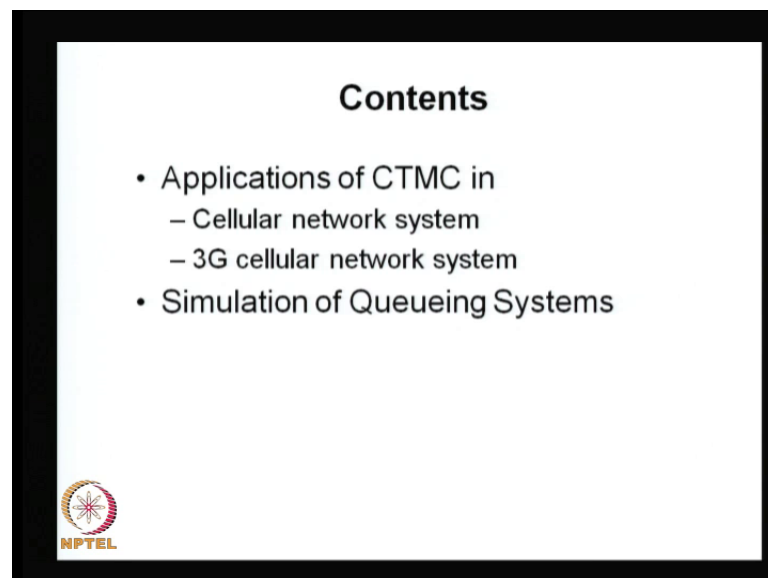


**Stochastic Processes**  
**Prof. Dr. S. Dharmaraja**  
**Department of Mathematics**  
**Indian Institute of Technology, Delhi**

**Module - 5**  
**Continuous-time Markov Chain**  
**Lecture - 7**  
**Communication Systems**

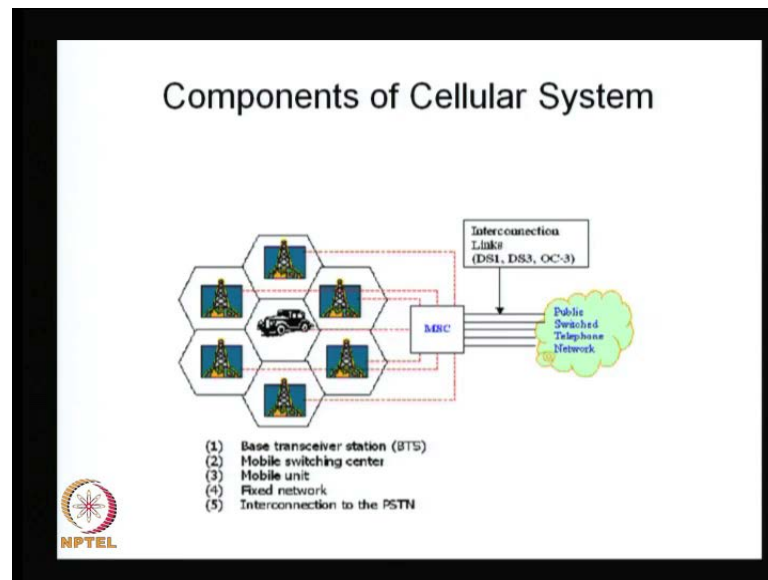
This is stochastic processes, module five, continuous-time Markov chain and seventh, this is seventh lecture, communication systems. In the last six lectures we have discussed the continuous-time Markov chain. And in the last two lectures we have discussed the application of continuous-time Markov chain in a queuing modeling. In this lecture I am going to discuss the application of continuous-time Markov chain in communication systems.

(Refer Slide Time: 01:10)



So, this stack divided into three parts. The first, I am going to discuss the application of CTMC in 2G cellular network system and then I am going to discuss the 3G cellular network system. 2G means second generation and 3G means third generation cellular network system and finally, I am going to discuss the simulation of queuing systems. It is that is basically a discrete event simulation.

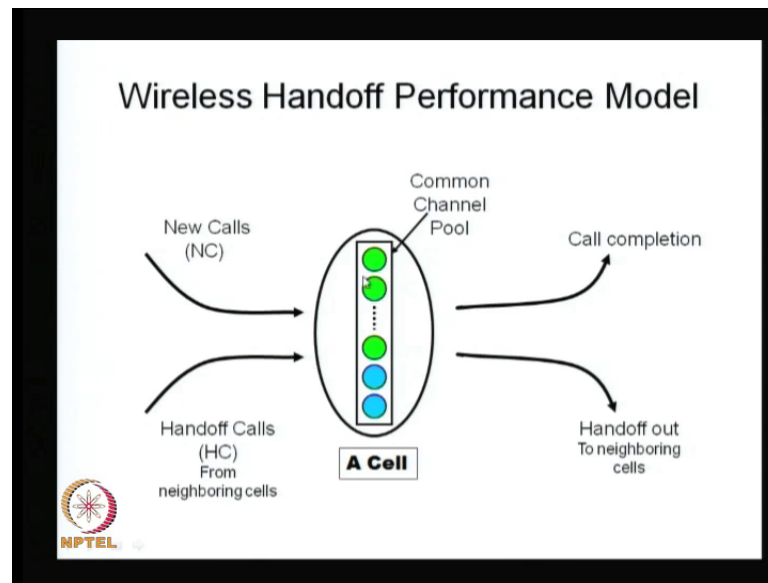
(Refer Slide Time: 01:44)



The cellular system or wireless system is nothing but a network system in which the calls are generated and calls are occupied for some time and leave the system and this consist of many cells. Each cell is of the hexagonal shape and our interest is to study the performance analysis of cellular networks using continuous-time Markov chain.

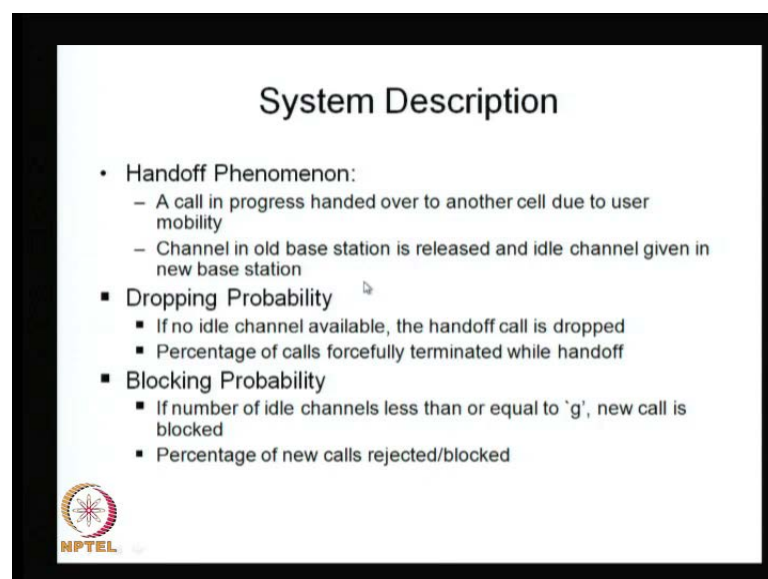
This is the center cell and all other six cells are called the neighbor cell. Either we can make the performance modeling for the whole cellular networks, or we can make the module for only one cell first, then we can apply the hierarchal modeling for the whole cellular networks. So, in this I am going to discuss the performance analysis of one cell in a cellular network.

(Refer Slide Time: 02:52)



Cellular networks are wireless networks. Our interest is to study the performance analysis. Suppose you think of bandwidth as the channels. Whenever the some calls originated it uses the one band, one cell for one channel for the communication. Suppose we assume, that we have finite number of bandwidth or finite number of channels in a one cell, suppose that you make it as some capital  $N$ , finite quantity. So, whenever the calls originated then it will take one channel from the cell and the call will be for random amount of time, it will keep going and once the call is completed, then the channel will be released.

(Refer Slide Time: 03:49)



We have designed calls into two types of calls, one is called the handoff call, the other one is called the new call. Handoff call is nothing but your call in progress handed over to another cell due to user mobility. In that case, the channel in old base station, that is in the previous cell is released and idle channel given to the new base station. So, that is the way the handoff phenomenon operates.

Therefore, I consider the call into two types of call, one is called the handoff call, the other one is new call. Handoff call is nothing but the call originated from the neighbour cells and coming into the cell where we are studying. That means, suppose the calls are originated in this neighbor cells and coming into the underlying cell, which we are considering, then that calls are called the handoff calls. So, either it can come from this cell to this cell or it can come from this cell to this cell and so on.

So, therefore, the calls originated from the neighbour cells and coming into the underlying cell, that is called the handoff calls, whereas the calls originated within the cell itself, that call is called the new calls. That means, the call, the channel will be taken, the channel will be given from same base station of the cell and once the call is get over, then the channel will be released.

Therefore, the calls originated from the neighbor cell is called handoff calls. The calls originated within the cells are called the new calls. These are all the two types of calls coming into the system. If there is a bandwidth or channel, then the calls will be allotted and after the calls are completed, the channel will be released.

So, there are two types of call completion also. Either the call is completed within the cell itself, not moving anywhere due to mobility. There is a possibility the calls can be originated within the cell or coming from the neighbor cell. It can last for whole time in the cell and it would have move to the other cell also, in that case it is called a handoff to neighbour cell. The way the calls are originated from the neighbour cell coming into the cell, the same way you can discuss the outgoing. Also, the calls can be completed within the cell itself or call, it keeps going to the neighbour cells, that is called as the handoff out calls.

Now, I am going to discuss what is the concept of dropping and blocking probability. Since the hand off calls are having the higher priority because the calls are already originated and keep coming into the cell and there is the possibility it can move into the

other cell also, therefore we should give a higher priority to the handoff calls. There are many ways of giving the higher priority to the hand off calls, the one easy way is reserve few channels for the handoff calls. That means, out of  $N$  channels users, few channels for the handoff calls, not physically few channels.

Whenever you have left out few channels in the system in the pool, then if that number is more than the reservation, then you keep allowing the new calls as well as the handoff calls. If you reserved some number of channels for the handoff calls and the available bandwidth or available channels at, at a time of arrival it is less than that, then do not allow the new calls. So, only the handoff calls will be allowed whenever the number of available channels are going less than or equal to the prescribed number, so that is called a fixed reservation channel policy.

So, suppose out of capital  $N$  channels you reserved some  $g$  channels for the handoff calls, that means, if number of idle channels less than or equal to  $g$ , then the new call will be blocked, whereas at that time if the handoff calls enter into the system, then it will be, channel will be allocated.


If no channel, idle channel is available, that means, all the channels are busy with the calls. Then, handoff calls will be dropped. Obviously, new calls also will be dropped if no channel is available. But if the number of idle channel is less than or equal to  $g$  itself, from that point onwards the new calls will be blocked. So, our interest is how to find the, our interest is to find the blocking probability and the dropping probability.

So, the dropping probability is nothing but the handoff calls are dropped because of no, no idle channel in the system. The blocking probability is nothing but the probability, that the new calls are blocked because the number of idle channels are less than or equal to  $G$ . Once we are able to find out this to performance measures, then using that one can design better cells, how to, for a given number of channels, how we can reduce the blocking probability, as well as, the dropping probability. So, these measures we are going to find out using a continuous-time Markov chain property.

(Refer Slide Time: 10:04)

### Basic Model

- Calls arrives in Poisson Processes
  - $\lambda_1$ : Rate of Poisson arrivals for NC
  - $\lambda_2$ : Rate of Poisson arrivals for HC
- Exponential service times
  - $\mu_1$ : Rate of ongoing service
  - $\mu_2$ : Rate of handoff of call to neighboring cell
- N: Total number of channels in pool
- g: Number of guard channels



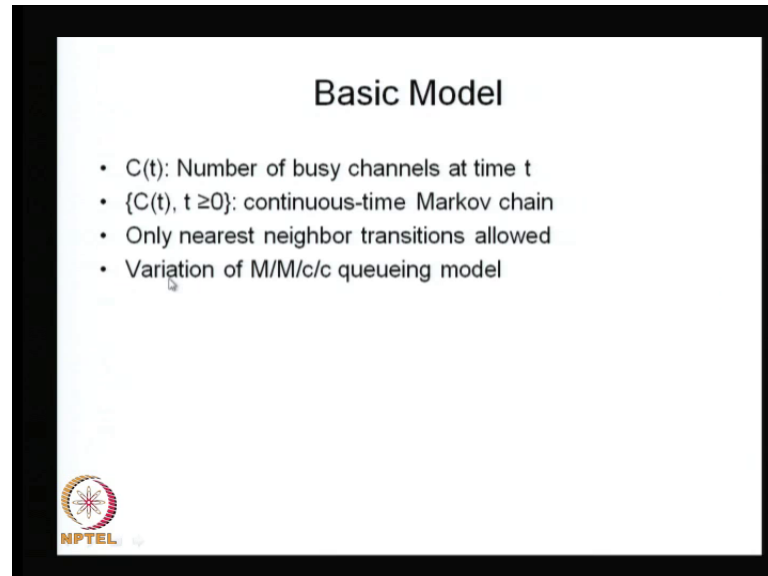
So, for that we are going to make a few assumptions, so that the underlying stochastic process will be a continuous-time Markov chain. And we have, here we are making a model for only one cell, not the whole cellular networks. So, we make the assumption, the calls arrive in Poisson process for the new calls as well as the handoff calls and both are independent. We make the rate  $\lambda_1$  is the arrival rate for the new calls and the  $\lambda_2$  is arrival rate for the handoff calls and both are independent. Therefore, the call arrivals, either it is new calls or handoff calls, that will be sum of two Poisson, therefore that is also Poisson process with the parameter  $\lambda_1 + \lambda_2$ .

And also we make assumptions over the call completion. Either the call completed within the cell that is exponentially distributed. the time for the call completion within the cell, that is exponentially distributed with the parameter  $\mu_1$  as well as the calls handoff to other cells, there is also exponentially distributed with the parameter  $\mu_2$ . And we make the assumptions, the call completion as well as the handoff call to other cells that is also independent.

And we have a total number of N channels is capital N, out of that we reserved g channels for handoff calls and these channels are called the guard channels. So, with this assumption, that means, we are considering only one cell and the new calls for that stream is the Poisson stream and this stream is also Poisson stream. And the call duration, either call completion is exponentially distribution or handoff to the neighbour

cells, that is also exponentially distribution, and we have total number of  $N$  channels out of that  $G$  is for the reserved channel.

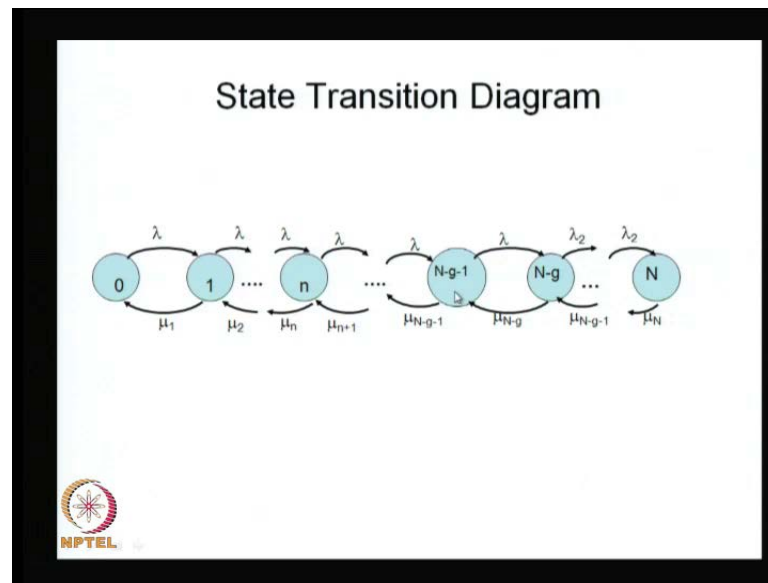
(Refer Slide Time: 12:10)



And accordingly, the Markov chain, first we are, you are going to make a stochastic process. Stochastic process is a  $C_t$ , that is nothing but the number of busy channels or each channel is the allocated for one call. Therefore, at time  $t$  how many, what is the number of calls, that is, the stochastic process, it is the random variable.

So, over the  $t$ , that is the stochastic processes and since it is number of busy channels, therefore it is a discrete-state continuous-time stochastic process with the assumptions we have discussed in this slide. The stochastic process will be a continuous-time Markov chain. Only nearest neighbour transitions are allowed, therefore this is not only a continuous-time Markov chain as a special case of continuous-time Markov chain that is a birth-death process also. So, it is a variation of M/M/c/c queueing model where the rates are different.

(Refer Slide Time: 13:18)



That means, the birth rates are  $\lambda_2$   $N$  minus  $g$  minus  $1$ , sorry,  $N$  minus  $g$ , after that the birth rates are  $\lambda_2$ . If you recall, this  $\lambda$  is nothing but  $\lambda_1$  is the arrival rate for the new calls and  $\lambda_2$  is the arrival rate for the handoff calls and  $\lambda$ , here it is  $\lambda_1$  plus  $\lambda_2$ . That means, whenever the number of busy channels, it is starting from  $0$  to  $N$  minus  $N$  minus  $g$  minus  $1$ , both the new calls as well as handoff calls are allowed into the system.

Therefore, the arrival rate is the  $\lambda$ , birth rates are  $\lambda$ , whereas the number of busy channels from  $N$  minus  $g$  to  $N$  minus  $1$ , the birth rates are only  $\lambda_2$  because the new calls are blocked. Only handoff calls are allowed whenever the system size is  $N$  minus  $g$  to  $N$  minus  $1$ . Whenever the number of busy channel is  $1, 2$  and so on till  $N$ , either the call is completed or the calls are handoff to the neighbor cell and both are exponentially distributed with the parameters  $\mu_1, \mu_2$  respectively and both are independent. Therefore, the rates are going to be  $\mu_1, \mu_2$  and so on where  $\mu_1, \mu_2$  is nothing but...

So, here is some notation problem. This should be, this is nothing to do with  $\mu_1$  and  $\mu_2$ , which I have defined earlier, that, that should be  $\mu_h$  and  $\mu_N$ , where  $\mu_N$  is equal to  $1$  times  $\mu_h$  plus  $\mu_N$  and  $\mu_2$  is nothing but  $2$  times  $\mu_h$  plus  $\mu_N$  and so on.


So, this is the death rates, therefore the number of busy channels at any time that is a birth-death process with birth rates  $\lambda$  is till  $N - g - 1$ , after that it is  $\lambda_2$ . And the death rates are  $N$  times  $\mu$  where  $\mu$  is nothing but  $\mu_1 + \mu_2$ . Either I can use this notation or the other notation.

(Refer Slide Time: 15:51)

### Steady-State Equations

Solve  $PQ = 0$ ;  $\sum_{i=0}^N P_i = 1$

$P_n = \lim_{t \rightarrow \infty} \text{Prob}\{C(t) = n\}$ ,  $n = 0, 1, \dots, N$

$$\begin{aligned} 0 &= -\lambda p_0 + \mu_1 p_1 \\ 0 &= \lambda p_{n-1} - (\lambda + n\mu_n) p_n + \mu_{n+1} p_{n+1} \quad n = 1, 2, \dots, N-g-1. \\ 0 &= \lambda p_{N-g-1} - (\lambda_2 + \mu_{N-g}) p_{N-g} + \mu_{N-g+1} p_{N-g+1} \quad n = N-g \\ 0 &= \lambda p_{n-1} - (\lambda_2 + \mu_n) p_n + \mu_{n+1} p_{n+1}, \quad n = N-g+1, \dots, N-1 \\ 0 &= \lambda_2 p_{N-1} - \mu_N p_N. \end{aligned}$$


(Refer Slide Time: 16:23)

### Steady-State Solutions


Let  $\lambda = \lambda_1 + \lambda_2$ ,  $\mu = \mu_1 + \mu_2$ ,  $A = \rho = \frac{\lambda}{\mu}$ ,  $A_1 = \frac{\lambda_2}{\mu}$ .

$$p_n = p_0 \begin{cases} \frac{A^n}{n!}, & n \leq N-g \\ \frac{A^{N-g}}{n!} A_1^{n-(N-g)}, & n \geq N-g \end{cases}$$

where

$$p_0 = \frac{1}{\sum_{n=0}^{N-g-1} \frac{A^n}{n!} + \sum_{n=N-g}^N \frac{A^{N-g}}{n!} A_1^{n-(N-g)}}$$

G. Haring, R. Marie, R. Puigjaner, K.S. Trivedi, "Loss Formulas and Their Applications to Optimization for Cellular Network", IEEE Transactions on Vehicular Technology, 50 (3) 664-673 (2001).



Once I know the, the underlying stochastic process is the birth-death process and that it is the finite state and also it is a reducible, so for all the states are positive recurrent. So, I

can solve,  $P_Q$  is equal to 0 and the summation of  $P_i$  is equal to 1. I can get the steady state or equilibrium stationary probabilities.

So, if I solve this system of the equation with the summation of  $P_i$ ,  $P_i$  is equal to 1. You will get the  $P_i$ 's where  $P_i$ 's is in terms of initially  $P_{naught}$  and the  $P_{naught}$  is 1 divided by this. So, this result is obtained, is available in this following paper. Here,  $\lambda$  is nothing but  $\lambda_1$  plus  $\lambda_2$ ;  $\mu$  is nothing but  $\mu_1$  and plus  $\mu_2$  and those  $\mu_i$ 's are the  $i$  N times  $\mu$  and  $A$  is nothing but  $\rho$ , that is,  $\lambda$  by  $\mu$  here and  $A_1$  is  $\lambda$  by  $\mu$  because the birth-death process with the birth  $\lambda$ 's and after that  $\lambda_2$ . Therefore, you have a different expression for  $A_0$  to  $N$  minus  $g$  and  $N$  minus  $g$  to capital  $N$  you have a different expression for the  $p_n$ 's. So, this is the steady state probability.

(Refer Slide Time: 17:19)

### Loss Probabilities


- **Dropping Probabilities**

$$P_d(N, g) = p_N$$

$$= \frac{\frac{\lambda^{N-g} A_1^g}{N!}}{\sum_{n=0}^{N-g-1} \frac{\lambda^n}{n!} + \sum_{n=N-g}^N \frac{\lambda^{N-g} A_1^{n-(N-g)}}{n!}}$$
- **Blocking Probabilities**

$$P_b(N, g) = \sum_{n=N-g}^N p_n$$

$$= \frac{\lambda^{N-g}}{\sum_{n=0}^{N-g-1} \frac{\lambda^n}{n!} + \sum_{n=N-g}^N \frac{\lambda^{N-g} A_1^{n-(N-g)}}{n!}}$$



Once you know the steady state probability, our interest is to get the dropping probability and blocking probability. Dropping probability is nothing but the handoff calls are dropped because of the system size is full, that means, all  $N$  channels are busy. That means, the system, when the system size is capital  $N$ , then the new calls are dropped. Therefore, the dropping probability is equal to the probability, that  $N$  customer in this system. Here,  $N$  is nothing but the number of channels; capital  $N$  is the number of total number of channels, all are busy. Therefore, the  $p$  suffix  $N$  will be the dropping probability.


Blocking probability, blocking probabilities is related to the new calls. So, the new calls are blocked, is not even entered into the system whenever the number of available channels are less than or equal to  $g$  in the system. Therefore, when the system size is from  $N - g$  to capital  $N$ , the new calls are blocked. Therefore, the blocking probability is running from  $N - g$  to  $N$  p suffix  $n$ , that will be the blocking probability.  $(( ))$  mean the summation over the steady state probability, which we got it in the previous slide, substitute those things. So, that will be the blocking probability.

(Refer Slide Time: 18:50)

### Description of 3G Cellular Networks

- Homogeneous Cellular Network with identical cells
- Network traffic:
  - Voice: New and handoff
  - Video: New and handoff
    - Hard QoS (e.g., video conferencing, live program) (say *video1*)
    - Soft QoS (video on Demand) (say, *video2*)
  - Data: New and handoff
- Priority: Voice > Video1 > Video2 > Data

S. Dharmaraja, Vaneeta Jindal and Attahiru Sule Alfa: Phase Type Models for Cellular Networks Supporting Voice, Video and Data Traffic, *Mathematical and Computer Modelling*, 47 (2008) pp. 1167 - 1180.



Now, I am moving into extending the same logic into 3G wireless networks. So, here also, we are making model for only one cell and not only, only one type of calls. So, we have three types of calls: one is the voice call, the second one is the video call and third one is the data call. In the 2G networks we have only one type of call, that is, the voice call and the voice calls are subdivided into two types, handoff calls and the new calls.

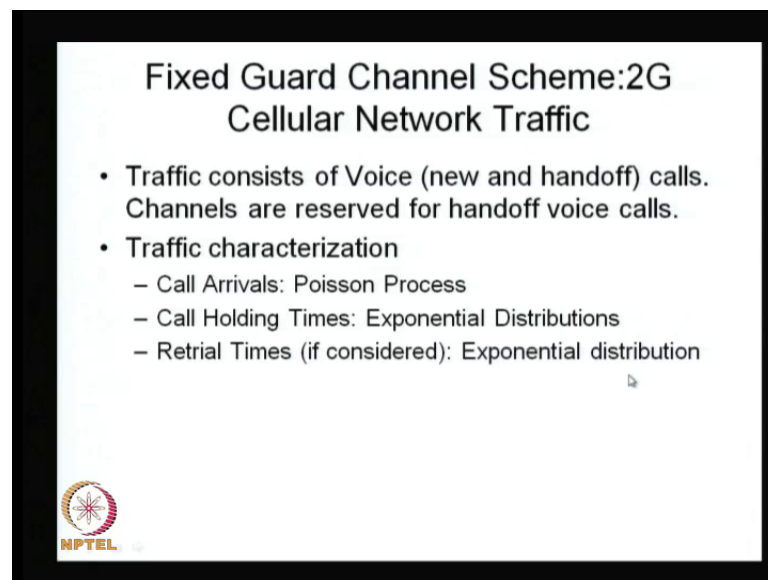
So, here also the voice calls are divided into new calls and the handoff calls because we are considering a performance model for one cell and the calls could be voice calls, could be new calls or the handoff calls. The video calls, again, we are dividing into two types, one is hard QOS video one type call and the soft QOS, that is, video two calls because there are some video calls, it is hard QOS. That means, it would not tolerate the delay, the video conferencing live program and so on. So, that type of video calls are the

video 1, whereas the video on demand, that can tolerate the delay, so that type of call is the called soft QO QOS calls. So, this is a video call, that type is called the video 2.

Then, the last data traffic, that is, also can be classified into new call and handoff calls. So, here the priority is, the voice calls are the higher priority, then the video 1 call, then video 2, then the data. So, the way we have discussed the 2G cellular networks, we are extending the same idea into 3G wireless networks. So, three types of calls, this priority and priority is taken care with the fixed (( )) channel policy. Earlier we have only one voice type call, now we have three types of calls. Therefore, the preservation also changes it accordingly and instead of only one voice call we have three types of... Therefore, the number of channels allocated to different calls also changes.


So, in this scenario we have taken voice calls users only, one channel, whereas the video calls that needs more bandwidth, therefore assumption, we have made an assumption, therefore channels are needed for the video calls and the data traffic needs only one channels, therefore voice and... So, this is explained in this paper in detail.

(Refer Slide Time: 21:32)



**Fixed Guard Channel Scheme: 2G Cellular Network Traffic**

- Traffic consists of Voice (new and handoff) calls. Channels are reserved for handoff voice calls.
- Traffic characterization
  - Call Arrivals: Poisson Process
  - Call Holding Times: Exponential Distributions
  - Retrial Times (if considered): Exponential distribution


 NPTEL

So, here we made the assumption, the call arrivals Poisson process, the usual way is 2G networks, call arrivals are Poisson process, call holding times or exponential distribution and so on.

(Refer Slide Time: 21:45)

### CAC and Resource Reservation Schemes

- **Assign**
  - One channel - Voice and Data traffic
  - Four channels - Video (1 and 2) traffic
- **Priority**
  - Order – Voice, Video 1, Video 2 and Data
- **Reserved channels**
  - $h_i$  – handoff calls (Voice, Video 1, 2 and Data)
  - $n_i$  – new calls (Voice, Video 1, 2)
- **Retry and accept**
  - New Voice, Video 2 and Data traffic




But what we are making the, extending the model with one channel is needed for voice and data traffic and the four channels are needed for the video 1 and 2 and the data traffic. And the priority is voice, then video 1, then video 2 and data. And since the priority is there, we are reserving  $h_i$  number of channels for the handoff calls in this order and  $n_i$  channels are reserved for the new calls of voice call and video 1 and video 2 and there is no reservation for the new data traffic also. And we allow the retry of the calls of new voice and the video two and the data traffic.

(Refer Slide Time: 22:32)

### The Proposed Model

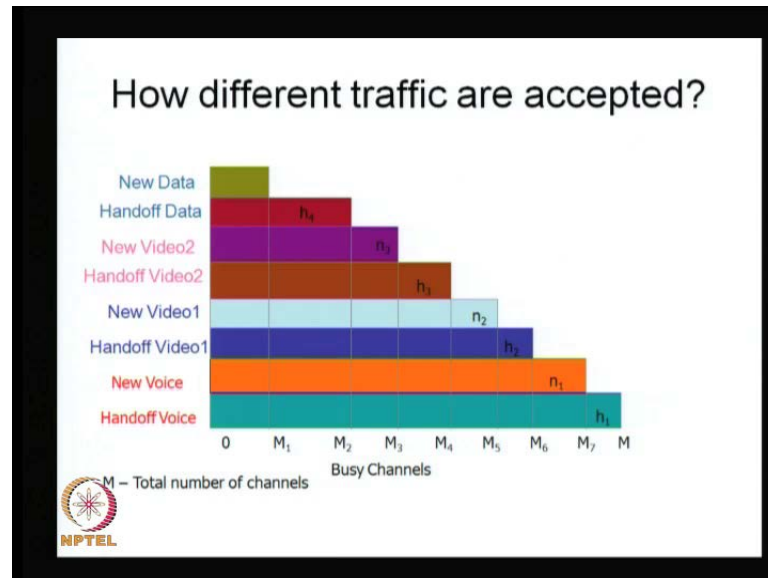
#### **Traffic Characterization**

- Traffic arrival: Markov arrival process
- Call holding times: Phase type distributions
- Calls that are blocked joins the orbit
- Call retrial times: Phase type distributions



Therefore, the proposed model, instead of a Poisson we have Amarkov arrival process for the data arrival and instead of exponential distribution we have more general situation, that is a phase type distribution for holding time, call holding time. And whenever the calls that are blocked, it can go to the orbit and it can be retried and the retried time also assumed to be Poisson distribution.

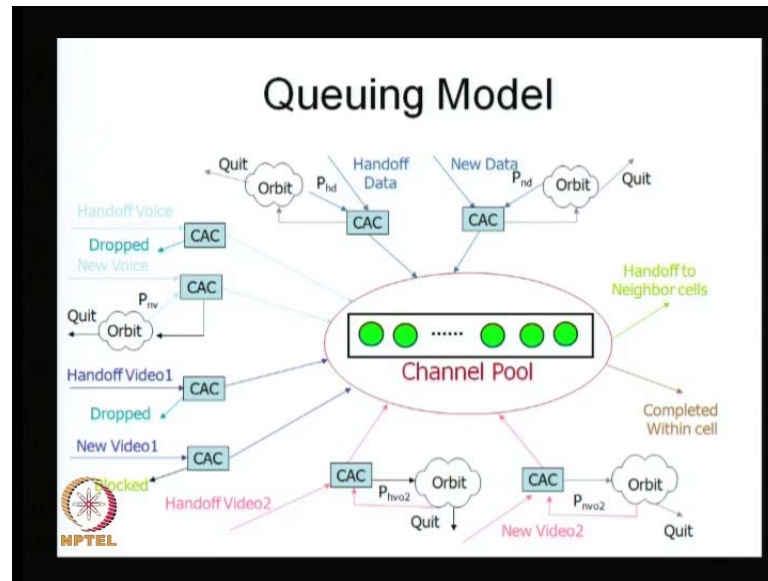
(Refer Slide Time: 22:57)



And since we have made the reservation for the calls with the  $h_i$ 's and  $n_i$ 's accordingly, these total number of channels are divided into many parts and the calls will be all, so the calls will be allotted according to this priority that means, from 0 to  $M_7$ . The handoff voice will be allotted from 0 to  $M$ , whereas the new voice will be allotted till the number of busy channels going from 0 to  $M_7$  because of  $h_1$ ,  $h_1$  is the reservation for the handoff voice call.

So, like that this is the, this is the higher priority, then next priority and next priority and so on. And this is the new data call and there is no priority for the hand, there is no priority for the new data, therefore this is the last lowest priority called traffic.

(Refer Slide Time: 24:00)



And this is the corresponding queuing model. You can compare with the queuing model of the 2G networks with the 3G networks. Since we have three types of calls, that is, voice, video 1, video 2 and data traffic, so we have different traffic is contending for the channels in the channel pool and after the calls are over, then it will either (( )) completed within the cell or handoff to the neighbour cells.

The first one is the handoff voice. So, the call admission control schema allows, then the call will be allotted otherwise it will be dropped. There is no quitting, there is no orbit the calls blocked one, go to the orbit and quit. Whereas the new voice call, if there is no channel, then it will go to the orbit with some probability, it will again retry after phase time distributed time, which I have, we made to the assumption, otherwise it will quit with 1 minus this probability.

Similarly, for the handoff video 1, either it will be accepted or it will be dropped, whereas the new video 1, the new video 1 also, either it is, we use the word dropped for the handoff and the blocking for the new type new calls, therefore new video call, video 1 call will be blocked if there is no bandwidth, whereas the video 2. If there is no bandwidth, then it will go to orbit with some probability. It will come back to 1 minus of this probability, it will quit.

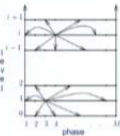
Similarly, the new video 2 also, either it will get accepted with the channel is available, the required number of channels are available, otherwise it goes to the orbit and retried.

The fourth type, that is, the data, both new and the handoff data, either it will be accepted or it will go to orbit and come back. So it is similar to video 2, but the difference is the video takes the four channels at a time, whereas the data and the voice takes only one channel at a time. So, this is the corresponding queuing model for the 3G wireless networks, cellular networks for only one cell.


(Refer Slide Time: 26:25)

### Analytical Model

- $\{X(t), t \geq 0\}$ : stochastic process to describe the state of the cell.
- It is modeled as a Quasi birth-death process.
 



- a generalization of the birth-death process
  - a Markov chain on the state space S, where the state space can be divided into k levels, and each level has m states (phases)
- State Space:
 
$$\{(i, j, k, l, u_{ve}, u_{vo1}, u_{vo2}, u_d, S_{i-j-k}^{ve}, S_j^{vo}, S_k^d, r_l)\}$$



So, here the stochastic process is the state of the cell. What is the number of channels are available at the time, that is going to be the stochastic process over the time and this will be modeled as a quasi birth-death process, whereas the 2G wireless, net, cellular networks is (( )) modeled as the birth-death process. So, here this is the quasi birth-death process.


So, what is the meaning of quasi birth-death process? This is a generalization of the birth-death process. Here, the Markov chain process with states space S and the state space can be divided into k levels and each level will have a capital N states or M phases. Suppose you have M, capital M phases and we have queue levels in it. So, it is a generalization of birth-death process, means suppose the system is in some ith level and the some state, then there is the possibility it can go to i minus 1th level with the different states or i plus 1th level with the different states and so on. Therefore, here the states phase is divided into k levels and each level has the some finite number of, need not be the same number, each level should have some finite number of states. So, it is a

generalization of the birth-death process. Accordingly, we have a state space with this type of vector and you can study in detail the particular paper about how to obtain the performance measures.

(Refer Slide Time: 28:11)

### Analytical Model

- $i$ : total no. of ongoing calls,  $0 \leq i \leq M$
- $j$ : Video calls,  $0 \leq j \leq j_m$ ,  $j_m = \min\{i, M/4\}$
- $k$ : Data calls,  $0 \leq k \leq k_m$ ,  $k_m = \min\{i-j, M/2\}$
- $l$ : No. of calls in orbit,  $0 \leq l \leq J$
- $u_s$ : phase of MAP for traffic type  $s$  ( $s = ve, vo1, vo2, d$ ),  $1 \leq u_s \leq K_s$
- $K_s$ : the number of states in the Markov chain of MAP for traffic type  $s$
- $s_{i-j-k}^{ve}, s_j^{vo}, s_k^d$ : the set of phases of service for ongoing voice, video and data calls respectively
- $r_l$ : phase of retrial of call in an orbit.



And here I am listing what is the meaning of  $i, j, k, l$  and what is the meaning of  $u_s$  and what is the meaning of  $s$  suffix and so on. And the last one, what is the meaning of  $r_l$  means, the phase of retrial of calls in orbit.


(Refer Slide Time: 28:29)

### Generator Matrix

$$Q = \begin{bmatrix} Q_{01} & Q_{00} & & & \\ \vdots & \vdots & \ddots & & \\ & Q_{12} & Q_{11} & Q_{10} & \\ & & \ddots & \ddots & \\ & & & Q_{M2} & Q_{M1} \end{bmatrix}$$

$$Q_{i0} = \begin{bmatrix} A_{01}^i & A_{00}^i & & & \\ \vdots & \vdots & \ddots & & \\ & 0 & A_{j1}^i & A_{j0}^i & \\ & & \ddots & \ddots & \\ & & & 0 & A_{j_m1}^i \end{bmatrix}$$

- $Q$ : Tri-diagonal block matrix
- $Q_{i0}$ : Increase in the no. of ongoing calls



So, here also we use the difference between that birth-death process and the quasi birth-death process. The Q matrix consists of block matrices, whereas in birth-death, birth-death process, the q matrix is tri-diagonal matrix and each entries are the elements. Whereas here, each entries are the matrix and the elements of that matrix also again will be a matrix based on the number of levels.

And if this model, we have four level quasi birth-death process. Therefore, we have a Q matrix with the block matrices in three levels, four levels and the fourth level, you have the matrix with the entries are the elements. I am just skipping this part of how the generator matrix, each element, each matrix exists.

(Refer Slide Time: 29:27)


### Steady-state Distribution

Assuming QBD is ergodic, steady-state distribution is obtained.

Let  $\mathbf{x}$  partitioned as  $\mathbf{x} = \mathbf{x}(i, j, k, l)$  with  $0 \leq i \leq M$ ,  $0 \leq j \leq j_m$ ,  $j_m = \min\{i, M/4\}$ ,  $0 \leq k \leq k_m$ ,  $k_m = \min\{i-j, M/2\}$ ,  $0 \leq l \leq J$  denote the steady-state probability vector for the generator matrix Q i.e.,  $\mathbf{x}$  satisfies  $\mathbf{x}Q = 0$ ,  $\mathbf{x}\mathbf{e} = 1$ .

Each  $\mathbf{x}(i, j, k, l)$  is a vector ordered in the lexicographical order.

Using matrix-geometric technique, steady-state probabilities are obtained.



And our interest is to get the steady state distribution for the quasi birth-death process also. So, here I am making the assumption, the quasi birth-death process is ergodic, therefore the steady state distribution exists. This is generalization of birth-death process, therefore the Markov chain, whatever the steady state distribution condition, you need the same condition, you can cross-check. So, here this QBD is ergodic, therefore the steady state distribution exists, then you can solve the  $\mathbf{x}$  times Q matrix equal to 0 with the  $\mathbf{x}$  times  $\mathbf{e}$  vector, that is equal to 1. Using that you can get the steady state probability vector. And since it is a QBD, you can use the matrix-geometric technique; that is a well known technique for the QBD. We can use that and get the steady state probabilities.


(Refer Slide Time: 30:26)

## Performance Measures

- Dropping probability of handoff voice calls:

$$P_{hv} = \sum_i \sum_j \sum_{k=0}^{k_m} \sum_{l=0}^J x(i, j, k, l) e$$

- Dropping probability of handoff video1 calls:

$$P_{hvo1} = \sum_i \sum_j \sum_{k=0}^{k_m} \sum_{l=0}^J x(i, j, k, l) e$$


And these are all the, once you know the steady state probabilities, you can get the blocking and dropping probabilities. Since we have three types of calls, you can analyze the blocking and dropping probabilities for the various traffics and also, we are discussing the retrial and orbit and so on. So, we can find out the probability, that the new voice handoff video two, new video two handoff data.

(Refer Slide Time: 30:43)

## Performance Measures


- Probability that new voice, handoff video2, new video2, handoff data and new data calls join the orbit:

$$P_{nv} = (1 - P_r(J)) \sum_i \sum_j \sum_{k=0}^{k_m} \sum_{l=0}^J x(i, j, k, l) e,$$

$$P_{hvo2} = (1 - P_r(J)) \sum_i \sum_j \sum_{k=0}^{k_m} \sum_{l=0}^J x(i, j, k, l) e,$$

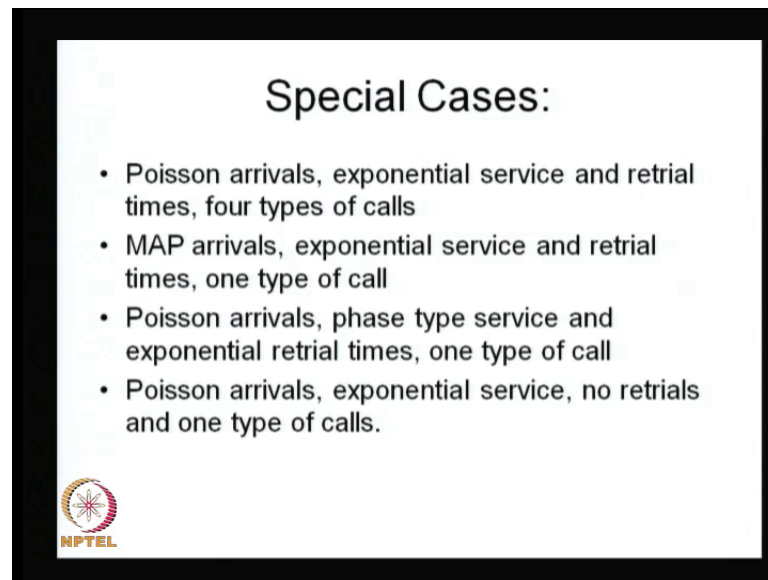
$$P_{nvo2} = (1 - P_r(J)) \sum_i \sum_j \sum_{k=0}^{k_m} \sum_{l=0}^J x(i, j, k, l) e,$$

$$P_{hd} = (1 - P_r(J)) \sum_i \sum_j \sum_{k=0}^{k_m} \sum_{l=0}^J x(i, j, k, l) e,$$

$$P_{nd} = (1 - P_r(J)) \sum_i \sum_j \sum_{k=0}^{k_m} \sum_{l=0}^J x(i, j, k, l) e,$$



And new data calls join the orbit, what is the probability for that you can find out after getting the steady state probabilities. So, you can one can see the paper to understand the complete paper.

(Refer Slide Time: 31:09)



**Special Cases:**

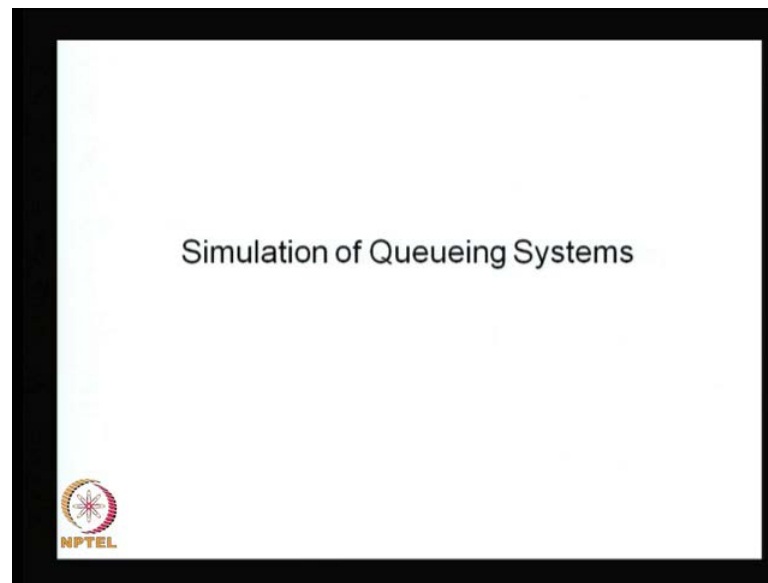
- Poisson arrivals, exponential service and retrial times, four types of calls
- MAP arrivals, exponential service and retrial times, one type of call
- Poisson arrivals, phase type service and exponential retrial times, one type of call
- Poisson arrivals, exponential service, no retrials and one type of calls.



And also, one can discuss the special cases by making the simpler assumptions. The last assumption is Poisson arrival, exponential service and no retrials and one type of call that will land up into the 2G cellular networks, which we have got it. So, if you make subsequent special cases, you land up with the blocking probability and the dropping probability, which we have discussed in the second generation cellular networks.

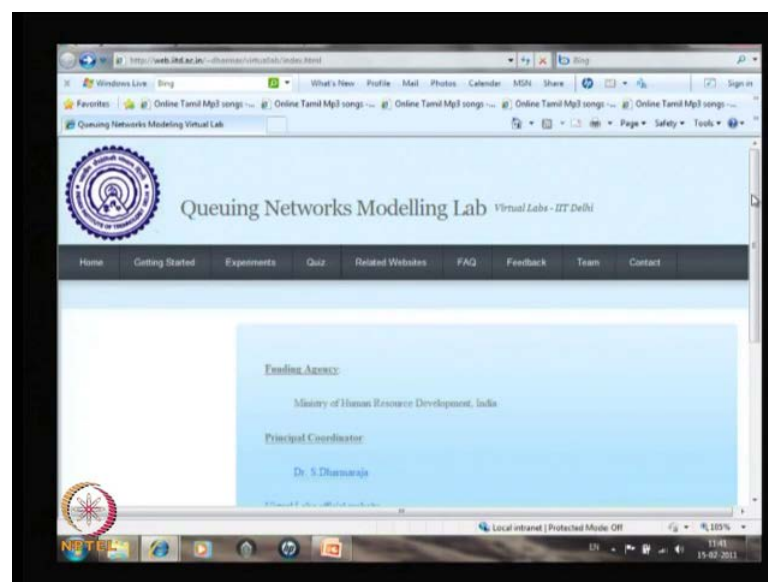
Now, I am moving into the last part of module five continuous-time Markov chain. In this I am going to discuss the simulation of queuing systems. So, the module five continuous-time Markov chain started with the definition and properties and so on. Then, I discussed the birth-death process, then I discussed the application of birth-death process in simple Markovian queuing models, then also I have discussed the queuing network that is also a multidimensional continuous-time Markov chain as application.

(Refer Slide Time: 31:44)



And finally, I have given few practical applications in cellular networks for the performance analysis. Now, I am going to discuss the desecrating simulation of the simple Markovian queuing systems.

(Refer Slide Time: 32:48)

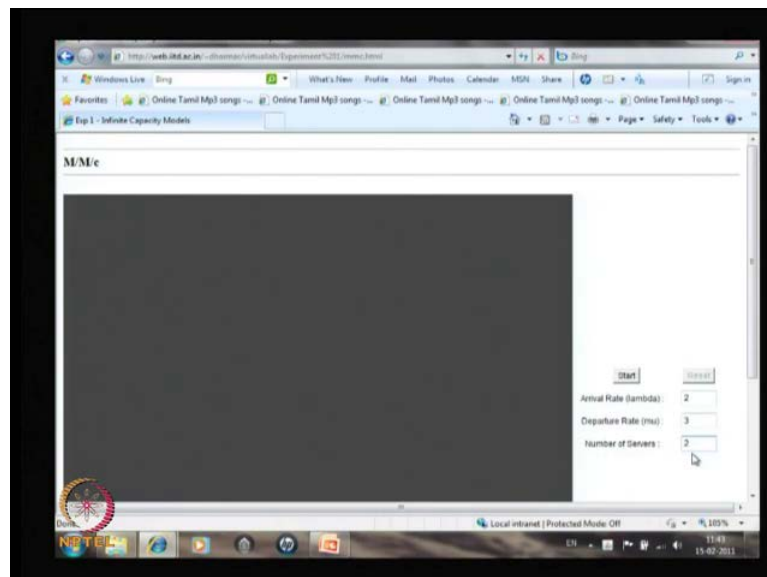


So, in this queuing network modeling lab one can do discrete event simulation and the discrete event simulation for the queuing network involves Markovian queues. We can do some experiment over the Markovian queues and we can do the discrete event simulation for the non-Markovian queues and one can do the discrete event simulation

for the queuing networks also. And finally, one can do the full queues also one can simulate.

Since I have discussed only continuous-time Markov chain till now, I am going to do discrete event simulation for the Markovian queue. So, that is in the first three experiments, then other three experiments are related to the non-Markovian queues. The first experiment consists of a simulation of M/M/1 queue, single server queue and the M/M/c of finite server on queues and the infinite server queue. So, in the, sometime I have done the discrete event simulation of M/M/1 queue.

(Refer Slide Time: 34:03)



So, let me go for M/M/c finite server queue. So, in this you need the input of arrival rate, input of the departure rate and the number of servers. It is a multi-server infinite capacity model. So, suppose you choose the arrival rate is 2 and the service rate is 3 and the number of servers are 2, that means, it is M/M/2 infinity model with the arrival rate 2 and the service rate is 3.

(Refer Slide Time: 34:41)



So, one can start the discrete event simulation by clicking the start, then you will get the window of... So, this is the sample path over the time. What is the system size? So, at this time point one arrival comes, then two, then service is completed and so on. So, this is the sample path over the time.

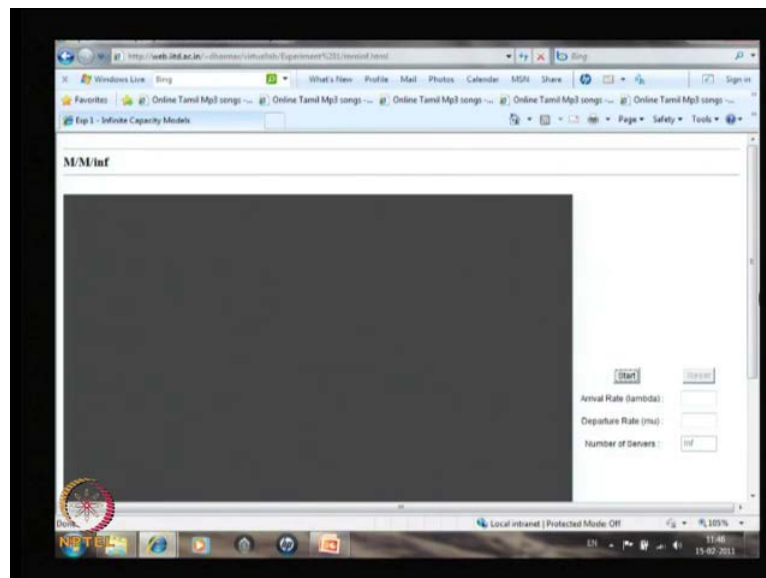
And here you can get the performance measures. So, whatever I discussed, the performance measures of a steady-state distribution and all other performance measures, one can get it with the theoretical results in the third column, whereas the second column gives the running time  $t$ , the discrete event simulation transfer 15 time unit and what is the result for the mean number of system and so on. And this will converge to this value for  $t$  tends to infinity.

So, I make sure, that this arrival and departure rates are satisfying, the conditions at the steady-state distribution exist. Therefore, as  $t$  tends to infinity, this will reach the steady state theoretical result. If you change the value arrival rate and the departure rate something else, then there is the possibility if conditions for the stationary distribution are not satisfied. Then, still the runtime results you will get, but that would not become averse to the theoretical and also the steady-state results would not be possible. It will get dash here. So, as long as the steady-state distribution, those conditions are satisfied, then you will have results in the third column, otherwise you would not have the results needed.

So, now you can see the throughput. Till this much time you are getting 2.0, whereas the steady-state throughput is true. Throughput is nothing but the number of customers reserved per unit time. Like that one can get the throughput utilization average response time or mean sojourn time and mean meeting time in the queue. So, using Little's formulas you use to find out this quantity, so this quantity you can get it from the discrete event simulation at any time as well as the theoretical result. And this is the mean number of customers in the systems, that is,  $e$  is  $N$  and the mean number of customers in the queue, that is,  $e$   $Q$ , which we have got it. So, this result, one can see the discrete event simulation as well as the, as  $t$  tends to infinity.

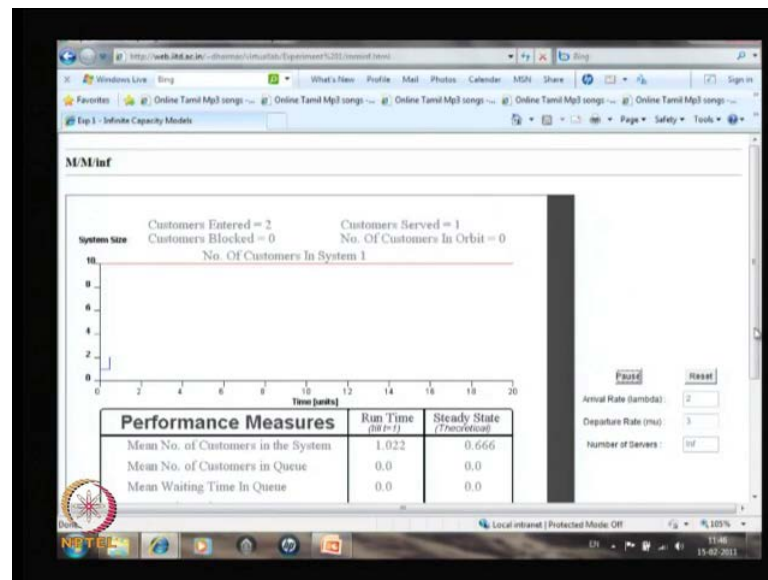
What is the theoretical result? If the conditions, conditions for the stationary distribution is satisfied and here this information is how many calls are entered, how many customers are entered into the system and how many are served and how many customers are blocked here. There is no blocking because it is infinite capacity system and we are considering the retrial orbit and so on. Therefore, here it says, the number of orbit customers is 0. This is not necessary; this is irrelevant information in the M/M/c queuing model.

(Refer Slide Time: 38:00)



Now, let me go back and do the life simulation of M/M/infinity. So, here we have to provide only the arrival rate and the service rate because the number of servers infinite, the cells service system.

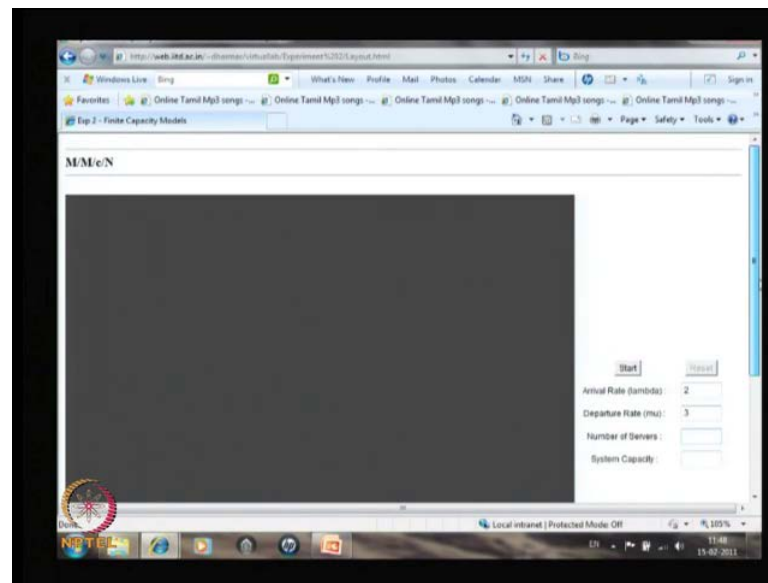
(Refer Slide Time: 38:13)



You can cross-check the theoretical results are correct and so on. So, here I am getting the steady-state probability, steady-state result based on the theory were, as this is the runtime results there is no blocking probability because the system infinite capacity. Mean number of customers in the queue, that is 0 and here also 0 because it is the infinite server. Therefore, who, the customers who enter into the system immediately, we will get the service.

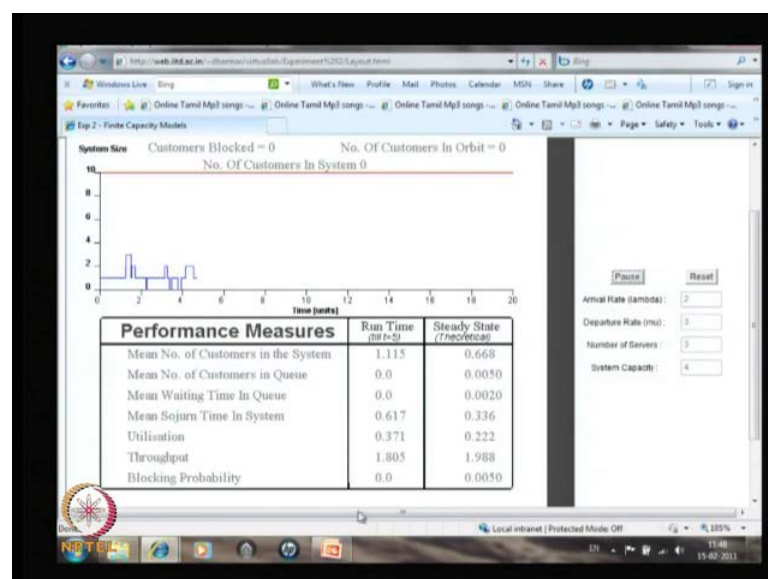
Therefore, the queue, average number of customers in the queue, that is 0 and the average time spending in the queue, that is also 0, that is correct. There is no working and the utilization is what is the probability that the servers are utilized.

(Refer Slide Time: 39:28)



Now, let me go to the next experiment, that is, the second experiment for the finite capacity queuing model, whereas the first experiment is the infinite capacity queuing model. So, in this we have, we need to give the arrival rate, service rate and number of servers. You can go for the multi-servers M/M/c/N model. You can give 1 also, you can give infinite servers also, you can give. So, here the number of, sorry, the number of servers, suppose if it is 3 and the capacity is 4, you can do the simulation.

(Refer Slide Time: 40:03)



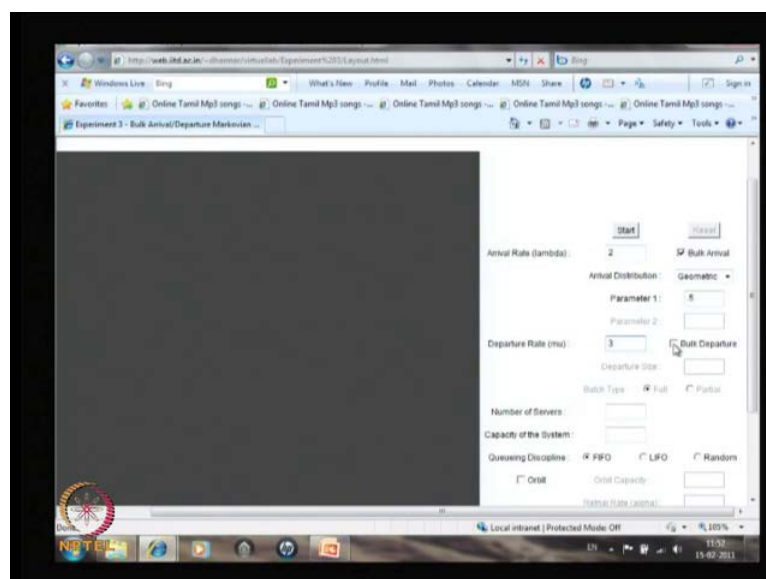
So, the blocking probability in the runtime, there is at this time if does not cross the number four, therefore the blocking probability 0 in the runtime result, whereas theoretical steady-state results is the blocking probability 0.005. So, whenever the system touches 5, then you will have the blocking probability in the runtime. So, this is the discrete event simulation.

See, number of customers blocked is till now 0, therefore you are getting the blocking probability 0 at this runtime. Suppose you run it again, run it again, let us see. Now also it does not crosses the system size by 2. So, maybe I can reset if the number of servers I can put at 2 and the capacity is 3. So, the blocking probability is this much, the capacity of the system.

So, there is no arrival after crossing, therefore still we are getting the blocking probability is 0, whereas the steady state theoretical results is 0.037. So, one can simulate with the different parameters and we can see the sample path. So, this is the sample path for M/M/2/3 queuing system with the arrival rate 2 and the service rate is 3.

Now, now we are getting the blocking probability because the system crosses, touches a capacity 3. Therefore, some calls are blocked, some customers are blocked, therefore the blocking. So, you can find out from here number of customers, 44 entered and 4 are blocked, therefore that ratio will be the blocking probability because it is a discrete event simulation.

(Refer Slide Time: 43:04)



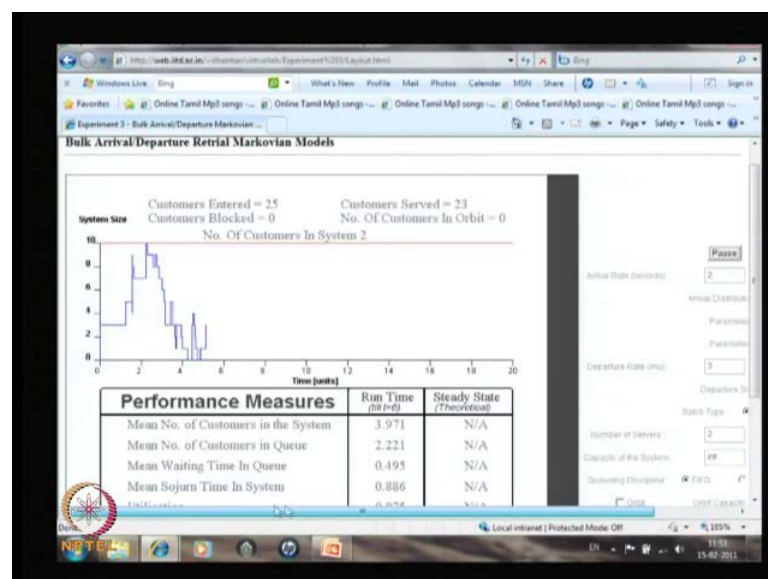
Now, let me go back and go to the experiment 3. Experiment 3, we have a retrial model with the bulk arrival and bulk service. So, now it is no more birth-death process simulation, this is non birth-death process because you have a bulk arrival and bulk service are possible and the retrial.

Therefore, let me show the simulation; you need some more information. So, what is the arrival rate you have to supply? Suppose if it is bulk arrival, then you should say what is the distribution? Whether the bulk arrival comes in a bulk, some constant number or it comes in some distribution.

You can choose it is a geometrically distributed and the parameter, parameter for the bulk arrival, parameter you can choose some 0.5. Then, the departure rate you can choose, the departure rate, it can be bulk departure or bulk... So, either you can choose the bulk arrival or bulk departure or you can choose both also.

The number of servers in the system, suppose 2 servers in the system and capacity system is 4 or we can choose infinite capacity of the system also and if you need orbit, then you have to click for orbit. And if you are changing the queuing discipline, the first-come-first served, last-come-first served are random order. You can choose the queuing discipline also according to the discrete event simulation goes. So, this not a birth-death process, is a continuous chain Markov chain simulation. With the different queuing discipline also one can go for it.

(Refer Slide Time: 44:53)



Now, once you start the simulation, yeah, since it is bulk, we have made a bulk arrival, whereas we made at, as we did not look for departure for bulk departure, therefore the customers keep going one by one whenever the service is over, but since it is a bulk arrival with the arrival distribution is the number of arrivals, that is geometrically distributed, therefore it just jump with the bulk arrivals.

Whereas, departure is by one-by-one and here is the performance measures is the runtime performance measures and we did not, till now we did not supply the steady-state theoretical results for this model and these are all the different results in the over, the runtime. And similarly, one can go with the bulk departure and you can change the queuing discipline also. So, this is the discrete event simulation sample path for this scenario.

So, I can reset and if I do not want the bulk arrival and if I choose the capacity of the system is 4, so this will be a no bulk arrival, no bulk departure. Therefore, this is a M/M/2/4 systems with a first-come-first serve and I can go for instead of first-come-first-out, I can go for the last-come-first-out also. The, the customer last entered is getting the service first.

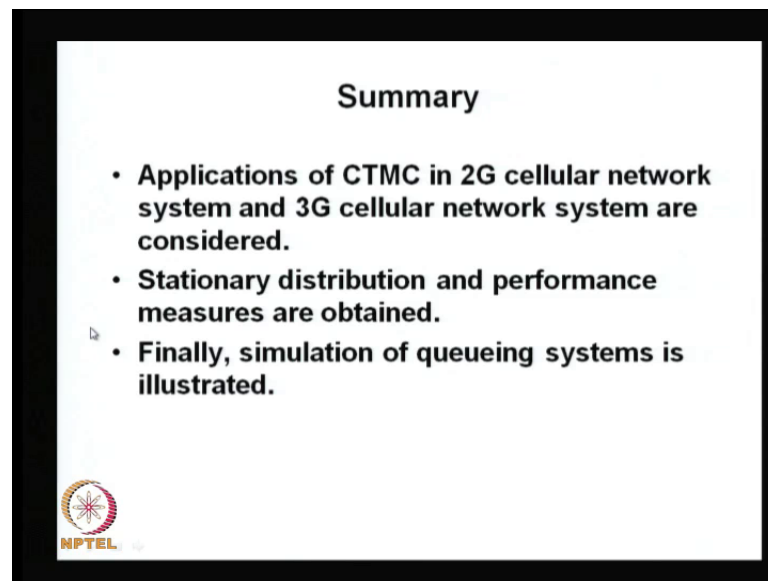
So, this is the steady-state and this is the time dependent result and I can change again (( )), that means, it is a two servers infinite capacity model with the first-come-first serve, still this is in the testing phase. So, some of the things should be removed, some of the things has to be edited. So, still it is in the testing phase.

So, this is the performance measures for the M/M/2 infinity model. So, that means, in the, in this experiment also you can remove the bulk arrival and bulk departure part and so on. You can try the simplest, simple Markovian queuing model, also one can do the discrete events simulation of that. Since if it is infinite capacity model, the blocking probability is 0 and the steady-state throughput is 2, one can find out the formula also, whereas the runtime is the... So, if, if discrete event simulation transfer a longer time, then this will converge two. Like that we can discuss all other cells also, all the results will be converged to the steady-state theoretical probability, theoretical results.

So, with this let me complete the discrete event simulation of a simple Markovian queuing model because we are discussing this under of title of the application of continuous time Markov chain. Therefore, I discuss only the Markovian queues and there

are some non-Markovian queues and so on. So, I am not discussing the non-Markovian queues at this stage, after discuss the renewal theorem and the Markov regenerative process and the semi Markov process and so on. I will be discussing the non-Markovian queuing systems also. So, here we have discussed only the simple Markovian queuing system, that is nothing but the applications of a continuous time Markov chain.

(Refer Slide Time: 49:05)



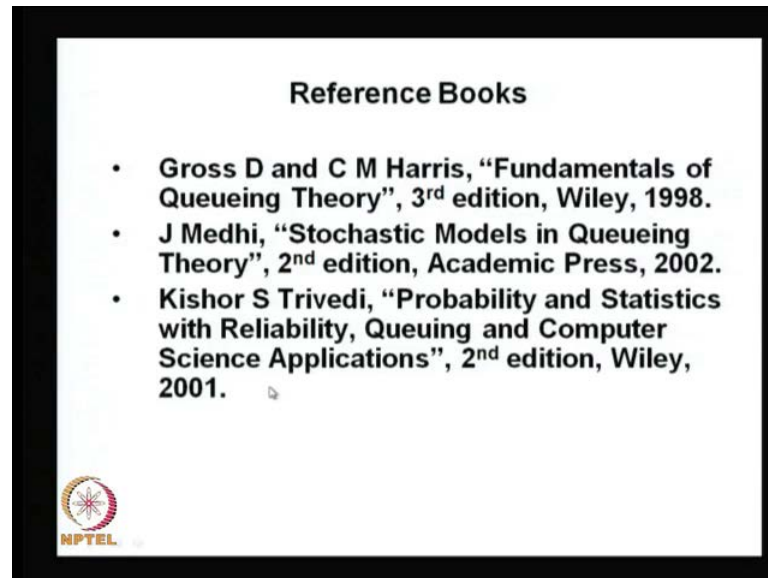
As a summary, in the last seven lectures we have discussed the continuous-time Markov chain from the definition and the properties and few simple continuous-time Markov chain. And starting with the Poisson process, birth-death process, pure birth process, pure death process, then application of CTMC in the queuing model starting, then M/M/1 queue, other example Markovian queues. Then, we discussed a few queuing networks, which has the product form solution.

Then, in the last this lecture we have discussed the application of, applications of CTMC in 2G second generation cellular networks as well as the third generation cellular networks and application of CTMC in 2 G network is the birth-death process, whereas the application of CTMC in 3 G cellular networks is the quasi birth-death process.

Even though I have not discussed in detailed the complete modeling, one can see it from the paper, my intention here is to explain the quasi birth-death process through the applications and we discussed the stationary distribution and all other performance measures for the birth-death process, comes in the second generation networks and quasi

birth-death process in the third generation networks. And finally, I have given the discrete event simulation for a simple Markovian queuing systems only. There are some more non-Markovian chain system, so that I will be discussing the other models.

(Refer Slide Time: 50:48)



The references are Gross and Harris books, Medhi books, Kishor Trivedi books.

Thanks.