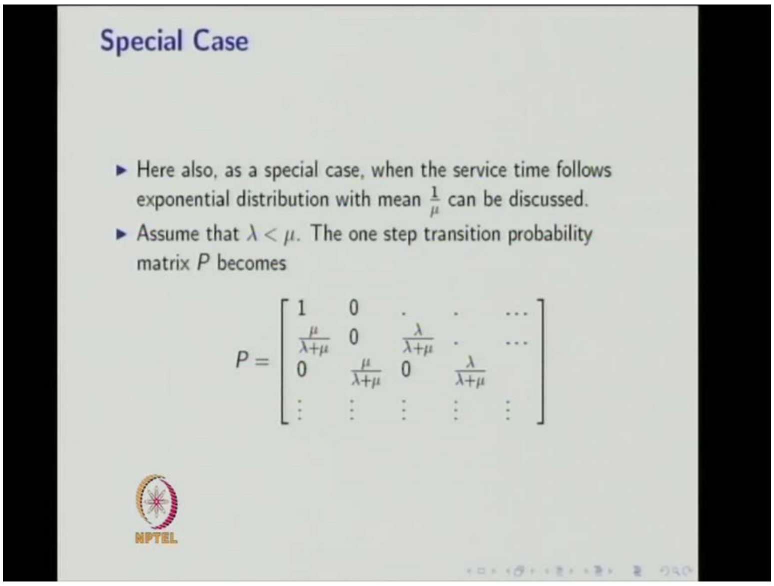
As a special case, if the variance is 0, variance of service time is 0 that means it is a M D 1 queue.

(Refer Slide Time: 22:48)



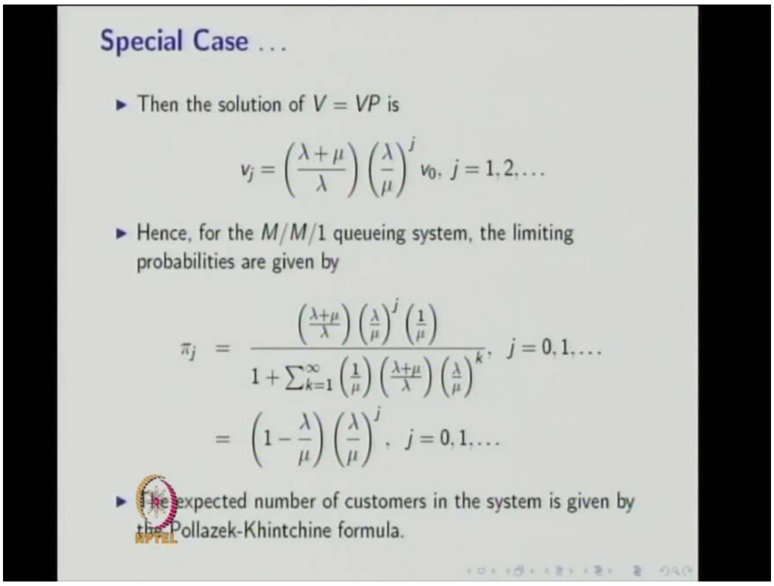
Then you can get the average number of customers in steady state will be rho plus rho square divided by 2 times 1 minus rho. Here the rho is nothing but lambda by mu where 1 by mu is a constant service time. Alternatively you can find out the steady state probabilities by solving V is equal to V p.

(Refer Slide Time: 23:31)



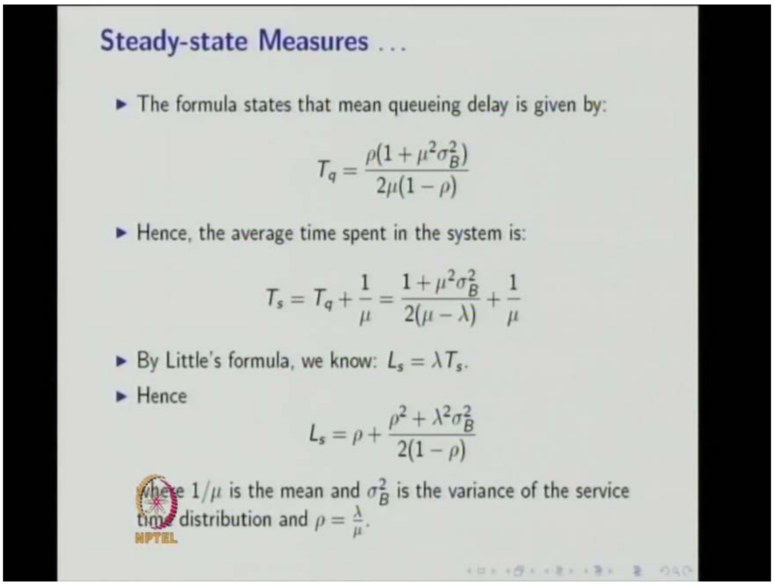
You can use the mean sojourn time that is 1 divided by mu j for the state j. Then you can find out the steady state probabilities P j's nothing but average sojourn time multiplied by the steady state probabilities of embedded Markov chain, the way we have done it in the semi Markov process.

(Refer Slide Time: 23:47)



As a special case, you can verify when the service time follows exponential distribution with mean 1 by mu you can get the 1 step transition probability matrix, in this form for the embedded Markov chain. Then you can solve V is equal to V p with the summation of V i is equal to 1, we will get V j's. Then if you substitute the V j's as well as substitute the V j's, then you can get the steady state probabilities for the M M 1 queue, that is same as one minus rho times rho power j.

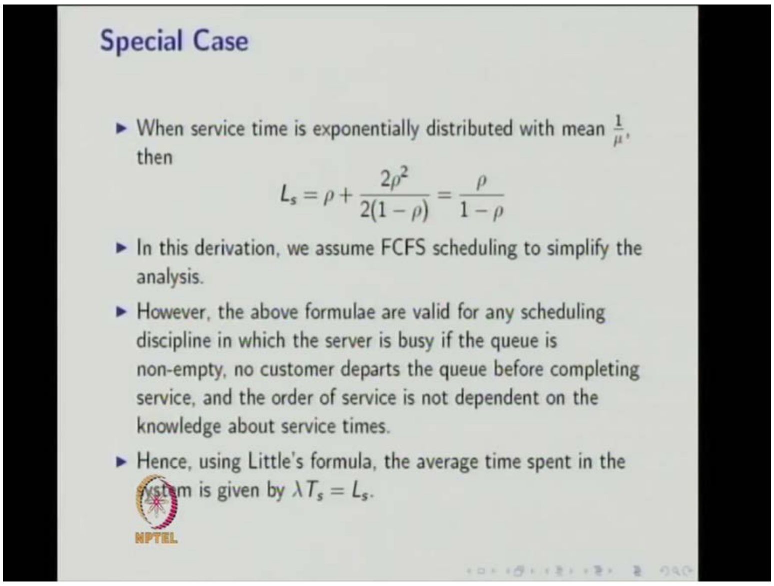
(Refer Slide Time: 24:22)



For any M G 1 queue, you can find out the average time spent in the queue. That is a mean queuing delay. Once you know the mean queuing delay, if you add the average time spent in the, you can get the average time spent in the system by adding average service time. That is one by mu, so the T s will be T q plus 1, by mu that will be the average time spent in the system. By Little's formula, you can get L s. L s is equal to lambda time T s. You know the T s, so from that you can get the lambda s, which is same as what you got it in the PK formula, PK mean formula.

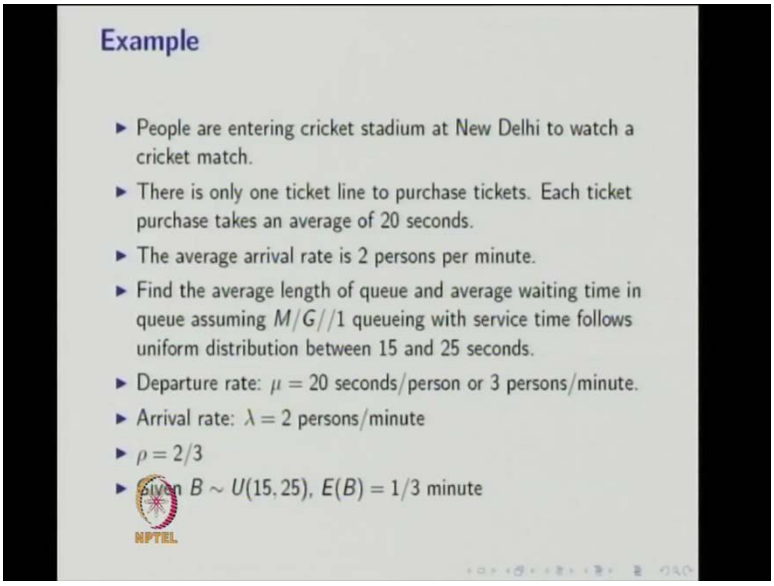
As a special case when the service time is exponential distribution with the mean 1 by mu, you can get the average time spent in the average number of customers in the system will be rho divided by 1 minus rho. This is same as the average number of customers in the M M 1 queue. In this derivation, we assume first come first serve scheduling to simplify the analysis.

(Refer Slide Time: 25:14)



But the above formula are valid for any scheduling discipline in which the server is busy, if the queue is non empty. Then no customer departs the queue before completing the service and order of service not dependent on the knowledge about the service time. If these conditions are satisfied, then for any scheduling discipline, you can use the above formula of the average number of customers in the system.

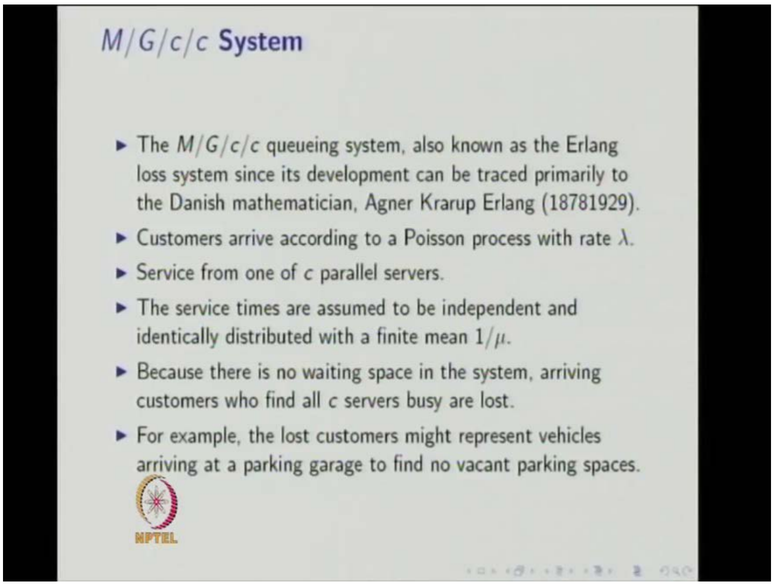
(Refer Slide Time: 26:23)



So, using Little's formula, you can find out the average time spent in the system lambda time t is equal to L s. As a simple example, consider the people entering cricket stadium at New Delhi to watch the cricket match. There is only one ticket line to purchase tickets. Each ticket purchase takes average of 20 seconds; the average arrival rate is 2 persons per minute. So, the question is find the average length of queue as well as average waiting time in queue.

Assuming the queuing model is a M G 1 queue. With the service follows uniform distribution between 15 to 25 five seconds. With this given information, you can get the departure rate, arrival rate because it is 20 seconds per person. Therefore, the rate will be 3 persons per minute and arrival rate is 2 persons per minute. Therefore, you can get lambda that is 2 by 3. So, is irreducible positive recurrent Markov chain, therefore the steady state probabilities exists and given the steady service time distribution is uniform distribution between the interval 15 to 25, you can get the measures of steady state probabilities as well as all the average measures.

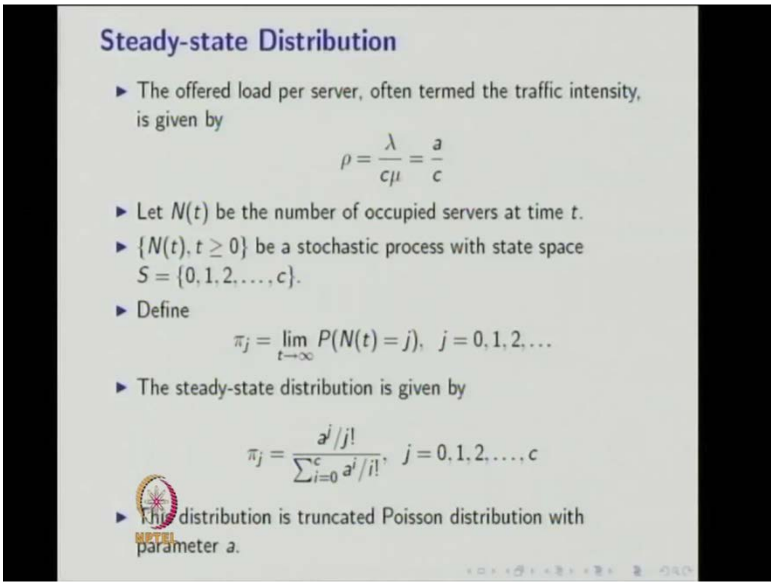
(Refer Slide Time: 27:58)



Now, we move into the second non Markovian queuing model not queuing model. Second non Markovian system, that is M G c c system, because in this model there is no queuing. The M G c c queuing system is also known as Erlang loss system, since its development can be traced primarily to the Danish mathematician Erlang. Customers arrive according to the Poisson process with the rate lambda service from of c parallel servers.

The service times are assumed to be independent and identically distributed with the finite mean 1 by mu, because there is no waiting space in the system arriving customers who finds all c server busy are lost. Therefore, this system is called a loss system not a queuing system. For example, lost customers might represent vehicles arriving at a parking garage to find no vacant parking spaces.

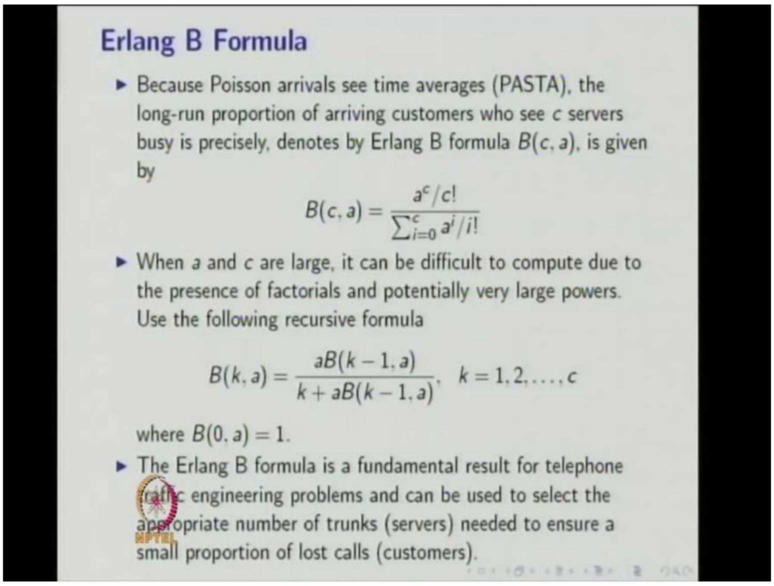
(Refer Slide Time: 29:22)



Our interest is to find out the steady state distribution, the offered load per server often termed the traffic intensity is given by lambda divided by c mu. If you denote a is equal to lambda by mu, then the rho is nothing but a by c. Let N t be the number of occupied servers at time t. So, N t is a continuous time discrete state stochastic process with state space. S define the limiting probabilities limit t tends to infinity probability. That N t is equal to j that is pi j are the steady state distribution.

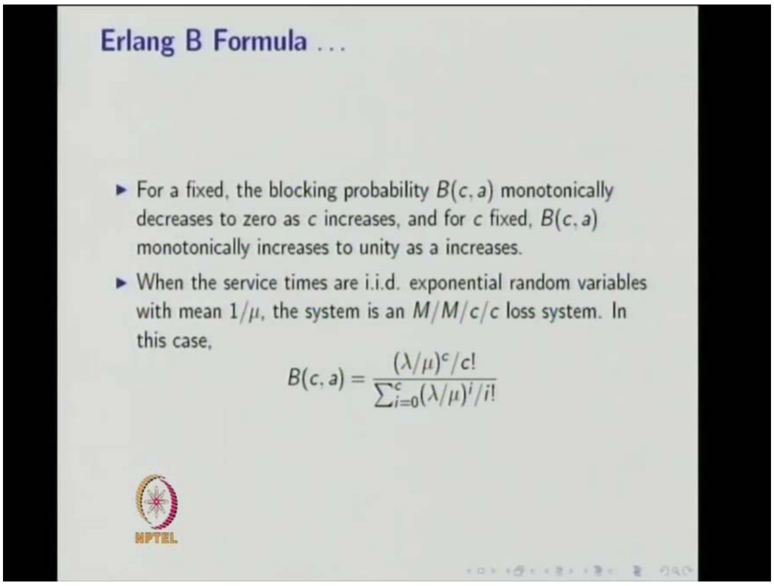
You can find easily that is nothing but a power j divided by j factorial divided summation i is equal to 0 to c a j a power i divided by i factorial. So, this is a steady state distribution of M G c c loss system. So, this distribution is a truncated Poisson distribution with the parameter small a, where a is nothing but lambda by mu. Once we know the limiting distribution, you can find the other measures the first measure.

(Refer Slide Time: 31:02)



Is Erlang B formula because Poisson arrivals see time averages, that is pasta. You can find the long run proportion of the arriving customers, who see c servers busy. That is denoted by Erlang B formula as the function of c and a, where c is the number of servers in the system and a is lambda by mu. That is nothing but the loss probability that is a power c by c factorial divided by summation i is equal to 0 to c a by i divided by i factorial when a and c are large. It can be difficult to compute due to the presence of factorials.

(Refer Slide Time: 33:44)

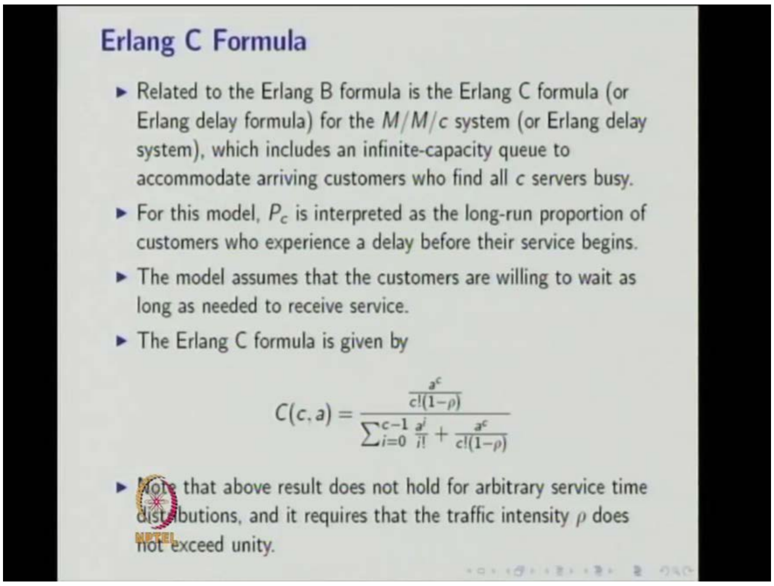


Due to the presence of factorials and potentially very large powers a power c, when c is very large, then a power c as well as this factorials giving trouble. So, we can use recursive formula, we can use the recursive formula to compute the Erlang B formula, that is in terms of that is B k comma a. In terms of B of k minus 1 comma k B of k k minus 1 comma a with the initial condition B of 0 comma a is equal to 1.

So, that means to find the value of B of 1 comma a, you use B of 0 comma a. Then to find B of 2 comma a, use of, use the value of B of 1 comma a and so on, so finally, you can get B of c comma a. In this recursive formula, we are avoiding the factorial as well as the large powers. The Erlang B formula is a fundamental result for telephone traffic engineering problems and can be used to select the appropriate number of servers need to ensure a small portion of lost customers. So, this is the way using the Erlang process, we can find out or we can select appropriate number of servers for the M G c c losses too.

For a for a fixed a where a is lambda by mu the Erlang B formula or the blocking probability or lost probability monotonically decreases to 0 as c increases. Whereas, for a fixed c, where c is a number of servers, the blocking probability monotonically increases to unity as lambda by mu increases. As a special case, when the service times are i i d random variables each having exponential random distribution with mean 1 by mu the system becomes M M c c loss system.

(Refer Slide Time: 35:00)



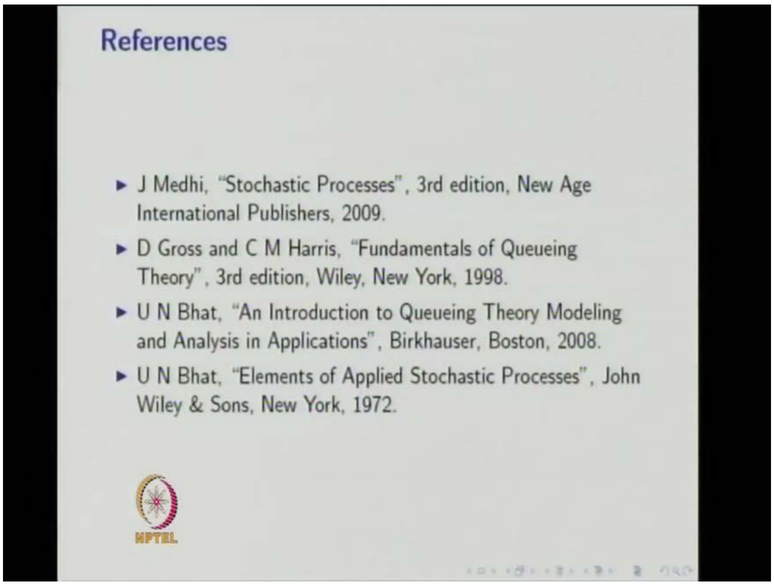
So, you can have Erlang B formula for the M M c c system also. With the function of service time of i i d random variables, each having exponential distribution with mean 1 by mu related to the Erlang B formula. We are going to discuss the other one called Erlang C formula. This is for the M M c c system not for M G c c system. This is for the M M c sorry, M M c system, which includes an infinite capacity queue to accommodate arriving customers who find all C servers are busy. That means this is the queuing system, queuing and delay system. We have seen servers and infinite capacity queue to accommodate arriving customers who find all C servers are busy.

So, corresponding to M M c queuing and delay system, we have the formula called Erlang C formula or Erlang delay formula. In this model the P c is interpreted as long run proportion of customers, who experience a delay before the service begins. The model assumes the customers are willing to wait as long as needed to receive service. So, the Erlang C formula is nothing but the blocking probability for the M M c queuing and delay system. That is in terms of that is written in this form a power c divided by c, factorial multiplied by 1 minus rho divided by this summation form.

Since, it is the queuing and delay system, you need additional condition to have the Erlang C formula, the additional condition is it requires the traffic intensity rho does not exceed 1. That means as long as rho is less than 1, the system is stable the corresponding M M c queuing and delay system will be stable. Hence, the steady state probabilities exists and once the steady state probabilities exists, you can find the loss probability and that loss probability same as Erlang C formula.

So, to have a Erlang C formula, it requires the traffic intensity rho has to be less than 1. Note that the above result does not hold for arbitrary service time distribution. So, this Erlang C formula is valid only for service times are exponential distributed not for arbitrary service time distribution, whereas, Erlang B formula is valid both for M G c c loss system and M M c c loss system. Erlang C formula is valid only for M M c queuing and delay system, with the restriction rho has to be less than 1. Whereas, Erlang B formula, the value of a is lambda by mu need not be less than 1, because that is a finite capacity and loss system.

(Refer Slide Time: 38:25)



In this, we have discussed non Markovian queues in particular M G 1 queuing system, M G c c loss system, M M c c loss system, Erlang B formula for M G c c loss system as well as M M c c loss system and finally, we have discussed Erlang C formula for the M M c queuing and delay system. With these, lecture 3 lecture 4 is completed and here is the reference for lecture 4.