

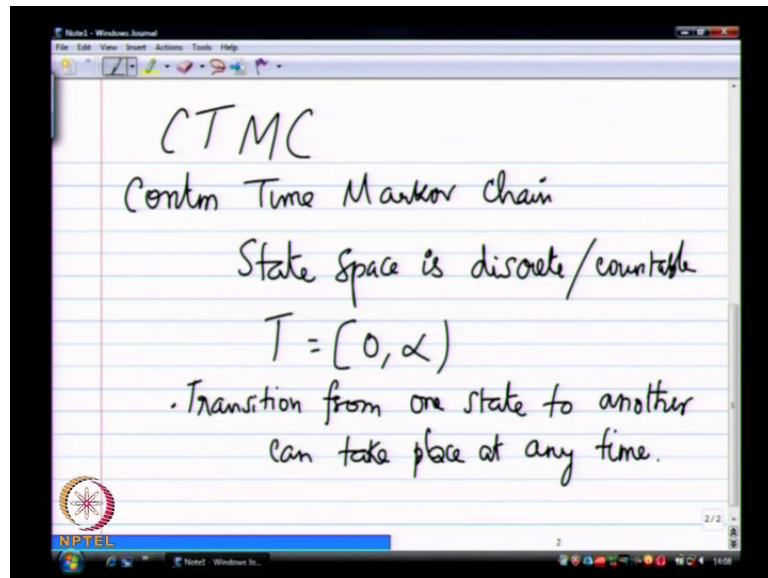
Performance Evaluation of Computer Systems
Prof. Krishna Moorthy Sivalingam
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture No. # 12 (A)

Continuous time Markov chain and queuing theory-I

So we stopped with the discrete time, Markov chain. The next part is our continuous time, Markov chains, right. So, again, I will just give the highlights of it, and then I will go along to the queuing theory; that is where we should have been last week itself.

(Refer Slide Time: 00:39)

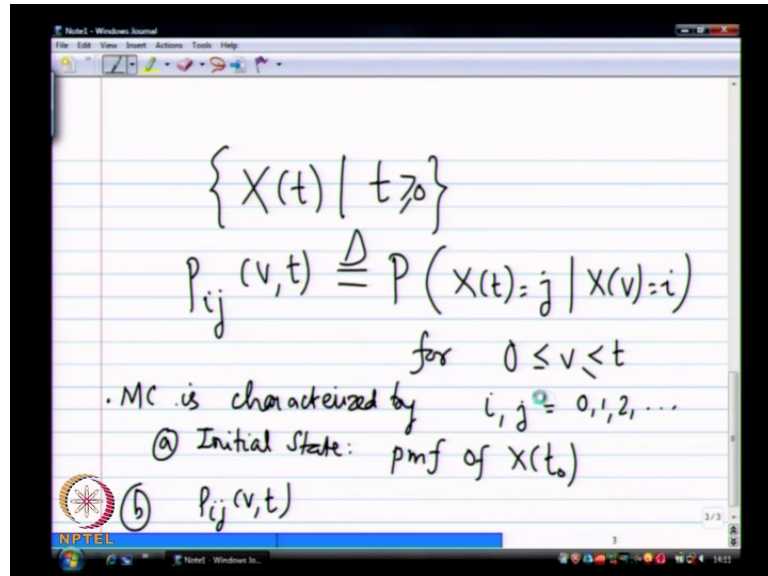


So continuous time, Markov chain; again, the state space is finite, right, or it is discrete and go to infinity, but with discrete set of values, right. And the time parameter is continuous, right. So the transition from one state to another can take place at any time; that is the basic definition right. Whereas in the case of discrete time Markov chain the transition was only at specified time ticks, time instance.

So the class, the standard definition of Markov chains again applies. The probability that you are at a current state, and the probability of going into the next state, that depends upon the current state and not any other past states, right. So or the past time... So the

future is dependent only on the present and not on the past; that is again the classic definition for that, right.

(Refer Slide Time: 02:14)



So if you want to represent this; it is basically this variable - x , right.. So x is the state of the system at time t and that can... right.. So that represents your Markov chain; the t is continuous.

Then, we talk about transition probabilities. So we used that last time, right. So here, we will again define transition probabilities first, but we will find that it is more useful to operate with something else. So this is probability or p_{ij} in terms of v and t , where v and t are time units, right. That is, what is the probability that x of t equals j , given that x of v equals i . So time v was in state i , and time j or time t now will be in state j , and this is for and even equal to is also, right.

So again, we talk about the instantaneous probability of transmission from one state to another state. So, the Markov chain is fully defined by, given Markov chain, this characterized by the initial state; in other words, the p M f of... So, the t naught is the initial time and that time what is the probability that x will take any one of these umpteen values, right.. So, that is the p M f, because x is the discrete set, right. So, state set is finite or countable. So, that defines, that is one part of the definition as to what will be the initial distribution for the system to be in one of those various states, and the second, is of course, our p_{ij} 's. Since I do not know how to change that, I will leave it.

So the book has some more other definitions, but time homogenous and all that; I will simply skip that. I just want to get to this so-called q matrix. You have heard of the q matrix - the infinitesimal generated matrix. I will just generate the matrix over here. ((Audio not clear)) So in the case of discrete system, we talked about transition probabilities; that made sense, right.. But in the case of continuous time space, we would, we actually, we would like to get the rates - the transition rates - of going from one state to another state. So, the rate is simply defined as the differential, right..

(Refer Slide Time: 05:49)

$i \neq j \quad q_{ij}(t) = \left. \frac{\partial P_{ij}(v,t)}{\partial t} \right|_{v=t}$
 \Downarrow
 Transition Rate
 Generator Matrix $Q(t) = [q_{ij}(t)]$
 $\sum_j q_{ij}(t) = 0, \forall i$

So we define for i not equal to j; so that is the differential or partial differential of the transition function with respect to t and then evaluated at v equal to t. So this q i j is referred to as a transition rate.

(No audio from 06:20 to 6:41 min)

Then, we will define a time t, right. This q t, which is....

(No audio from 06:48 to 7:02 min)

For i equals j there is slightly different definition, but I will not go much into that. So, we know the state offrom one state to another state what is the probability, but there is also a small probability will stay in the same state, right. That is looking at this very tiny instant t plus h, right h is very, very small h approaching 0, there is still a probability that it will still stay in the same state.

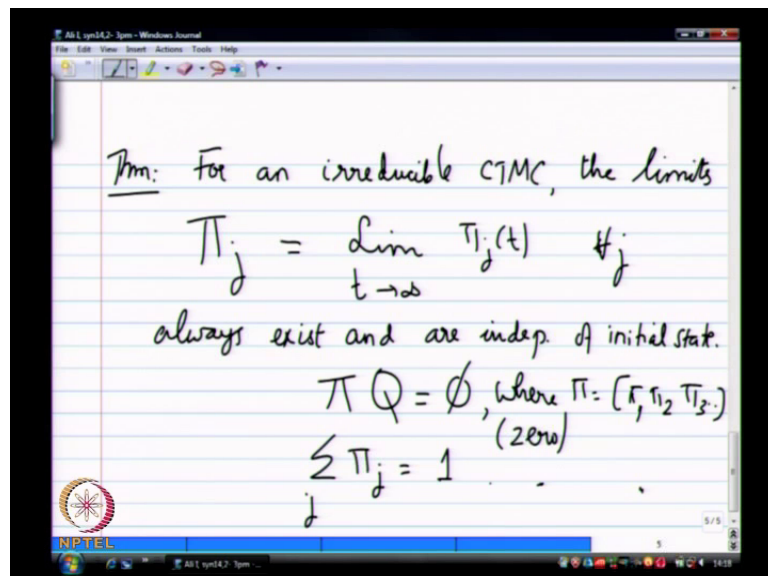
So the only condition that in this particular matrix that we have to observe is that...

(No audio from 07:38 to 7:50 min)

So from a given state, right, the sum of all the transitions rates should equal 0, which means the probability of staying in the same state will actually be negative or the rate for the staying in the same state will be negative; we will see, we can actually derive when we look at the birth-death process later on, but for now the q_{ii} will be all negative, right, along in this matrix. The diagonal elements will be all negative; there that is the rate of staying in the same state will be negative, such that, if I take a row, and sum at all the values, it will end up being 0, right. Because is a rate of transition from one... In the case of probability it adds up to 1, in the case of rate it is going to add up to 0; that is your basic balancing equation.

(No audio from 08:29 to 8:48 min)

(Refer Slide Time: 08:49)



And then, for this the theorem that we will look at by itself. Again, I am using the term irreducible without the full definition. So what I am interested in is, for this continuous time Markov chain, given the transition probabilities from which you can derive the transition rates, what is the probability of being in a given state at time t equals to infinity, right, what is the steady state probability of being in a given state. So that is again defined by....

π_j is the probability, is the steady state probability, of being in state j , right..

(No Audio from 9:40 to 9:48 min)

Technically, at every instant of time, this probability will change, but at time approaching infinity this reaches steady state, so there is going to be no variation, right. So this is the steady state probability just we calculated for the discrete D t M c's right. So this π_j 's will always exist, and this is for....

(No Audio from 10:04 to 10:42 min)

So how do we determine this π_j 's? We had in the other case π into p equals v oh sorry v into p was equals to v , right. That is the steady state probability in the case of D t M c, because there we used the transition probabilities. Here we are using this generator matrix; so π of π into q , where π is your vector, right.. And so on... and so π is

(No Audio from 11:00 to 11:22 min) the π is changing....

So that is the normalizing, right.. ((C)). So steady state probability should sum upto 1; this is the probabilities, and so I basically need to solve this equation. Just like we solved the other case, $\pi v v p$ equals v , here we simply solve πq equals.... Again, I am skipping all that. I am sorry; that is actually 0 and again if you are keen on the derivation, you can actually go through that, this book here; I am bypassing all that. I just want to get to the fact that if give you a Markov here, continuous time Markov chain, then I give you the transition rates of probabilities, then from that we can construct this q matrix and then bring over here. And then, if you can solve this and if you have solution for this, for this particular equation, then you have your steady state probabilities for a given system, from which you can derive lots of other things, right. So this is what you want to do.

(Audio not clear from 12:19 to 12:33 min)

Which one? Which one?

(Refer Slide Time: 12:41)

$$(\pi_1, \pi_2, \dots, \pi_n) \begin{bmatrix} q_{11} \\ q_{21} \\ \vdots \\ q_{n1} \end{bmatrix}$$
$$\sum_{i \neq j} \pi_i q_{ij} - \pi_j q_{jj} = 0.$$

Yes, so if you look at the steady state probability, because if you....

(No audio from 12:38 to 12:45 min)

So this is π_1 , π_2 , and so on, right **right** And this is let say q_{11} . So this is... and then this is probability of going from state 2 to state 1 and so on up to state n , right.

So this will be **your**, the rates of transitioning from.... See π_2 into q_{21} is the probability between this state and the rate which it will transition into this state, right. So the sum of transitioning from all these states. So let us see where it is? I have a separate equation for the time. Without the equation, you cannot get the expression. This is....

(No Audio from 13:34 to 13:48 min)

So given state j , right, the rate of transitioning to this state from all the other states, right. So π_i represents state of being in state i , q_{ij} is the rate of transition from that state to this state j . So the summation of all those states should equal the probability of staying in the same state, which is actually I am not defined it. This is $\pi_i q_{ij}$ that i is equal to j . So that is again a balance equation; let us say that it is the rate of which... when you are in a given state, right, the rate of entering the state, the rate of leaving the state is going to be the same. And the rate of entry depends upon the probabilities of being in the different states, and transition with this particular state, right; that is what this is trying to tell you; makes sense.

Sir q i's are actually not...

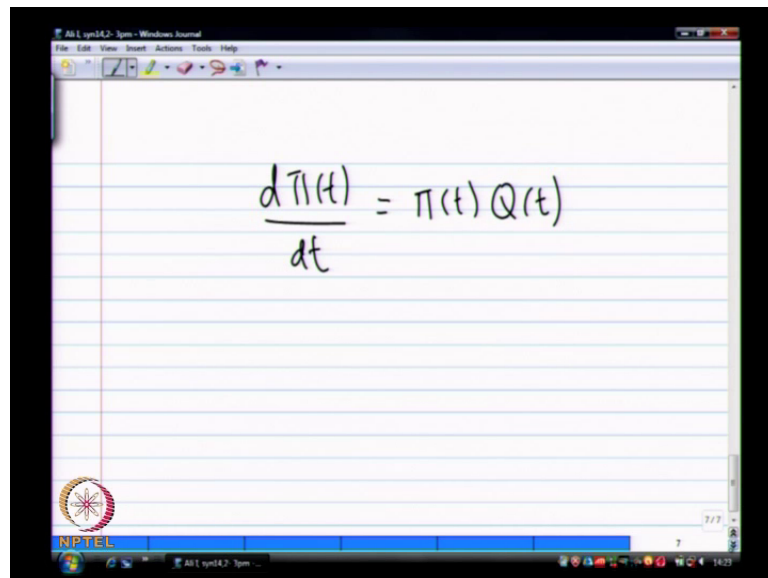
Sorry this is q jit...so what happens?

Q's dependent right...

(Audio not clear 14:45 to 14:50 min)

This q's will be time dependent; we assume it time homogeneous, and then a lot of assumptions I am skipping there. So if it is the case that is the time homogeneous system where at when, when...it is, it is time dependent yes, but if it is, then we say that the system stabilizes after some long time, then this holds good, but it need not happen. In fact, the more general equation would $D \pi(t) = \pi(t) Q(t)$. This is if you are keen, if you do not want to, if you do not want to no this, do not know this, but this will be essentially.

(Refer Slide Time: 15:33)


$$\frac{d\pi(t)}{dt} = \pi(t)Q(t)$$

So this is actual equation out of which, if pi f k stabilizes, right, to some steady state, then $D \pi$ by $D t$ goes to 0, right.. So the trade of transition of the... there is no more change, right, when pi becomes constant. So that is why pi q becomes 0, but this is the actual equation, which we simply skipped, because we are mostly interested in steady state.

(Audio not clear from 16:05 to 16:12 min)

For the irreducible or for irreducible or did I skip some condition?

Actually, it says if...yes this... that is the short proof for that, but that is... if it is irreducible, then that is what they will claim. I have to go back look at the proof for that, right. But intuition for that is.... For now, we are just assuming that if the, if you can get from any state any state, you understand that the irreducible definition is that you have a set of communicating states, right, which you cannot reduce any further. All states you can reach to any state, from any state you can go to any other state with some non-zero probability after some infinite amount of time or some long time in that case claims that it is a (()). I am sorry...

q equal to $\pi x \dots$

But q is independent...

πq equal to 0.

Yes, Yes.

But q is time dependent

So that is what I am saying, if it is steady state, then we are saying that the π changes will not be there and the q also will not change with respect to time. Again if it is a time homogeneous system, we actually look at more in terms of the time intervals than the times the time itself; the system change, the probability of going from one state another state, this p_{ij} we take $v t$ will depend upon the difference between v and t than actual t itself, right. That is called the time, again that I have skipped all those definition for the sake of not defusing a further, but for time homogeneous system these things will apply. If it is time homogeneous, where the transition probability and rate depends only on the interval elapsed, right, then, then these thing will hold good, because most of the systems we look at, sort of meet this conditions, we are generally side stepping all those definitions, because we want just get to the main research.

(No audio from 18:03 to 18:19 min)

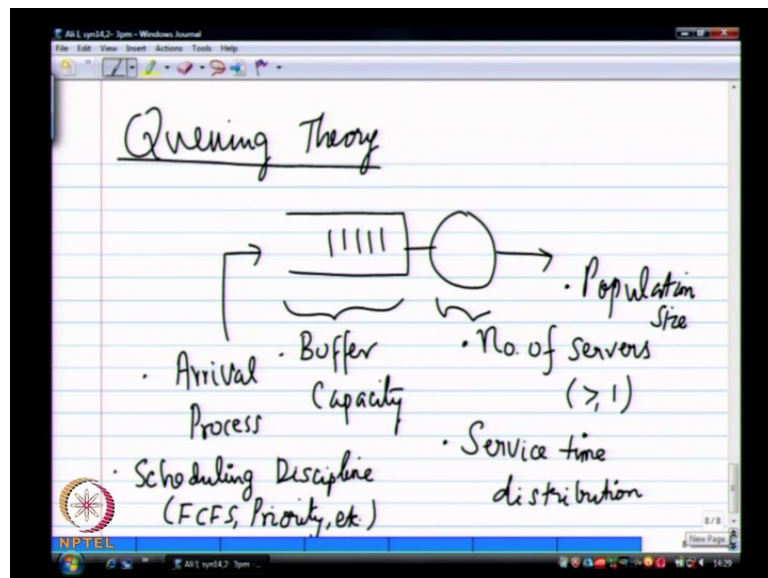
That is the first first....

So if you want to solve the....So we have now $D t M c$ for which we can solve that equation, right. v into i minus p or p minus i equal to 0 or this one, right. πq equals to 0 .

So one of these two we will have to solve for pretty much each of these systems. So besides this Markov continuous time and discrete time, then there is also, in again continuous time Markov chain, the other definition which is that the amount of time spent in every state is exponential, right. Because now if I look at my system state, we will look at m , n and q , right, to begin with; so then some of these things will....

So now let us actually get into some queuing models.

(Refer Slide Time: 19:02)



So now, we are getting into the queuing theory. Then I come up with $M M 1 q$, these things will sort of make sense at that time, as to why each state the time spent is exponential, because in the discrete $D t M c$ as I said, the time spent is geometric with probability that p_i as what we defined. Likewise here the same thing will apply; the time spent in each state in a continuous stay Markov chain is going to be exponential; and the exponential parameter will we will see how we can derive that as we will along. It will depend upon simply that q_i is the probability the rate of staying in the same state; that is what will define the parameter for the exponential time spent in a given state.

Now we are going to do a slightly quantum leap to queuing theory all of a sudden. So queue we will have define some basic concepts here.

So queue is simply represented by buffer and a server, right. So you have a few things: one is the arrival process; so that is... so a queue is characterized by whole bunch of things. So one is the arrival process, and then the second is the buffer capacity. Sometimes we talk buffered capacity, sometimes we talk system capacity; system capacity the total number of customers in the system, right, including once being queued and the one that is in service that is also refer to system

So if you look at a bank teller, then including the guy who was serviced we will say total capacity of each queue is 5, right. So then, this, as far as the server is concerned, we can classify the system in terms of the number of servers, right. It can be more than one, and for every customer what is the service time, right. So we need to define the service time distribution. So what is the service time for per customer, right. That can be either a constant or a totally random or it could be exponential or some other, right.. We are get high exponential something of that nature; we do not know this stage.

Then also, one other characterization would be the scheduling discipline, right. Usually we assume FIFO, but there are results for some other specialized scheduling disciplines also.

(No audio from 21:34 to 21:59 min)

Then there is one more characterization and that would be the total number of customers allowed in the system, which is actually the buffer capacity, no, that is already there; then we have service time distribution; there is one more, which I forgot.

(No audio from 22:20 to 22:43 min)

So besides that, there is also notion of population size, which is what I said. We will see how that is going to make sense.

(No audio from 22:50 to 23:05 min)

When we get to this population size, we will see how it is different from the capacity of the system.

Demands are coming from or request is coming from total people in system.

The total thing is total number of jobs in the system. For example, in a closed queuing network, it is going to be finite, right. And so we might have a smaller capacity, but that that no, let me go back look at that; I am not able to intuitively, right, explain that population case again. For closed queuing, it makes sense; where as in this case, they have buffer capacity, which tells me the total number of customers in the system. Besides that having one more parametric size....

Numbers of servers are more; number of servers are more.

Numbers, that comes, that comes, in your number of capacity, the server capacity

Buffer plus people getting service states, is a population sets, something like that...

I will at those....

Including the ones that are getting serviced....

Including on service

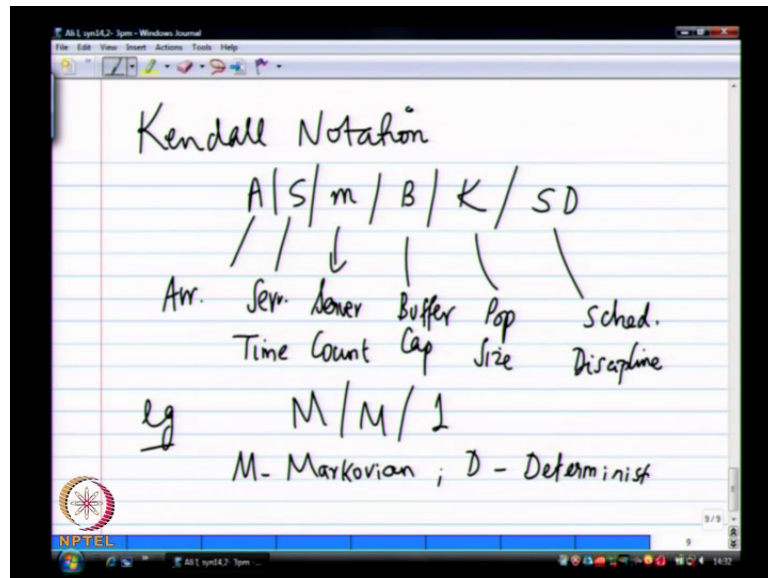
Or is it that the things that entering queue or entering from this population or something like that.

This if it is consider a system, you have a system buffer

Multiple regarding service time so this the total population it will...

The total population if I have a guess is the total set of customers that can enter the system, out of which your queue can only hold some $b + m$, number of buffers plus the number of servers, but now let me find queue where the population size is actually necessary, right, and then it will make sense.

(Refer Slide Time: 24:46)



So we have this so-called Kendall notation, right. So, it is arrival, service, some number of servers, buffer capacity and then our....So this is for arrival, this is the service time, this is the servers....

(No audio from 25:20 to 25:42 min)

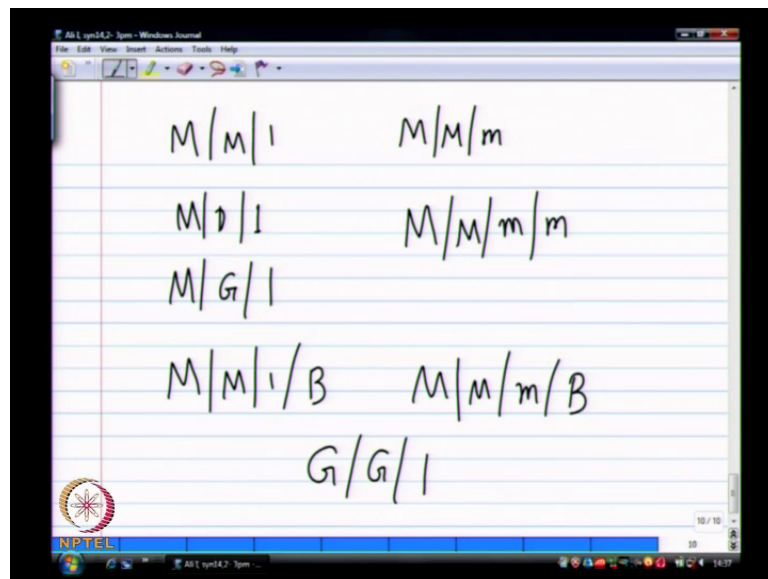
So many times, you only will only see three of them; first three is the one that you will normally see, because often we assume infinite buffers, especially theoretically analysis we want to see asymptotic capacities for performances, delay, throughput, so on and so forth. B is often set to infinity; K again is set to infinity, so it does not come into the picture; and scheduling, if it is not specified, is simply FIFO - first come, first served, right. So, for example, right. And then, we use M for memory less. So if arrival process is Poisson, then, that is represented by M because it is Markov again, and this service time is also exponential, that is what the second m stands for. So if the packets arrive according to a Poisson process and are serviced, the service time is exponential, then we have an second ((C)). And one is the number of service. So your set of classical queue is a M M1 queue, right.

So we use the terms M is used for Markovian or memory less, and then we have, D for deterministic. So deterministic is when you have fixed length packets. For example, service time is fixed, so that will be D or if your arrivals are uniformly spaced, right, there is no randomness; you have every 5 seconds a packet arrives, that is also

deterministic. So $D d 1$ will be deterministic arrival deterministic service time, right. Let us take, for example, if it is voice over, if it simply voice, voice sample gets a, generates every 100 milliseconds. So, therefore, that is D and the packets are always 100 bytes long, and therefore, you have fixed packets. So that is $D d 1$ queue and sometimes we simply call it general, g .

So you find combinations. $M M 1$, $M D 1$, $M G 1$ - these are the things that will normally we are interested in, because we can always pretend that the arrival is Poisson, and then we have closed form results for a $M M 1$, $M G 1$, and $M D 1$, that we always use as the first cut analysis. Even if you do not know if it is v 's $M G 1$, we simply need to know the mean and the standard deviation of the service times. If you can measure that over some sampling, you can simply use that and plug that values into $M G 1$ queue to get that delay results and so on. And $G g 1$ is the worst case; your arrival is also unknown and some general distribution is being followed and this service time is also some general distribution.

(Refer Slide Time: 28:11)



So you are sort of family of would be say $MM1$, $MD1$, then we go for $MG1$, right. And then here again we have some additional flavors, where I might go for MM is also more interesting, right.. Where we have looked at M servers, we saw some of these results, only the results, right, in the networks class, and the special cases where... and then or buffer, right..

So, MM 1...

So from M M 1 B we can derive them M M B, right, all those various.... We have M M 1, so M M B where B is sort of unrelated to M, but this is special case where M equals the number of customers in system equals to the number of servers, and of course, so GG 1. It is when **your arrivals to.... goes** to in a bank **define that you find that**... because you can measure the mean and the standard deviation, but there is no particular process that it is fitting into, right. And simply have those **(())**. Likewise the service time also. Service time, for example, if looking at a queue, if you look at the packet length distribution nobody is, there is not necessarily exponential service time, right. There may be some distribution is being observed and let us say only 10 different packet is possible, and for each of those there is some histogram that is given to you. And that is a histogram and that is we only measure it, that is about how you write over a one our period of finding that, and so for that, simply use the D G 1 result in terms of the mean as for as the standard deviation of what you measuring. So that is usually when you have set of samples and you cannot fit that to any particular distribution.

There is some other category, right, that has to go at the end. We cannot fit it in M D or G, right. Now if I am able to, if am able to analyze the

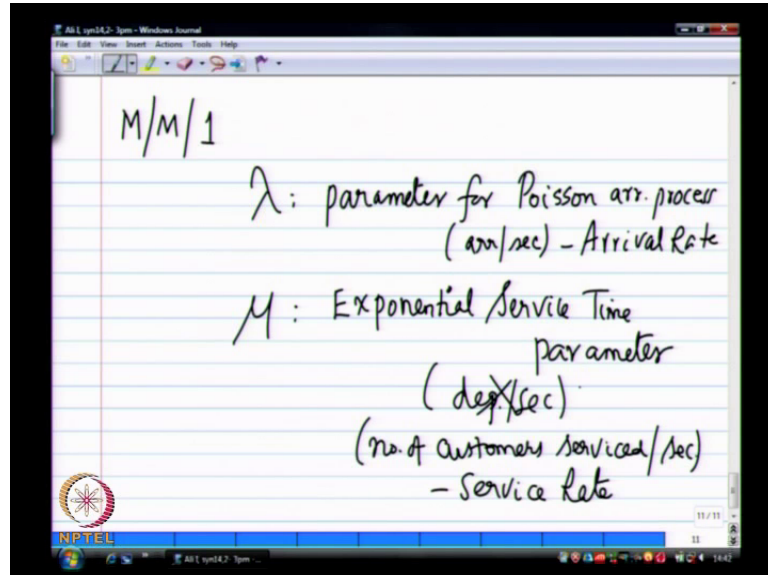
Yes, if it is able to still analyze, but what we have usually is closed form for M and D. If it is something else, like if it is hyper-exponential, then I will simply use again the mean and the standard deviation if you compute that. So even if it is a known distribution, but closed form is usually available for one of these. I do not know of closed form for other types of distributions, right. The classic ones are these. So, M G 1 could be, even if it is N or some other known distribution, but for which you cannot get the close **(())** right..

So you just want to get the average distribution; you the mean is what you are interested in, but M M 1, we can also find out other secondary parameters. We can look at the service time, distribution or the yes, the delay distribution. We only calculate average delay, right, but what is distribution of delay, that also you can define it in some cases; other some cases it is not user defined. So it depends on whether first level, first order metrics are **(())** or if you want to go into more detail, in terms what is a percentile of the delay, right. What is the, what is the, what is the probability that 90 percent of packets

will have less than, right, 100 second delay; those things are harder to define as other distributions. So that's our so basic classification of queues.

(No audio from 31:33 to 31:48 min)

(Refer Slide Time: 31:49)



So, now let us see what we can do in terms of analyzing M M 1 queue.

(No audio from 31:58 to 32:12min)

So in M M1 queue we needed figure out what will represent state of the system and what are the transition probably the rates, right. Is it continuous time? Discreet time, what is this like? So here, you have customers getting added to the queue and then customers leaving this system, right. And so, therefore, we will use these terms; so lambda is the parameter for the Poisson arrival, right. So, this is number of packets that come per second. So, in general this is number of arrivals per second.

Then mu is the parameter for the... so we will assume that the process... this is again M right; so, therefore, the service time is exponential; so this is exponential.

(No audio from 33:15 to 33:28 min)

So for an exponential distribution I used, we used μ before, because we did not want to confuse; sometimes we also use lambda, lambda 1, so on, but from now on, it is going to mu for this service; this is capacity of this server. So the mu for a queue will depend upon

two things, right.. One is the average time that this person, that each customer will need in terms of whatever, packet size and things like that; it also depends upon the capacity of the server, right, in terms of number of bits per second it can send; that is the simplest way you can look at it. It depends upon both things. Somebody would just see **their case...** So this will be departures per second, right, or service per second. So it is actually not departure per second, but that is roughly we will call that, right. So this is the number of customers...so we will not use that... serviced per second. So that λ and μ are both in same units, right. Something per second, so this is the rate, right. So this is the arrival, right. This is departure or service, rate, right. That is what we will call this. So this is the arrival rate; so this is the service rate. So the rate is basically the number of customers it can handle per second.

So, some teller will be very fast, if that is only simply cash giving teller or if it is a DD teller then more processing is needed there, therefore, there will be only three customers per ten minutes or something like that, right. So that is simply capturing the capacity of the system; the arrival right and the service rate, right, this is captured and there is only one server in the system.

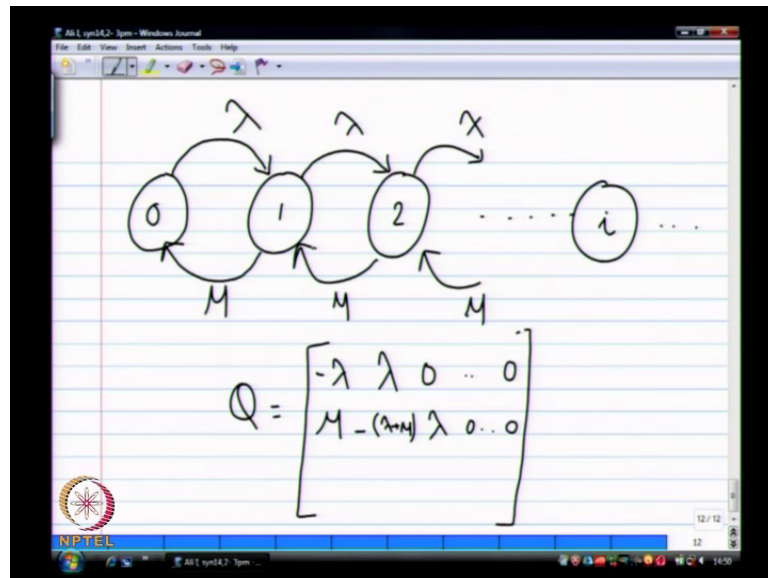
(Audio not clear from 35:18 to 35:26 min)

Where? yes, yes that is correct?

In this, we assume this only one server, so I am just looking at one particular say DD counter is one counter, that is all we are looking at. When you look at the multiple contours that we will have little bit of analogy. This is only a single counter that you are looking at and we know this counter can handle x number of customers per second; that is all.

Now for the, now to represent this as a Markov chain, right. Somehow we know Markov chain; so everything is Markov chain, right. So what should be the state of the system? What is the state and the queue 0? The number of customers; the number of customers, we saw this earlier on, right. So, the number of customers in the system is the state. So the state is discrete set. So, the possible states are 0, 1, 2, 3, with infinite queue, you can go up to infinite customers in the system. Therefore, it satisfies my Markov chain; well at least the chain property, right. That set of state is finite or countable.

(Refer Slide Time: 36:33)



So then we start of with state 0, state 1, all these fellows we will draw and this goes on up to some state i and then so on, right.

(No audio from 36:40 to 36:58 min)

Now, is this continuous time or discrete time? Only continuous time or can it be also discrete time?

I can also model

I can also model this as a discrete time system, where I take a very small interval Δt , right. Then if I am in state 0, from any state to any state we will use something like what we saw in that **(C)** probability, right. So, I need to know what is probability of getting one packet in some small Δt , right. So Δt is my time, they even call it t to be simple, right.. So t is approaching 0. Then I can also model this as a discrete time. So every t , where every one nano second we look at the system's state, right. So, mostly it will be in the same state with some probability; small probability you will have one packet, right. And the probability of getting one packet, more than one packet, in this one nanosecond is presumably very small, right.. If my λ is sufficiently small; if my λ is a million, then of course, taking it is a several packets in 1 nano second or 10 million.

So I can model it that way, right. As we have looked at the system, every nanosecond and modulate as a discrete time process. That, then we can simply have to compute this transition probabilities, then we can simply solve, right. That is one way of doing that. Then we will try both; let see if this first, this second one; so that is one. The other is to look at this as continuous time system, where at any point in time I can go from state 0 to state 1; so what is the time spent before the.... So then, we have to look at this whether this Markov property holds or not, right..

So what is the distribution for the time spent in a given state? Let us just look at state 0; there is no customer, right. So I am looking at the time required to go to state 1, right. There is some, there is what is that inter-arrival time, right. There is simply interest, when to go from one state to another state, the time is simply the inter arrival time between such subsequent customers a packet. We know that the source, the arrival packet is Poisson, and we know the relationship that if it is a Poisson arrival, the inter arrival time between, right, two consider arrivals is exponential with parameter lambda, right.

So, therefore, this state 0, I can... so this basically a continuous time. If you look at this continuous time system, the amount of time spent in state 0 is exponential with parameter lambda. So that is one part of the story that we have so, right.

So now, here, actually, we will find it easier to define the transition rate. What is the rate at which I will go from state 0 to state 1? I can simply say lambda, but we have to do some justification for that, right.. So the rate of transitioning from state 0 to state 1 is simply defined by the arrival rate, right. So with rate lambda I will be going from state 0 to state 1. We can later on derive the discrete ((C)), probably it will a little more intuitive, but let us look at this, right.. So here, I am looking at a system where I go from state 0 to state from 1 with probability lambda. Likewise, in state 1, I am servicing a customer, right. The service time is exponential. So the amount of time I will spend in the same state is simply exponential. So, if any time I can simply go back to saying that I have finished, the customer has finished service and I switch over to state 0 with rate mu.

So why do you call this the rate? Because....actually for that I should go back to the definition. But, if you look at it, the number of customers, right, no, actually this should be the probability, fine; next time I will try to work out the probability, right..

The rate of arrivals is the rate at which the lambda arrives?

The rate at which, yes, λ is rate of arrivals λ ; the rate of $(i \rightarrow i+1)$; the rate of transition from one to the other it is simply λ . So for now we will just... So to go from here to here, again it is λ ; from here to here, it is λ . So, now this is your simple birth-death process. So to go from one state, you only go to neighboring states, always, because there very right. Because only one arrival that will come, right. So one arrival comes, you go to the next state. And the rate at which an arrival comes is defined by this parameter λ , right. And likewise, you will go from 5 customers in the system to 4 customers, when one customer finishes service, and the rate of finishing the service is simply the service rate of this system, right. So, with rate μ you will go to neighboring, to the one state below. So this is...

So, but unlike the example that we saw for the $D_t M_c$ birth-death process, there is no self-loop in this particular case, right.. Why is there is no self-loop?

Rate is not 0...

Probability is not 0.

Rate...

Rate is not 0; rate is not 0; the rate summation of all the rates is going to be 0. So I am not showing the self-loop, because that is implicit in the definition that this is a discreet. So the it spins a certain amount of time in a given state with the particular rate, right. And that we will see how we can actually get that.

So this is quite basic like, right, rate equation. So if I were to look at this matrix, right. So I will, from 0 to 1 at this λ , this will be minus λ , everything else is 0. This I am simply using that this is the row elements, right. The rows of the or the sum of elements in a row will add up to 0. So, the rate of transitioning from 0 to 0 is given by to minus λ . So, the rate of staying in the same state is minus λ . The rate of going to the next state is λ . So, from 1 to 0 I will go with rate μ ; 1 to 2 I will go with rate λ . So, what is the probability of staying in the same state is?

(Audio not clear 43:24 to 43:42 min)

Now, this $\lambda + \mu$, right. So this is of course, intuitively same; same, makes sense, because the rate of being in a state, right. The transition rate should all add up

to 0, but this, this lambda plus we have some other significance also. Other than simply say - I know this lambda means this, therefore, it adds up to 0; therefore, I am choosing my mid element to the minus lambda plus mu.

So let us look at this state 2. State 2, I have two processes happening - either an arrival comes, therefore I go to state 3 or a departure takes place, therefore I go to state 1. And these are both exponential processes, right. This is with parameter lambda; this is parameter mu. So I will leave the state 2, when the min of is the, right, at the when first of these two events takes place, I will go to the, I will leave this state, which ever state I go to, lambda plus mu is, right, lambda mu, or right, so I am going to either this state or that state respectively. So if you remember the tutorial, right, min of two exponential variables, we worked it about two weeks ago, right. So if x equals exponential or x 1 is exponential with rate lambda 1, x 2 is exponential with lambda 2, right..

(Refer Slide Time: 45:03)

$$X_1 = \text{Exp}(\lambda_1)$$

$$X_2 = \text{Exp}(\lambda_2)$$

$$X = \text{Min}(X_1, X_2)$$

$$= \text{Exp}(\lambda_1 + \lambda_2)$$

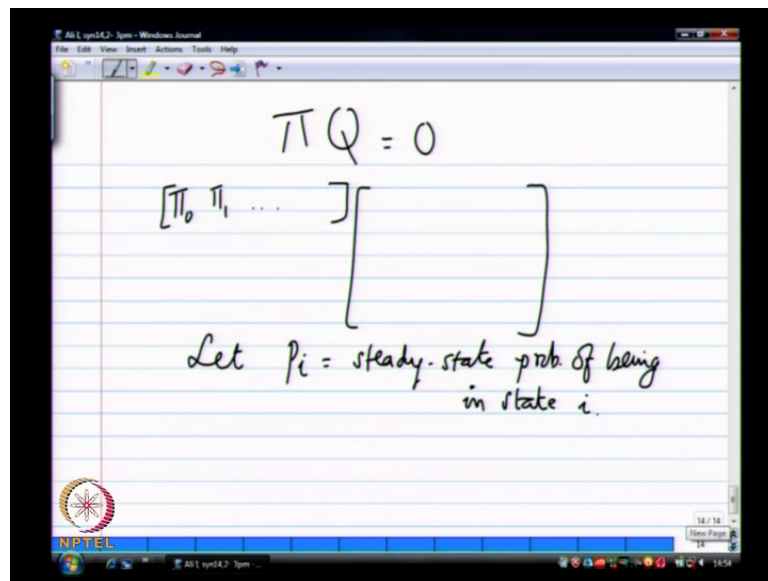
So we saw this one. If I have two such variables, right, so two events taking place, and if I define x to be the min of x 1 and x 2, right, so what was x? Actually the... we show it is also exponential with rate lambda 1 plus lambda 2, right.. So that is what...that is why here, so actually the rate of leaving this state or staying in the state, is the rate of departing is the lambda (()); therefore, the rate of staying is simply minus lambda plus mu, right.. So with rate lambda plus mu, I will either go to one of those states with that; that is what these arrows indicate, right. And therefore, because I am talking of a rate,

rate of staying will be simply negative of that. So same, just like the case here, right. $\lambda - \lambda = 0$.

(No audio from 46:29 to 46:39 min)

So now, this is my queue, right. So, assuming that steady state probabilities will exist for this particular system, I can try to solve my $\pi Q = 0$, right..

(Refer Slide Time: 46:48)



$\pi Q = 0$

$[\pi_0 \ \pi_1 \ \dots]$

Let $\pi_i =$ steady-state prob. of being in state i .

So basically I will say π naught, actually, it should be... technically speaking it should be finite matrix, but with assume it is somehow finite.

((No audio from 47:07 to 47:37 min))

I am sorry, which one?

If assume a finite ...

Finite last one will $(\lambda - \mu)$; last one will be simply...

μ

$\mu - \lambda$ minus one.

So just to be consistent with the book's definition we will use from now on, right, that I want to use that let π_i be the steady state probability of being in state i .

((No audio from: 48:03 to 48:13 min))

So, if I simply multiply that matrix and you start, right, you will see a bunch of equations popping out; one for each state, you will have an equation that is popping out, right. So, what is the equation for the first state?

(Refer Slide Time: 48:30)

The image shows a digital whiteboard with handwritten mathematical equations. The main equation is a matrix multiplication:

$$[p_0 \ p_1 \ p_2 \ \dots] \begin{bmatrix} -\lambda & \mu & 0 & \dots \\ \lambda & -(\lambda + \mu) & \mu & \dots \\ 0 & \lambda & -(\lambda + \mu) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = 0$$

Below this, the first row of the matrix is expanded into an equation:

$$-\lambda p_0 + \mu p_1 = 0$$

From this, the relationship between p_1 and p_0 is derived:

$$\therefore \mu p_1 = \lambda p_0$$
$$p_1 = \frac{\lambda}{\mu} p_0$$

Finally, a definition for ρ is given:

$$\text{Let } \rho = \lambda/\mu \text{ and } \rho < 1$$

The whiteboard also features an NPTEL logo in the bottom left corner and a system tray at the bottom with the date '02 September 2011' and 'Friday'.

So for the first state, this is p_0 up to 1; so this is μ , right. And then 0, μ , so on. This equals 0, right. So what is the... so from the... for taking this... first step of multiplication would be minus λp_0 plus μp_1 equals 0. So, this is your first so-called balance equations, right. So you... this is basically telling you that the rate should always sum up to be 0.

Therefore, μp_1 equals λp_0 equals... right. Now, of course, if we do some definition, which we kind of skipped in the beginning. So we will let, right, ρ equals λ/μ and $\rho < 1$. So your arrival should always be less than the service; otherwise, what will happen to this system? If just λ is larger than μ , then you will... all then **for** along time you will never come back to the steady state, right.. You will always be approaching infinity; there will always be more customers arriving; you will simply keep jumping to the future, to higher state, and that you should never come back to a steady state. So you will always have infinite queue, right. Infinitely large queue and the probability of being in any state is going to be essentially 0. So that this

condition is needed. So my first equation gave me this balance form, right. My second equation.

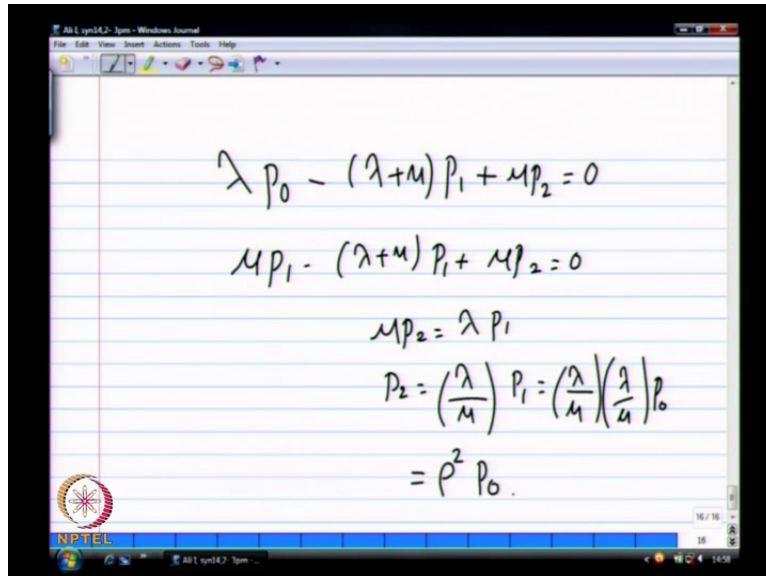
(No Audio from 50:56 to 51:16 min)

So this is my second balance equation.

(Refer Slide Time: 51:02)

So, what is this? λp_0 minus $(\lambda + \mu) p_1$ plus μp_2 equals zero. We saw time or equal to 1; so, this is μ minus.

Therefore, μp_1 equals λp_0 ; p_1 equals $\frac{\lambda}{\mu} p_0$.



The image shows a digital whiteboard with handwritten mathematical equations. The equations are:

$$\lambda p_0 - (\lambda + \mu) p_1 + \mu p_2 = 0$$
$$\mu p_1 - (\lambda + \mu) p_1 + \mu p_2 = 0$$
$$\mu p_2 = \lambda p_1$$
$$p_2 = \left(\frac{\lambda}{\mu}\right) p_1 = \left(\frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right) p_0$$
$$= \rho^2 p_0.$$

The whiteboard interface includes a menu bar (File, Edit, View, Insert, Actions, Tools, Help), a toolbar with drawing tools, and a status bar at the bottom with the NPTEL logo and a page number 16.

So, μp_1 is also λp_0 . That last μp_2 into p_1 .