Indian Institute of Technology Kanpur

National Programme on Technology Enhanced Learning (NPTEL)

Course Title
Digital Switching

Lecture-25

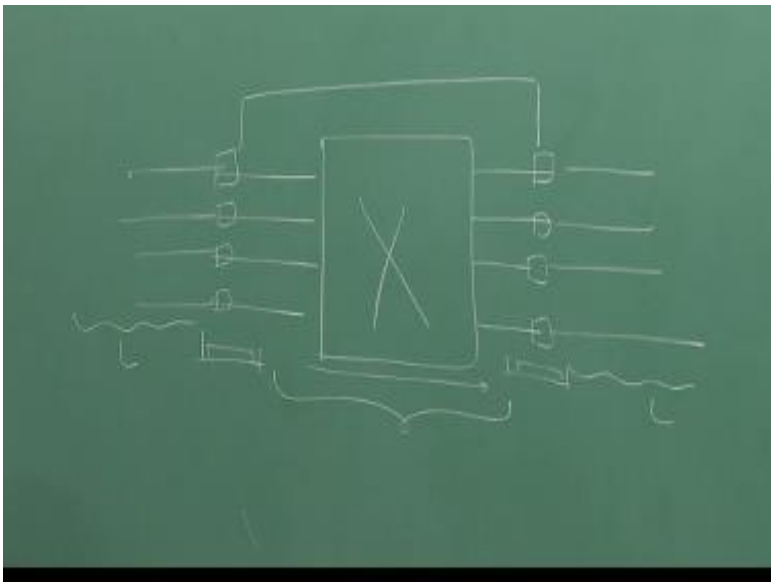
by
Prof. Y. N. Singh
Dept. of Electrical Engineering
IIT Kanpur

So in the last video what we are actually talking at was a cross bar as a packet switch and I have actually told in the end that there will be interface boards as well as the output ports. And all packets at the interface input interface boards will be analyzed for the header, \and this analyzing formation should be passed to the controller and this controller will essentially will now be regulating the cross bar at every time slot.

And as a result the packets will be now moved on to the outgoing ports and then there will be transmitted further, to other switches or the users whoever will be connected. So, in this case, the way I have assumed was that this cross bar, which is being used to do the switching, is operating at the same speed as the input line or the output ports or output lines. So one slot takes one packet here, actually take one slot to transmit packet all the way to output port.

So internal speed of this is same as the line speed, so that was the assumption, but switches actually can operating slightly different fashion, I can actually build up, if I put I can actually show the interface boards here, what I can do is, if I actually provide this input and output interface boards, so my line rate is different, is going to be same but from this interface input interface board to the output interface board, in this switch I actually can transmit at higher rates.

(Refer Slide Time: 02:53)



It need not be the same rate, of course certainly it cannot be smaller it has to be there, equal to the line rates, so line rates are being prevalent here, I call this line , so I can actually operate at higher speed. So suppose if I operate at double the speed, then what is going to happen? So it is operating at double the speed, in that case, in the slot period, the slot period here in switch, we offer a half duration, half of the this particular duration which is at the input or at the output lines.

Internally this will be two slots, so these are technically means, if I actually have ,I can transmit, in this one slot I can transmit two packets to the same output. So that should be feasible, so to find this particular rate, the factor there is a k times the line rate, so I call it the speed of factor. So this speed of factor, if it is to this minute, two minute slot is used within the same period, for transferring the packets from input interface to the output interfaces.

Which actually means one slot period this cross bar can be configured two times, so let's the simplest case is when k=1 which was the original case, so we will now be talking about when k=1, when k=n, where n is the number of ports here and by n switch and k can be something

which is greater than 1 and less than n, so there are three cases. So let's see the k=1, what is going to happen.

So will define something called art of the line blocking for this particular case, so how do the line blocking happen in this case and in this case, it will never be happening for k=n, so when you have this case, technically the interface board will be like q, it will be storing the packets, so I am showing four of them, the four lines and the packet which are there in the front, there headers has been communicated to the controller, and this headers controller will now configure the cross bar.

As a result the packets from here now will be switched out to the outgoing ports, for example when this is going to happen is when each one of them want to go to a different outgoing ports. But there is a possibility that, that these packets which are in front may be two of them or three of them would like to go to the same outgoing port. Then what is going to happen, so in this case for example, these are two packets or three packets which want to go to the outgoing port 1.

So when this address will be analyzed by the switch by this controller, it will configure the switch but it also has to decide that all this three cannot be pushed to the outgoing lines at the same time. So two of them have to be stopped to these two interface boards, have to be told that don't transmit, only one of them will be permitted to transmit, so there is going to be a feedback, which is going to now these control boards, so also.

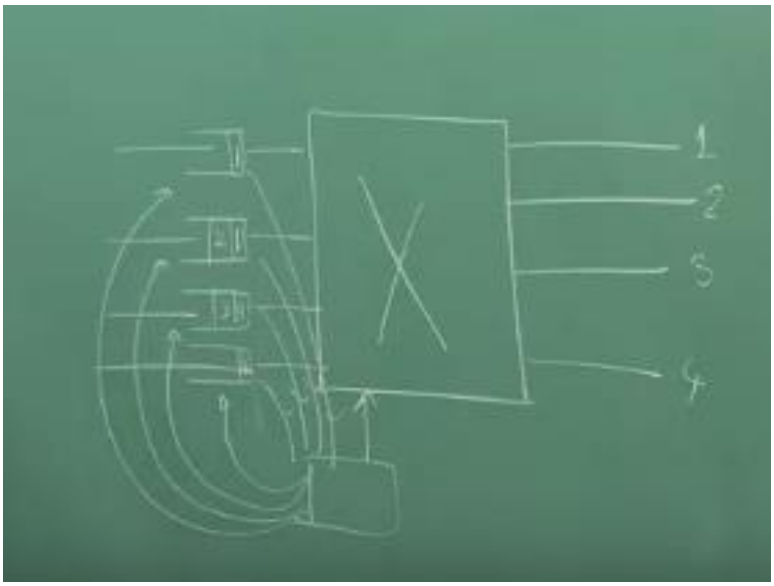
So there will be essentially now, a throttled or regulated by the controller as well the switch will configure and these header in formation are also need to come back to the controller, so controller has to do the computation and based on that it will decide what is going to be the next state of the switch and what should be the, it will give a command to interface board whether a transmission should happen or should not happen.

Now in this scenario, when three packets, which are at the front , which are supposed to be transmitted to the current slot, want to go to output 1, only one of them will be going out, so but there may be possibility might have 2 and 3 here, this might be a 4, so 4 goes to 4, there was no

conflict for 2and 3, I could have transmit 2 and 3 to the port number 2 and 3, but I could not do so, because I actually take first come first serve policy.

There is only one single q at the input, so then there is no transmission here, there is no transmission here, there will be transmission here, there will be transmission here, packets will go out so this particular packet will move further and this packet will be in rest, so whatever was x will come here ultimately, whatever was y will come in the front ultimately, this will remain as it is, 2 and 3 could not be transmitted.

(Refer Slide Time: 08:28)



Now this is happening because of, there was some, what we call, packets which is going to have a contention and that packet was lying in the front of q, which has not allowed the packet which was not having contention at that slot, that packet could not be transmitted. So this king\d of blocking is known as header blocking. We simply usually simply write as actual blocking, so this is what was happening in k=1.

Now interestingly you can also observe, the movement packet is transmitted, this transmission is happening at the transmission rate as these lines are operating, so packet will come and they will

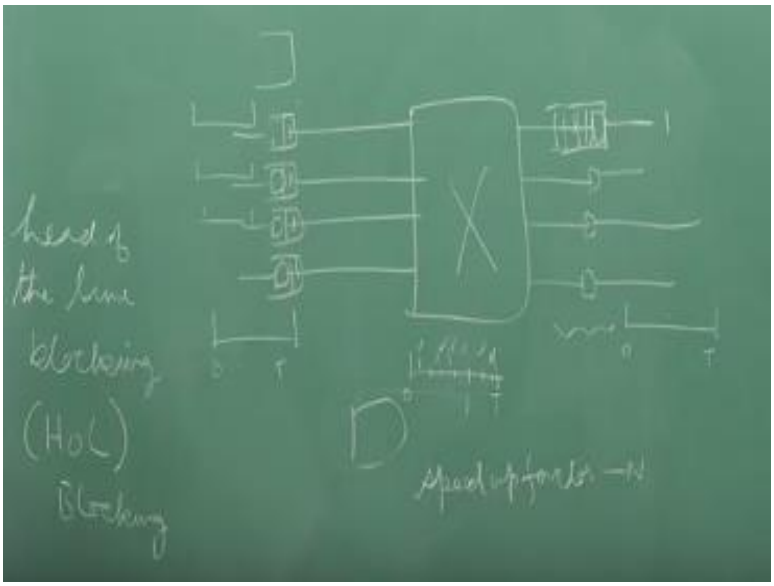
be immediately, transmitted out to the next switch or next user or next network or whatever it is. As a result I need to buffer at the packets and the inputs, there is no buffering at the output, So I need to essentially queue up or store the packets at the input.

And this need to be stored , because of this art of line, this does not happen, if we can guarantee the transmission, that all four packets which will be coming, will never ever be having a conflict, there will be never be content\ding for same port. In that case you only need one buffer, because whatever will be coming will be transmitted out, and next slot next packet will be coming and it will be transmitted out.

Better it will be a lie and does not happen, we need to have a queue and this queue actually build up to very larger queue, in fact if we analyze that what's the maximum throughput which we can achieve, in this kind of case because of this art of the line blocking, I cannot certainly operate at 100%. That can only happen if all the packets which are coming in the same time slot are not contending for single, for same output port, they are all, being mapped to different outgoing ports.

This permutation kind of thing which is happening out there, but that we cannot guarantee, now let's look at the second case k=n, what is it mean? That the switch, these are the interface boards, I do require the interface boards in this case on the outgoing side, so this is switch is operating, this slot period here, this same slot period now is being divided into, n sub slots. So you will receive one packet in this time period, in this same time period, the transmission from the interface boards and packets can actually be transmitted.

(Refer Slide Time: 11:11)



And these can be received here, but then there will be again transmitting only one packet, so in a time slot when the packets will be arriving, so there is a queue here, which gets formed, so I can actually draw a queue on the board, so even if all packets are want to go to the same outgoing port. Since there are n possible slots, even if all n packets which are there at the art of queue want to go the same port, I can do so; I have to just say the controller will get all this information.

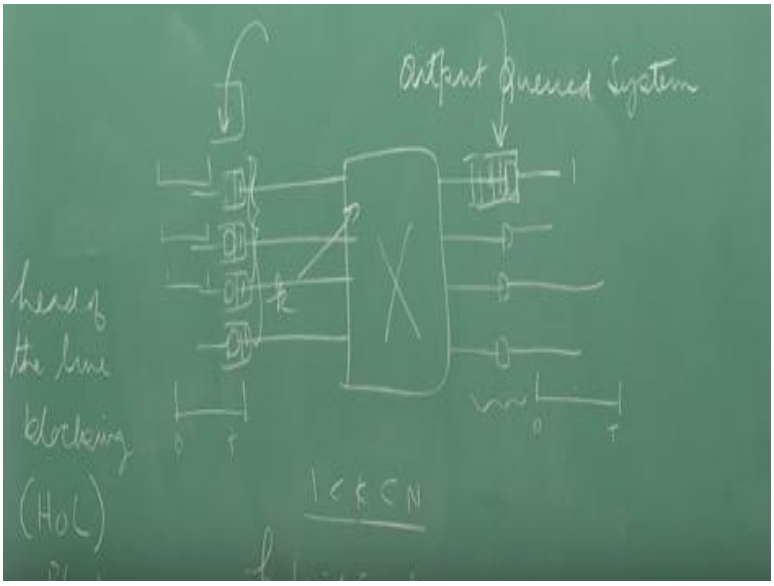
It has to tell all of them back that they have to do it, in this mini slots they have to the transmission, so all of them will be given separate sub slots or separate mini slot, because this system is operating at much higher speed, at n times the speed. So all this packets can be transmitted and they will be received here. But your receiving at a faster speed, you cannot transmit at that rate to the outgoing line.

You have to transmit only one packet in the full slot, not n packets in a slot, so as a consequence you do require, a buffer or a queue here, so all this n packets can arrive and the one by one, you will keep on will be transmitting out, but do you require a queue here, you actually don't require, in all the synchronized slots all the packets comes in, all of them will be pushed out, queuing will be happening only at the outgoing line not at the input.

When k=1, when the speed of factor is n, so when speed of factor is 1, you will end up in input queued system and when k=n you will end up in a output queued system, there is no art of the line blocking in output queued system. Another important thing is the line rates are same and so far your probability of arriving here is less than 1, and there is a uniform routing, so there is a equal probability that packets from port I will be going to any one of the ports on the output side. So these are equally likely happening.

So because of this, this system will always remain stable, so whatever the arrival rate and departure rate, they can most be 1 and they will be same, arrival and departure rate is statically, so this queue will always remain bounded.

(Refer Slide Time: 14:22)



It will be always finite and stable system, you can also have a case when k is not equal ton, so this particular scenario, what actually was happening is k is not equal to n and k is not also 1. So far k number of packets containing they will be able to push the outgoing to buffer, but the moment you have k+1, contending packets for the same outgoing port, so k + 1 will be passed on to the buffer and the remaining one's have to be still.

You have to hold them in the input buffers, you will requiring in this scenario, input buffers as well as output buffers, but normally the chance is that, all n, if n for example, 16/16 switch, all 16 packets at the 16 buffer would like to go the same outgoing port that chances will be very low, may 2 or 3. After that the probability will keep on falling down drastically, so even if you k=2 or 3, the buffer requirements at the input can be drastically reduced.

You will require the outgoing output buffer system, it will work as well as output buffering system actually, so and this technically not input queued not output queued, this is a hybrid system. And very soon even if the small values are of k, this will be working well as a output system. So let's come to the input queued system, in this case there will be k packets, which can be directed as same port.

So let me just clear it up, so no more an output queued system, I am considering a input queue, so you will have, an out of these, the k of them are being directed to the, one single port. So there will be k packets which will be coming, at any point of time, and remember, sorry this k is 1,and I should actually, the speed of factor is1, but there is let me call it, l number of packets which are coming in, so these l packets when they are coming in, only one of them will be actually routed to this side.

(Refer Slide Time: 16:54)

And remaining l-1 has to be stored here, so every time when actually you're going to have, this kind of contention, so next time there may be new packets coming for the same port, you may be having l-1 backlog from the earlier transmission, earlier slot and more new packets. Now which packet you should choose for sending it to the output, this is the important question because, this leads to the performance improvement.

So you actually choose, very simple strategies, you can actually say that there is no strategy, I am going to pick up, so even if there is , out of the l contending packets, n packets from the earlier slots, which could not be transmit earlier, but they are slanging around. I will not give them a priority, I will pick up anybody randomly out of this thing, this could be one strategy, so l packets, randomly being transmitted.

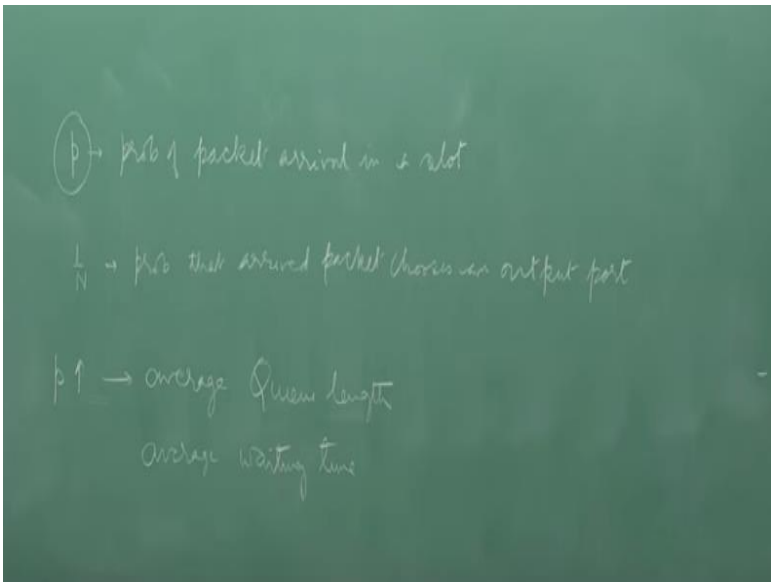
You can actually take packets from the longest queue, so that packet which haven't backlogged for the longest period of time is actually giving them a priority in this case, technically you can also have an strategy, you can define the priority actually of the various inputs, packet from the highest priority queue, so you can actually define the priority among the input ports. So in fact remember all these switches, in real life we need to implement priority classes.

Because, based on quality service, I should give priority to certain packets and should give less priority to others, so now the question which we need to ask, now coming back to, first of all let me analyze the output queued system an d then will come back to input queued one, so these are basically the issues which are there, for an output queued system, the question which we are going to ask is, so we have to define called p, which is the probability of packet arrival.

WE don't use arrival rate like landau, and in a Poisson's statistics, now everything is slotted, in a slot either a packet will come or will not come, these are only two things which will happen, so packet will arrive in a slot will be p, so that is going to be our parameter, with which we are going to look into the performance of the switch, so and we also going to assume, when a packet comes at the input, it can be equal probability can be directed to any one of the outgoing ports.

So probability it's going to directed to port will 1/n, so it's basically uniform traffic condition, that's what we are assuming, now what we want is, as you're p will increase, p will change as a parameter, then what is going to happen to a average queue length at the output queue, so I am actually talking about the speed of factor of n, so what happens to the average queue length and what happens to the average waiting time.

(Refer Slide Time: 21:17)



Contentions or the packets may try to go to the same outgoing port at the same time, packets has to spend some time at the switch, so when I am especially worried about , how much time it takes to reach the destination, this particular factor and queuing delay also need to be incorporated. Remember when I was talking about, circuit switching and packet switching in the previous video, I have drawn the timing diagram, I have only given the header processing time.

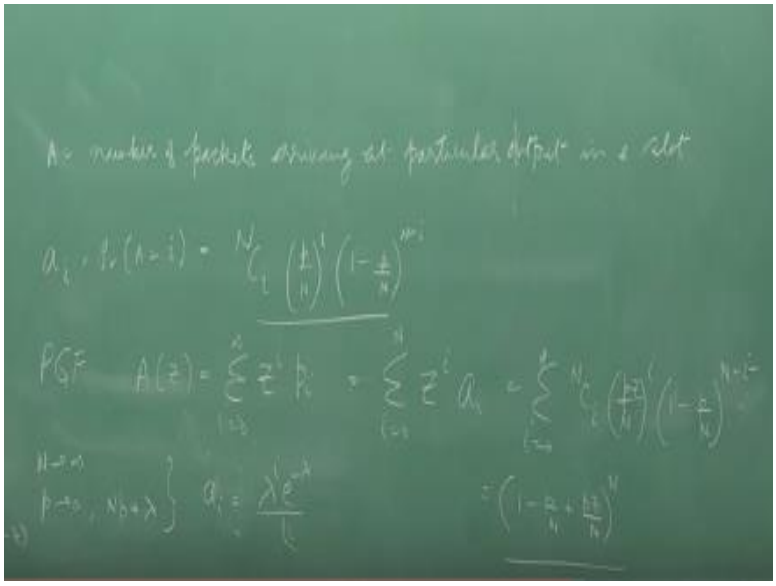
I have not actually included the waiting time in the queue, so waiting time in the queue should be, very small especially if it is a real time traffic. Otherwise, the roundtrip delay will be larger than required; I cannot transmit the voice or video, live video streams over it, over the network. Let's define, let's start analyzing it, so if these are n packets, each one of the time you will get packet with probability with 1/n probability, it can actually go to each one of the port.

So p/n, I can get a packet from here, with p/n, I can get a packet from here, with 1 – p/n, I won't get packet from this particular input ports, so what's the probability that in output, the thing I will get, no remember this switch is operating at k=n, There's an output queue switch, that is not input queue, so what's the probability that I will get I packets in a slot?

So I define this as a parameter, which we call it as the number of packets, arriving at a particular output, so actually tagged a output and analyzing that and whatever I am analyzing going to be true for all other outputs also, because I am not differentiating, we are all behaving in the same fashion. So this will call it A_i, is the probability that A is equal to I, so this will be given by, because there are n inputs, out of which I of the inputs will be having the packets will be directed to this particular switch.

Then only you can get I packets, and this can happen in $N_{Ci,}$ N commute to I way is, N factorial by I factorial by N-I factorial, and p/n is the probability that the I packets will be directed and 1-p/n will be the probability, that packets will not be directed, so that will be the $A_{i.}$ So once I had this probability distribution, so this is the discreet probability distribution, where I can take value from 0 to n. I can actually find out the probability generating function of this.

(Refer Slide Time: 25:09)



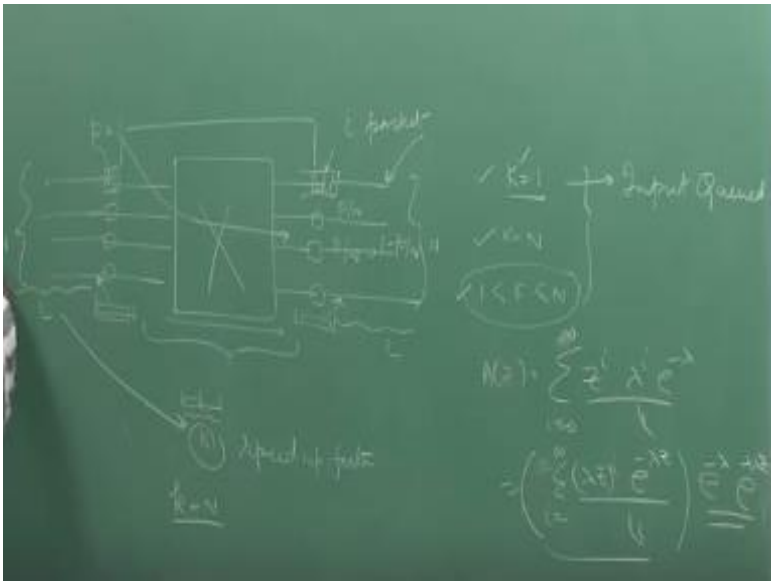
So this will be given by a z remember is nothing but the z transform that's way it will be return I go from 0 to infinity or whatever the valued values so in this case this will be $\sum$ of I goes from 0 to m z is for I and a I ok so we can actually put in the value a I and then solve it, so this z^i actually I can take inside with p /n this is a complete binomial so I can this as 1 –p/a n + z p/an

power n so that should be the value of the a z and that n will go become very large limit went n go's to infinity so this will actually can be approximate as an explanation so it will be like when n go's to infinity and e go's to 0 as such that and p go\s to some consents equal to n in that case this az will become e raise power.

So probability of a this one this will get converted to let me first of all ai will be converted to λ raise power I – λ e factorial this get converted to the point an once you solve it you will actually will get the similar so you will get p g f which will a of z will be ∑ of z I λ e is power of - λ of factorial so this n will become infinity and you can actually write λ z is power e

By I factorial e is power - λz so we will have this so we need to actually take this 4 λ out and e is power λz out so this thing will turn out to be one so which is an indentify and ultimately your solution will be this which will give you e is power - λ1 – z so that will be probity generating function for an infinity dimensional switch so for now such a case for this one this probity generating function a z is

(Refer Slide Time: 28: 56)



Is known to me now we have to also understand that this was congaing this many number of packets went the m slot is going to happen then the new arrivals with are going to come in m slot was actually am and there was one pocket which is gone out because it was transmitted out so which actually mean q of n leftover packets will be nothing but q m – 1 +a m – 1 and off course if this value cannot go negative so I have to write down it here that q of m is maximum of this or 0 okay so went for example this value is 1 and no arrival one go out it will become 0

When it is 0 and no arrival you cannot have a pocket going on out that is y there is a maximum operator which is to be happening so it will be either this or this which is maximum so this q will be defined by the that this m will have a probity generating function the distribution of this the p gf will be given by this expedition of infinity size which for the size switch this is what will be the expression so in fact we can even draw the mark of chain for this so you can refer to any book on Quinn theory and stranded text book for this one and I am actually using the results so.

So will have something like 0 state when there on pocket in the q so number of pocket which come if there only one packets which comes in that pocket will come and the output will be going out you will still remain 0 if there is no packets which arrivers then also this is going to happen so a 0 +a 1so that is the probity of arrival of zero for four pocket or one pocket ok so from 0 you can transit one I stay 2 packets arriving you can do a transition to two if there is there is packets coming in ok ultimately you actually have n packets coming in.

So you will end up in n- 1 okay so these are the multiplies remember mark of chains which we did for fare simply okay so this one similarly for 1 I want go to 2 a 2 coming in a3 I will come here and form here if I want to go also this will necessarily will recurs in this fashion this is what is going to happen similarly for this ok for this one will be happening when there is no packets coming in when there is no packets coming in and so on so this is what will be the mark of chain so this off course you can write it down the equations balance equations and solve it all the state

Probability okay once you have uvula the state the probability you can actually use this to compute what is going to remain is q length and there is a aula native way also correctly use an

expression can which is for this kind of q son if I actually take this to direct result for an p g f for q so q z is the p g f q z is the p g f of this kind queuing system can actually can be return z and an expression so this will be given by this and the beautiful thing about qz or any probity generating function is I can estimated any movement of the distribution using this so first

Movement is average value which we would like to actually note what was the question with what we start so if I know for example is that I want to find out meaning value so this was given as ∑ of z is equals to I where I go to 0 to infinity if I take this derivate then I will get I z is power – 1′ of I and if I take z is equal to z limit goes to one in that case this will turn out to be and this nothing but a – okay, the average value so I can similarly find out what is going to be the average value for the q if I find out q – this will turn out to be n – 1 / n this again actually using

A direct result this there in all text on quinoa theory so this will be given by p square / by 21 – p now interestingly this is an expression for average q length of a m d 1 infinity q what does it mean this expression the circular one for this only thing in which comes because of this batch arrivals so which the packets is coming in to the output qr coming in batch's depending on how many are trying to contain are so if the because of that you get this thing batch arrivals okay, then it will become md infinity q so when it actually becomes large this will cancel n- 1/ a n

Will all most can be approximate - 1 and q – q will expertly behave if it's a mark conterminal arrival deterministic departure single server and infinity buffer q what does it mean is that there is q and only one single server so there bias only one guy ok every pocket will take a exactly the same time is an deterministic thing okay, it is actually constant is there an constant departure post every packets takes exactly one slot time to go out and the packets which are arriving are happening because of the mark conterminal process.

So whatever is the average value so number of packets arriving in time will be given by λ power n e is power - λ/ t s set royal so this is the probity of n arrival is in time p now interestingly went I am actually talking about the switch at the inputs packets are coming in slots there not coming in un but the q will starts behaving in that fashion and becomes very large ok as if I actually having that kind of q so this is because of this mark conterminal thing determinates means

constant thing you know the value the single server and buffering the infinity capitals it will never gets over flow ok m that what it means it will axially behave as if it is a m d 1 infinity q. We would also like to know what will be the waiting time ok so n what will be the waiting time of an output queuing system so this waiting time of any pocket will cost of two parts so the first part is there will be a when the pocket will be arriving it will be arriving in batch of I packets so your packets which we are trying to take some packets trying to find out what's the delay it's is coming in a arrival in a batch of I size but before the batch arrival there already packets' which

Have present so theses packets anyway have to go out so the delay because of this I will call this belay has double 1 and there is a delay because of this tagged packets which you have can be placed here can placed here equally it can be placed any where there is no order there is no preparers in all the packets which have being pushed out the port okay because it is an output quinine switch so it depending where it is been placed I call it double 0 so total delay double 0 will be now given as w 1 + w 2 so what we need to find out is statics we will be using if I take

z transfer or generating functions for this delay times I can always write a w and z into w 2 set this again is a stand result from profanity theory and I am actually using that okay so using this we need to find out what is w1 what is w 2 and then make a estimate of what is going to be my delay in the system okay so we will stop now and we will in the next video we will be looking into how the delay will be estimated in the system.

Acknowledgement

Ministry of Human Resources & Development

Prof. Satyaki Roy

Co-ordinator, NPTEL, IIT Kanpur

NPTEL Team

Sanjay Pal

Ashish Singh

Badal Pradhan

Tapobrata Das

Ram Chandra

Dilip Tripathi

Manoj Shrivastava

Padam Shukla

Sanjay Mishra

Shubham Rawat

Shikha Gupta

K. K. Mishra

Aradhana Singh

Sweta

Ashutosh Gairola

Dilip Katiyar

Sharwan

Hari Ram

Bhadra Rao

Puneet Kumar Bajpai

Lalty Dutta

Ajay Kanaujia

Shivendra Kumar Tiwari

an IIT Kanpur Production