Communication Networks

Prof. Goutam Das
G. S. Sanyal School of Telecommunication
Indian Institute of Technology, Kharagpur

Module - 05
Queuing Theory (Contd.)
Lecture - 20
Little's Theorem

Ok. So, today in lecture 20 we will be continuing still our quest or means discussion about queuing theory, but in the queuing theory, a specific topic called Little's formula probably Little's Theorem is something we will be discussing.

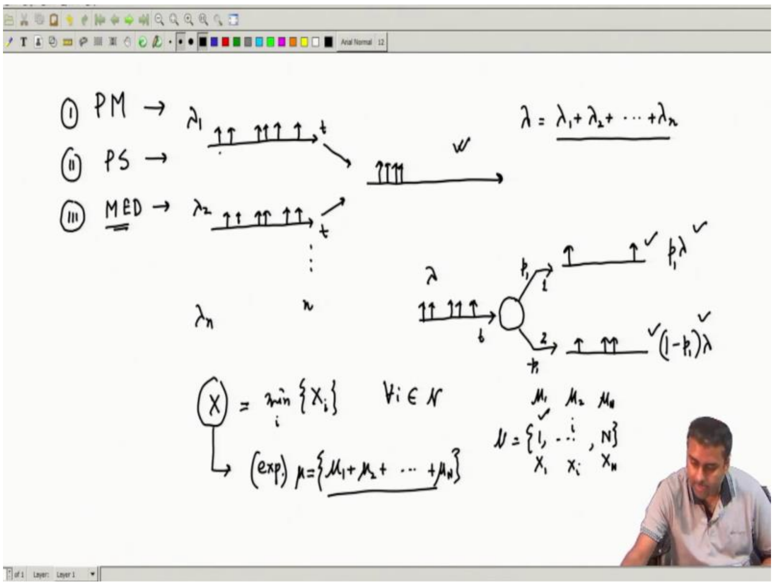
So, this is we will we will discuss in due course what that means, but before that probably we will still this memorylessness process that we discussed in the last class, we will continue on that little bit some more properties of Poisson and exponential and some more relationship with the queuing theory. So, this is what we will be doing.

So, let us see what we have been so, far doing we derive the exponential distribution we have also told how it is related to Poisson distribution. We have been told that the service and arrival process of a queuing will be dependent on this particular distribution probably at least for the analysis of our trunk switch or analysis for our voice network or TDM switched interval.

So, that is something we have discussed we have seen that there is practical evidence also that these things follow a similar kind of distribution. So, that is all done we have also derived some property of memorylessness of the exponential distributions. So, these things we have already done.

Now, let us give some theorem, of course, this is not the full course of queuing theory. So, that is why all theorems will not be proved, but let us now give some theorems associated with Poisson or exponential distribution which will be very useful in our derivation of later formulas.

(Refer Slide Time: 02:09)



So, the 1st one is called Poisson merging, the 2nd one is Poisson splitting and the 3rd one is minimum of multiple independent exponential distribution ok. So, what do we mean by Poisson merging? So, Poisson merging means it is like this, suppose let us say multiple streams of independent Poisson arrivals are happening.

So, its two streams of Poisson arrivals are happening each has a timeline and in time they are doing the arrivals like this. So, of course, the associated distribution is Poisson and there is independently another timeline is there. So, there also some Poisson arrivals are happening and these two are independent of each other.

So, when this arrival happens that does not influence this one and vice versa ok, and like that I can actually take multiple streams. So, not only restricted to two I can actually take n number of streams n can be any integer ok. If I take that and then merge these two and construct a new process; that means, for each arrival whatever time stamp I have given I will be taking one arrival then if in between there is an arrival from here that also I will be considering.

So, I will just actually merge these two things and construct a new one what will be happening to this one can we comment on this ok? It turns out it can be proven very easily that this also remains to be Poisson independent of all the input that is fitting to this one.

So, this is Poisson with the strength suppose this has a strength of lambda 1; that means, lambda 1 was the arrival rate this was having lambda 2, and so on lambda n. So, this will have a strength, of course, the strength will be all added. So, it will be lambda 1 plus lambda 2 up to lambda n.

So, this is a simple theorem of Poisson merging that if two independent Poisson processes or multiple independent Poisson processes are merged to generate one process the arrival process also will look like this, which means it will be a Poisson process which will be independent of all the inputs and the strength or the average rate of arrival will become the addition of all the input ingredients or input Poisson flows.

So, this is the Poisson merge process. What is Poisson splitting? So, like merging now I will be splitting. So, suppose I have a Poisson source that is coming over here to a scheduler ok? So, the scheduler is where I or somebody is sitting and what they are doing. So, let us say I have this Poisson arrivals which are happening in time.

Now, this scheduler picks every arrival at that time whenever it happens that it will be actually generating two flows, and how he constructs these two flows, he actually puts a probability be he tosses a coin and see what this one it is, means it's of unbiased coin sorry it is a biased coin.

So, it can have a head probability of let us say 0.3 until probability 0.7 or something like this, and from this means biased coin tossing if he sees head is coming he will put things over here in the first stream and if he sees it still he will put in the second stream like this he can also construct multiple this one.

So, then coin tossing maybe he has to do multiple coin tossing to generate multiple events or he has to roll a die, or some other things he will have to do associated means random experiment he will have to do. So, that he can stream out the probabilities, and with those probabilities, let us say with p 1 probability he puts it over here, and 1 minus p 1 probability he puts it over here.

So, it has to be done with probability you cannot just say ok the first one I put over here second one I put it over here. So, that kind of thing logic you cannot do has to be a random thing. So, when you are splitting them you are splitting them randomly.

So, with that what will be happening two streams will be generated, and exactly whoever will be present over here that will not be present over here the next one might be present over here, next one also might be present depending on the probability, next this one might be present over here so like this, it will be split.

Once they are split now these two we can again comment again it can be proven that they are also Poisson independent of each other and independent from the input. And what is the associated means rate? The rate will be probability multiplied by this lambda if it is p 1 and 1 minus p 1 multiplied by lambda.

So, again these two are Poisson arrivals independent of each other and they have a strength, which is derived from the input strength and the associated probabilities. So, this is called the Poisson merging process; that means, a Poisson this one with probability if you branch them out remain Poisson and they are independent.

If 2 Poisson processes you merge them they still remain Poisson and the strength is the addition of them. So, these two things we have seen already. Now, the next very important thing so now the exponential. So, let us say I have multiple exponential distributions and from there, I draw values.

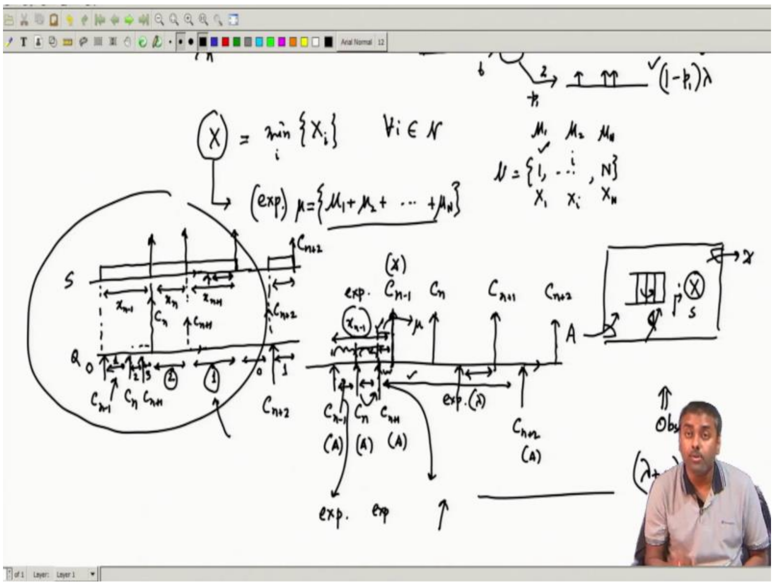
So, these X i's are taken from I sorry for all I belonging to some set N where N has N number of indexed this one. So, each one of them is actually a Poisson sorry data taken from exponential distributions. So, this will probably get at a particular instant X 1 this will get X n and similarly the i th one will get X i. So, if I know for all these I, I do minima overall I and whatever that I note it as a new random variable.

So, basically multiple exponential I take with a different kind of rate probably so, they might have this might have rate mu one this might have rate mu two and so on ok mu N. I can take them, I can take a minima of all those things that will happen to this X now it turns out again that it can be proven that this X is also exponentially distributed what is the associated strength or the rate that will be addition of all this rate.

So, mu 1 plus mu 2 should be the mu, and it is exponentially distributed. So, the minima of exponential are also exponential and they are all distributed means this is also distributed with the addition of all the individual strengths that is what happens ok? So, with this, we conclude with our special property of Poisson or related exponential distributions.

So, we know now that if we merge the Poisson process what will happen, if we split the Poisson process in a probabilistic manner what will happen, and if we take minima of Poisson sorry exponential then what happens ok you will see that all these things have application ok.

(Refer Slide Time: 10:07)



So, now let us try to go a little forward and let us try to do the same experiment that we have been doing in the queuing. So, I have a queue, I have a server and then let us say this is C n minus 1 be nobody was there so he started serving. This is exponentially distributed with some let us say x n minus 1 in between some arrival has happened and then he waits he joins over here.

So, this is C and this is C and he starts getting service. So, this might be n in between another arrival has happened, then he waits for this amount of time also his service and then at this point he goes to the service and this is let us say x n plus 1 this is C n plus 1 this is also C n plus 1 ok. So, this is how the timeline progresses as we have already demonstrated.

Now, as you can see there are multiple events that are happening so if we just mean from the outside from outside of the system. So, basically, our system is something like this we have a Q followed by a server so, this is the server this is Q and this is the whole system ok? So, this is probably two doors I am just constructing. So, through this arrival happens and through this departure happens ok.

Now, if I just from the outside I put an observer. So, what he will see internally whatever is happening he is not bothered about that, that somebody joins in the Q he moves up in the Q he joins the server and so, all those things he does not he is not interested This observer is not interested about those things.

So, for him events are some arrival is happening some departure is happening ok. So, let us try to characterize these events this sequence of events and what happens to those events ok let us try to see. So, if I just summarize this whole thing this whole thing from the perspective of arrival and departure I will put arrival in the bottom half and departure in the top half.

So, this is one arrival so let us call that as C n minus 1, when that is departing this is when C n minus 1 is departing. So, that is happening over here so this is the departure of C n minus 1. So, this I can call as D departure and this is the arrival ok, in between C n that arrival has happened over here in between another arrival has happened.

So, this is C n plus 1 then the next departure has happened over here another departure has happened over here. Let us say I continue like this no arrival has happened and one arrival happens over here then immediately because the Q is now free everybody that has arrived they are they have departed.

So, it is free so he immediately goes and joins over here he will start his service. So, this is probably this is C n plus 2 and that is where the same C n plus 2 will be going and this is where the departure happens ok just to demonstrate. Now, the next arrival so basically nothing happens over here.

So, this is where another arrival happens, and then after sometimes probably the departure ok. Now, let us try to see this inter event we are interested in the distribution of this inter events ok? So, how they are distributed can we now comment on these things ok? So, inter-event distribution if you now try to see this one. What is this? This is between two arrivals that are exponentially distributed.

So, whatever this factor might be that is exponentially distributed; that means, that is memorylessness that has memorylessness property. What about this? This is also between two arrivals so it's again exponential. What about this one? This is one arrival and one departure.

Now, what can I say about this particular part, can I say anything from whatever discussion? So, far we have made let us try to understand this particular part and concentrate on this part probably this is the most important part of our journey of queuing theory with exponential or Poisson arrival and service ok.

So, let us try to understand what is this. So, the first thing is this is where the last arrival has happened and the next arrival will be happening over here that is exponentially distributed with strength lambda. What is this? This is when this is about a service that has begun over here.

So, this was x n minus 1 and this is the first time that is departing. So, this is again exponentially distributed, right? So, this x n minus 1 is exponentially distributed from this point to this point so that is exponentially distributed this is also exponentially distributed. So, basically, this is actually part of two exponential distribution this one and this one, but for both cases it's partial.

Now, the partial property comes with the memorylessness property. So, if this is exponential from this point to the rest of the things, we have already discussed in the last class that the residual part remains exponential for exponential distribution, of course, we have not derived this we have just done the memorylessness property from there we could understand that ok.

But exactly why that happens that requires means we need to do renewal theorem and then we need to derive the residual distribution, but that will come to exponential with the same parameter as whatever parameter it has the original one has. So, this is something we have demonstrated so, this will still have distribution with mu so the rest of the portion ok.

What about over here? Over here also so, when this is happening. So, basically, as you can see this particular point can happen either there can be one arrival or there can be one departure so basically all the time what is happening from a particular time instance.

Suppose from here if you even try to see this guy what is happening one service is going on. From here that service time is exponentially distributed that has not finished in the future somewhere that will be finished. From here one inter-arrival has started which is also exponentially distributed with parameter lambda.

So, among these two there is a competition whichever finishes first will declare an event, if the arrival happens first; that means, he is still in service once arrival happens then that will be the first one as this event has occurred. From here to here one service is still in service another arrival has occurred ok.

But from here to here what has happened? So, basically, one service was going on that is service has finished before the next arrival has happened. So, basically, two exponentials are having now competition, and whoever is the minima among them now that principle of minima of exponential distribution comes into the picture. And what will be this distribution then? This is still exponential which is the beauty of it and with a parameter which is lambda plus mu.

That is the beauty of this whole thing. So, every time whatever happens we see that it is all exponential. The next event is again exponentially distributed; that means, all the event separations are actually memoryless. Therefore, from the observer’s point of view it does not matter where I start counting I can still say the rest of the things are all exponentially distributed whatever happens.

So, that is the beauty of this queuing it is a, very fundamentally different thing from all other things that we know. So, that is why it can be actually described with the number of customers they are there because at any point in time if you go in you will be able to see how many customers are there inside. So, over here as you can see at this point if you try to target.

So, basically, there were 0 customers over here from here to here it actually has 1 customer in the system from here to here it is having 2 customers in the system from here to here ok. So, it's actually having 3 customers in this system their one departure happens so it again comes down to 2, here it comes down to 1 and here it comes down to 0, and again this is 1.

So, this evolution of the number of customers is happening because the intermediate part whatever is remaining that still exponential with the same parameter therefore, I do not have to characterize how much time has elapsed and how much time is left. So, those things are not required.

Every time just this number I can keep on recording and the next thing that will happen because of this exponential memory-less property will be only dependent on the current number and not previously what has happened.

So, all that history is not required I can summarize the whole thing just concerning these state variables single state variable, that is just the count. A number of customers that are inside ok who are still waiting for their service they might be waiting over here taking already the service or they might be waiting over here for their term to come.

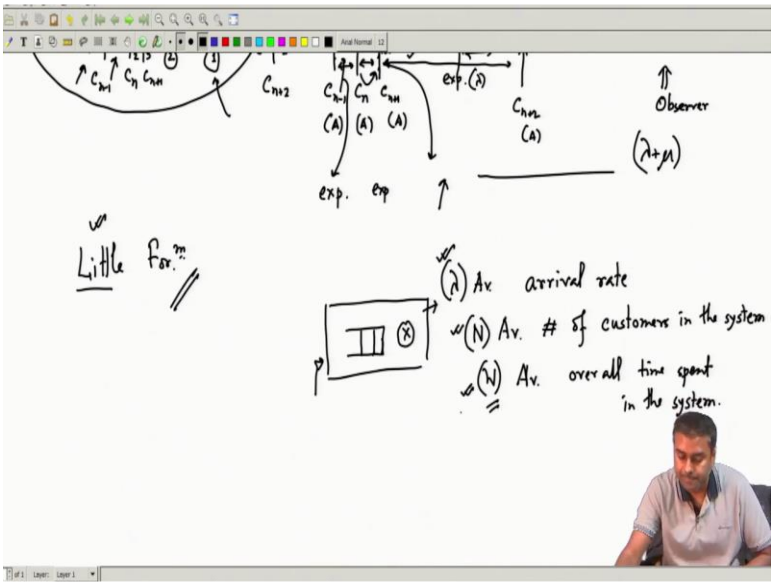
Whatever happens, I do not discriminate also among those customers or among those packets if they are packets. I just count the overall number and I comment on that, this is the beauty that will be happening as long as I take this exponential distribution for arrival and for sorry for service ok.

These are both things if they are exponential then we have understood because of this nice property of exponential distribution that minima of exponential are exponential, and exponential has a memorylessness property. So, therefore, from any point in time, you cut the exponential, and then the rest of the things are still exponentially distributed with the same parameter.

So, basically, I have all this information and I do not need any other things to describe this system. There is a whole interesting aspect of queuing analysis or Markovian analysis that you will learn later on. So, I hope this is clear to everybody so with this I can actually probably go towards the first part of our.

So, far we are just discussing generic aspects of distribution how they interact with queuing systems, and all those things. Now, we can go ahead and derive some more things about the actual queuing.

(Refer Slide Time: 22:28)



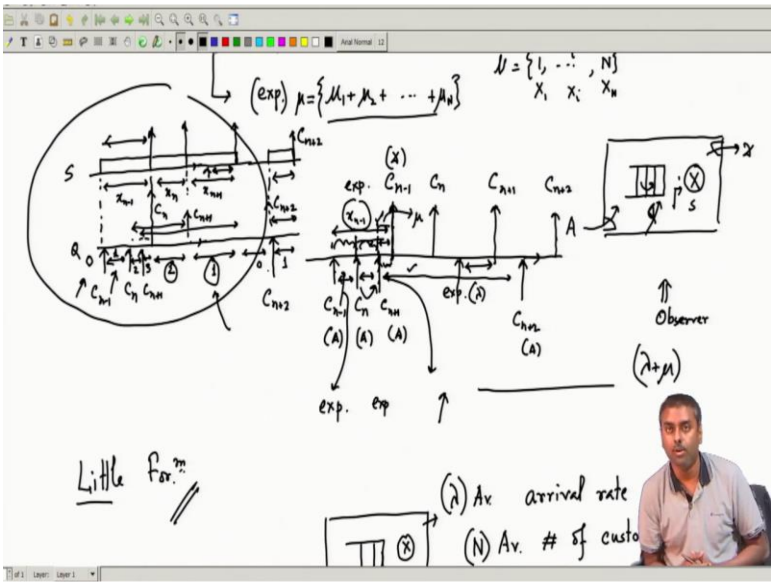
So, what we will do, we will have promised we will first derive Little's formula. So, what is Little's formula? Let us try to understand this part. This is actually telling that suppose I have a queuing system followed by a server it might be a single server it might be multiple servers.

So, how these servers are serving them might also mean you might have some particular way of serving them that you have a priority and all those things, it is all these things can be there even queuing discipline also can be anything it might be random it might be LIFO, it might be FIFO, it does not matter it's it might be smallest job fast and all kinds of things can be happening over there.

So, whatever happens, what I am trying to understand now is can get some average relationship between 3 parameters, one is the arrival rate the other is. So, this is the average arrival rate then, the average number of customers or packets or calls in the system remember in the system. Whatever we call a system we will also talk about that and the average means overall time spent in the system by all these customers ok?

So, can I get a relationship? So, we will represent this we already this we have characterized that is lambda ok. So, the customer in the system will be represented as N and this will be represented as W waiting time ok. So, basically, this means actually the customer enters over here and whenever he departs it is the time between two the gap between these two is the overall system time that he requires to get service. So, if I just see in my these things.

(Refer Slide Time: 24:57)



So, for the 1st customer, this is the overall system time, for the 2nd customer this is the overall system time, for the 3rd customer this is actually from here this is his arrival this is his departure so, that is the overall system time, for the 4th customer this is the overall system set. So, like that if I do bookkeeping I will get all this system time, and then I can take an average over a very long duration. So, I can observe this whole thing for a very long duration and I can take average over them.

So, once I take those averages I will be able to get that value of waiting time. Again number of customers so basically you have seen for this over here it was 0 for this amount of duration it was 1 we have already recorded that for over this duration, it was 2 this was 3 this was again 2 this was one 0. So, like this, we can take and do a time average.

What is the average means, what is the average number of time average amount of time over the whole observation period for a very long observation that I have seen him in state 1 state 0, and all those things and I take the means I take associated probability and then I take the average value of that. So, I get the average N for these things.

So, the average number of customers is generally in the system present in the system so, this is not instantaneous it is the average taken over the whole duration. Without doing any queuing analysis can I give a relationship between these two, it turns out and was proven by Little that this is achievable this can be done.

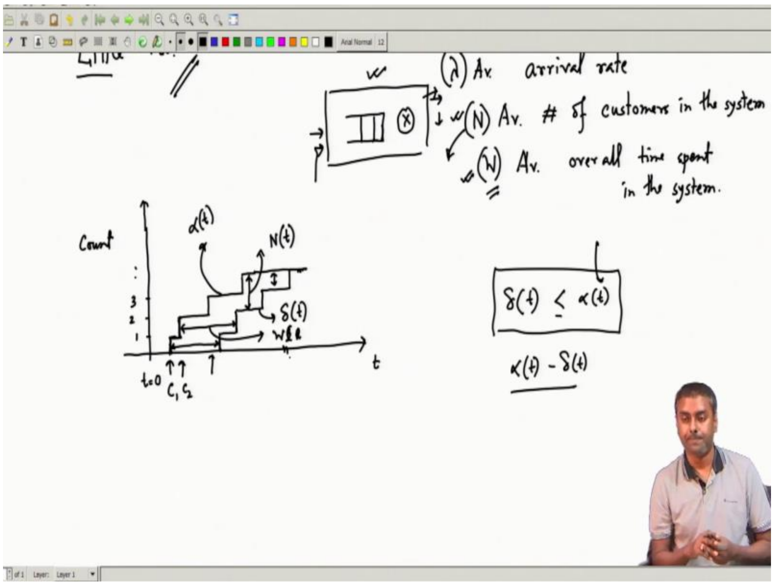
So, let us now try to see how we derive it without doing any queuing. So, no queuing analysis no means whatever the arrival rate that might be exponential that might be other things also means inter-arrival might be exponential inter-arrival might be any other distribution service might be any other distribution. So, we are not restricting ourselves right now, when we started we started restricting ourselves, but right now we are not restricting ourselves we are trying to derive an average formula.

So, we are not going to discuss the whole statistics just an average formula that will tell us the relationship interrelationship between these average quantities. This is a very strong tool because if one of them will be generally given if one of them you derive the next one you do not have to derive that that is the beauty of it.

So, basically, statistically, you derive the whole thing and then you take the average this is generally what we do right, but over here what we are trying to do is we will try to develop a relationship between them so that one of them I derive the other one is already means readily available because of this formula. This is a strong formula because it is true for any distribution of arrival and service process ok.

And irrespective of any queuing discipline that will be putting over here or any kind of a number of servers you put all those things how the servers are serving so, all those protocols that you will be implementing inside its not dependent on that, that is the beauty of this formula and you will see when will be deriving it is not taking any such things into account. So, in this class what I will give? I will first give the sketch of this derivation and next class I will solve this one.

(Refer Slide Time: 28:20)



So, again we will just help ourselves and do bookkeeping again event bookkeeping I call it. So, in time you are just trying to see, but over here we will draw instead of just even bookkeeping we will do a little bit more informative event bookkeeping actually, and we will also count the events that are happening ok?

So, over here I will be putting the count over time and this is the time. So, as time evolves we will we will keep that count. So, basically, what do I count so, be particular things from the system perspective there are only two things one is the arrival which is the event and the other one is the departure.

So, these are the only two things I will be counting over here. So, let us say from time t equal to 0 I start at this point one arrival happens. So, immediately this is the count so basically the arrival count will go up. So, it will go up by 1 so this is an integer this one is 1 2 3 and so on I am doing so this one will go up.

Now, this is where C 1 happens this is where probably C 2 this next arrival happens immediately the count overall arrival count I am just counting right. So, the count will go to 2 this is where probably the 3rd one happens. So, the count will go to 3 and so on it just goes on like this wherever it goes. And this will be as you can see it will be monotonically increasing because the count will only increase.

If nothing is happening for a longer duration it will be flat whenever another one another arrival happens immediately it will jump. So, it is such a monotonically increasing staircase kind of thing that goes to infinity right? So, it just keeps on increasing so, this is what we call the arrival curve or arrival count curve over time.

Similarly, what we will also do, is count the departure. So, C 1 comes over here probably let us say C 1 departs somewhere here. So, then whenever it departs the other counter from here it will just keep counting that and it will increase the counter. So, that is the first departure which has happened it is a that is the next departure this is the next departure ok this is the next departure.

Now, let us try to understand. So, this is what I call it this is arrival this is Delta t so departure count. Can I say anything about these two alpha t and delta t? What is the relationship? As we have seen alpha t and delta t can touch each other that is possible this is where it has happened; that means, as the count is the same; that means, as many have arrived many have departed. What does that mean?

So, in this instance, there is nobody in the system, everybody who has arrived has finished their service and has departed and no fresh arrival has happened so the system is going empty; that means, nobody is there in the system. Can this be happening that the delta t curves overpower or it goes beyond alpha t that can never happen, always delta t must be less than alpha t.

This is because as many is a count right if the arrival count can the arrival count be less than the departure count that cannot mean as many customers have come only they can depart. So, there cannot be new things that will be generated from the system that will be departing it cannot happen.

So, therefore, this delta t can never go and surpass alpha t this will never happen. Occasionally they can actually close down their this one ok. Now, some more observations over here. What is this thing? The height of this one that is actually exactly tells how many are there in the system, that is over time how it is evolving. Can you see this?

Because at this instant this difference says it's equal to alpha t minus delta t what is that actually, as many have arrived minus as many have departed rest of the things will be there in the system. So, that is actually I can call that N t over time how many are there in the system? So, every time instance it will give the count good, so this graph is very useful.

What does this give me? This actually tells me this is the first arrival and the corresponding second sorry first departure, this is the second arrival and this is the instance of second departure. So, this is actually the waiting time of let us say the nth customer this exactly gives me the waiting time of the nth customer.

So, therefore horizontally and vertically we can actually make some sense out of it in the next class what we will try to do we will try to take these things into account and try to derive the relationship among themselves, we will try to see that they appreciate that in the next class.

Thank you.