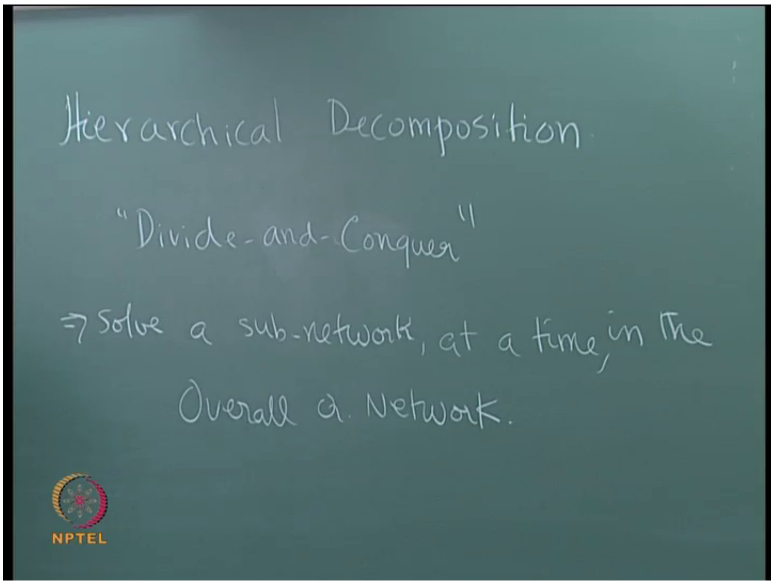
Performance Evaluation of Computer Systems

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

Module No.#01
Hierarchical Decomposition

This is the last topic in our last, but one topic. So, we have seen m m one systems, we have seen now also m m m systems right, load dependence service interest.

(Refer Slide Time: 00:24)



So, this topic is what is called as hierarchical decomposition. So, the basic need for this is when you have very large queuing networks queuing networks. Then it can be very time consuming to go through all this calculation especially when your n is large and the number of queues is also very large. So, you are trying to find ways to cut down on the computation time, to do it as sequence of steps than this massive queue that you are trying to solve or queue of networks that you are trying to solve. So, this based on the, so-called divide and conquer philosophy. And what you do is you try to find out the solution for a subset of the original queuing network, then use that solution replace that back in the original network and then do this in a hierarchical manner.

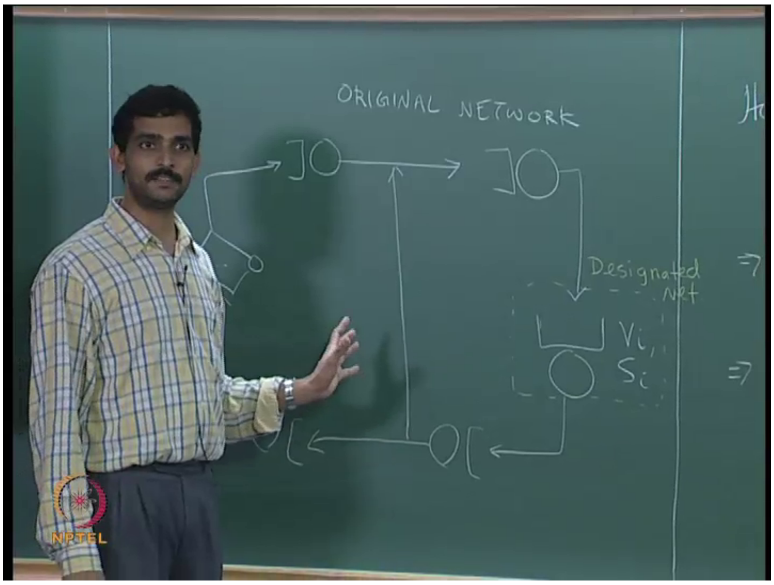
So, you start with smaller networks find the solution for that into grade that is rest of network and then the keep building this, so at a bit by bit to get the overall system solution. So, this is a divided conquer a divided and conquer; got it they where in got it. So, the idea is that you solve a sub network at a time in the overall queuing network. The idea is to handle smaller problem sets than try to have a very large problem set.

Especially, when you talking of one million if n is million, if n is ten million and so on then this might be able to, so all the reduce the complexity of the system itself. So, the basic idea goes back to electrical systems right, those from w should know what is, Norton's theorem? Norton's theorem.

Current is equals, because current going on current coming out sum is 0 (( )) is Norton's (( )).

Norton's, there is notional notional equivalence circuit, where you can short a circuit then find out the effective current flowing through that than we replace that in the larger network and so on. So, the exact same thing is what there going we will talk about we will talk about in this particular case.

(Refer Slide Time: 03:14)



So, let us look at so this is the original network, I have a queuing network that has these queues. Start with the queue here, another queue here.

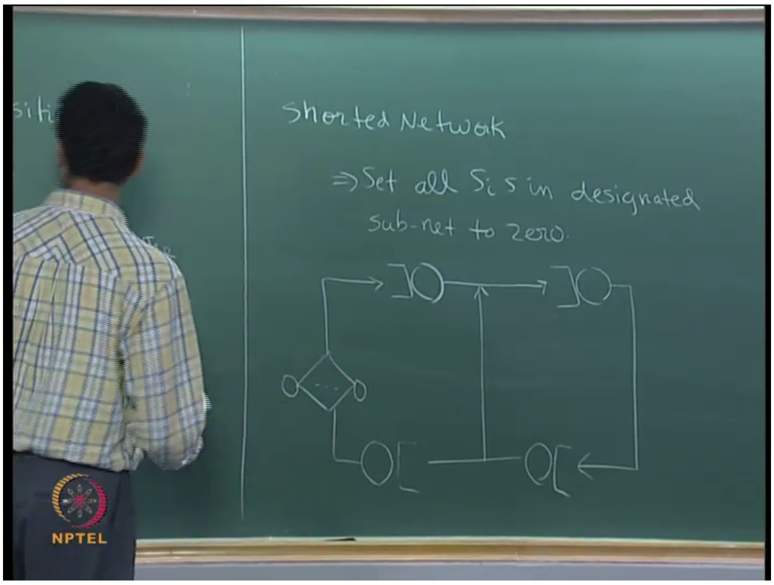
That is one queue that flows here here is in other queue, this flows here for another queue and then this goes to a delay sampleand then there is extra. So, that is let say one sample queue that I want to analyze I can of course, take this and run what we have seen, so for one of the each methods exact or approximate orthe queue in the convolution rather than taking place algorithm. And then I can get the answers that I want to but let say that I want to solve thiswith this a sub network mechanism.

So, I have two basically take this network and prove whateverrelational that I have break it of into two networks one we call us the designated network, and one is call the aggregate network. So, let us say this queueI can look at either one queue or a combination of queues, I can basically look at a complete sub network I could take this two queues and that from that my designated network, and then or could be all this are threealso the designated network and so on.

So, a break up the original network into the so-called designated sub network, and then the the rest of the network which is the aggregated. So, let us say that in this the example, this is my designated queue basically network. So, this is described by the two parameters V i and S i. So, the number visits to this queue i V i and S i is the service time average service time.

So, this is my designated network. Yes, you can. You take the subset also.Then what we do is in this designated network we set this S i to zero,so basically as like shorting this queue just like you short in electrical circuit. So, you short this basically remove this from the system then solve the rest of this queuing system with whatever technique that you have m v awhatever and then what we will do is this aggregate network will be replaced by a so called FEC, a flow equivalent circuit. And then I will have this V i S i back this flow equivalent system also put together that is my equivalent network as we will come.

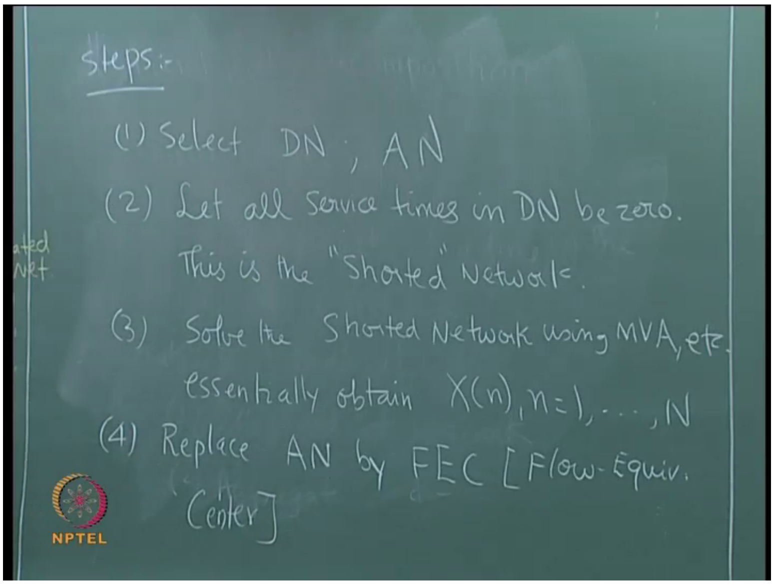
(Refer Slide Time: 07:04)



So,let usso that shorted network running out of space hereknow.So, for the shorted network set all theall the designatednetworks or queues service times to zero that so we short.

So, this shorted network will basically look like is is going to be essentially the redrawing the whole thing but it is gives that for the sake of (no audio from 07:41 to 08:13) actually the book example as one more queue here which I forgot drop does not matter when you go to that. So, there is a queue here than that is also get retained which I dropped my drawing here, but that is basically what happening your designated network is shorted and that is your, this youraggregates of network which you can solve with the help of whatever techniques that we have.

(Refer Slide Time: 08:39)



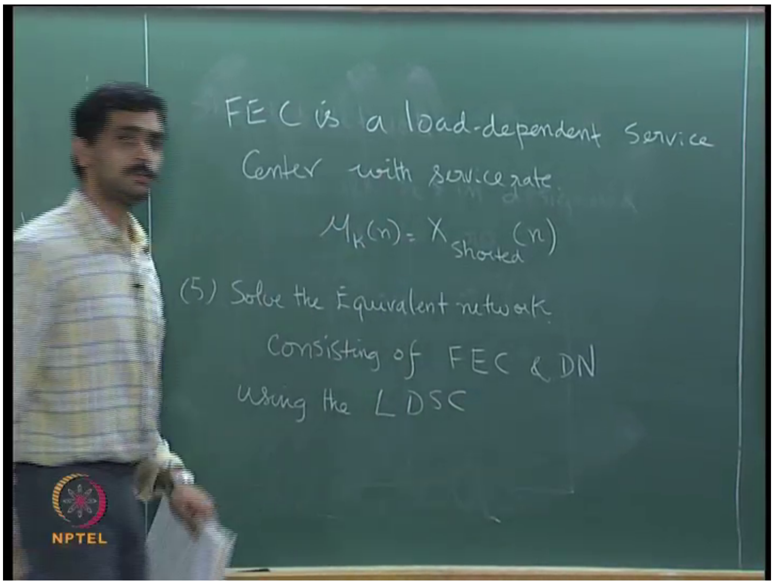
So, the basic steps we will come to what happens next the basic steps are select yourdesignated networkand then the aggregated network. So, first step which we saw alreadyand then let all the service timesin the D N be zero. This is the shorted network.

So, we have that know than solve that the shorted network for parameters that we want. So, this solve solve many find out the throughput and response times of things like that but mainly throughput is what we have interested in. So,I will get all the X n, so basically.

Essentially, so you know the N that you are trying to solve for the original network so try to find out the through put of this shorted network for various value of n 1234 and so on we may will get that. Because we do that any way in you require some. So, then we go back to this original system here, and then replace this aggregate network with the flow is so called FEC.

Now, it is not circuit is it.So this network gets replaced by a single queue, no single server which is represented by this FEC and for this FEC what we need to know we need to know two things V i and S i. So,V i is said to oneand S i is said to that iswhat you say for this FEC.

(Refer Slide Time: 11:38)

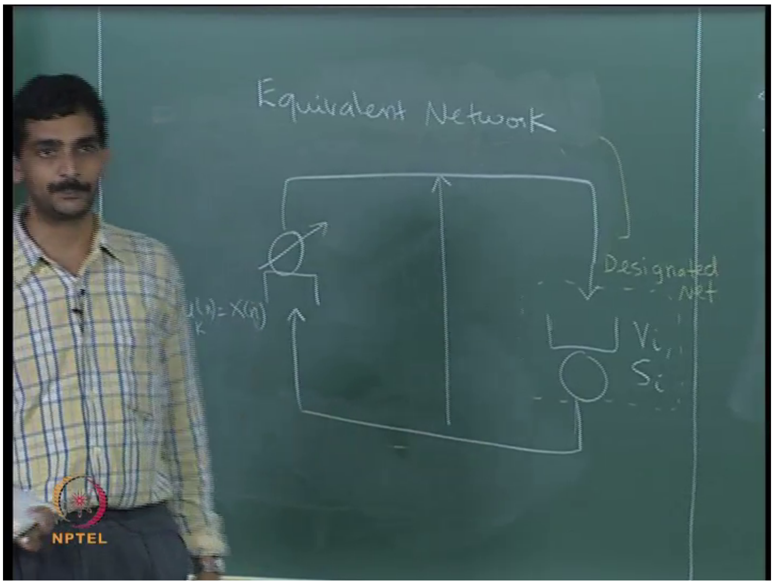


So, the FEC is basically a load dependent service center. And the service rates for this, so when there are n customers in this queue the service rate is simply the throughput that you see in the the shorted network.

So, what you found here this is x of this is the x shorted n so for n the there are n customers in the particular FEC the corresponding service side is simply the throughput. Well, this is we are we are just our taking at blind faith for now you want to really have go back to that chandice paper from75, it is beensome time proving that. But for now will just accept this as a fact, I am sorry aggregate network. This is the aggregate network.

Everything else is the aggregate network. So, everything butthe designated systems are queue is is the aggregate network. So, now I have two queues are the original designated network plus this FEC that forms my new network, so called equivalent network. I will show that picture for that. So, then solve the equivalent network consisting of this FEC and your D N using the the same thing LDSA. The load dependent service center modal that we just know saw. So, in this case may be its possible that this is completely mm one queues so for. I can use this mm one, but here I will have to use my l d's. So, essentially what we donow is go back to this original diagram.

(Refer Slide Time: 14:47)



Replace all this for us.With an efficiency, so packets come here go to thisFEC which is.

Remember, we draw the FEC like this, a tunable server. And here this of this of the service rate is simply X of j will be used that and let us we… And then doing back and then this self loop is needed. So, you could so that you can go back V i times to this first queue. And that is for you which it, for this is your equivalent network.

Without this loop you will what happen is a packet that queue was the job that comes in will have to visit this queue multiple times. This designated network queue. Without that, you will end up with having mistake errors in your calculation,if we do not have let us.

V i times by going through other queues.

That is only one queue.

In a original queuing network, the job will visit V i times one queue by.

After going through operations

Operation going to than going through then...

That is correct that is correct. So, here that is same thing is retained here. So, your queue definition now, this is again V i Si, this is what the original definition was the, that is not change.

This queue is represented byV i equals one and then this is service time. This is the service rate for that particular case that solve. So, you will know solve infact we solve this example before office of last class we basically this one remember there was an FEC that we solve with only one queue.

Equivalent queuing network thanit will be.

We see that this is V i.

For both of both of them V i.

I do not.

That is coming there V i times, that means that it is going through that two V i times.

Sorry is I know that is, I do not think so.I will come back to their but it should be one but I will will try to clarify that .This was part of the aggregated network also itmeant that there was the self loop. From so this self loop to the book says that many books do not have this self loop.

And therefore, because of that thereis errors in the calculation, but therefore the book is insisting on this one here, but I am not check the other books what they are rush. But this was also there in the original one. Yes, but you do not. If this is not there, then what will happen is yourpacket will simply so what we wantto see here is a job that comes here could either go to this queue or go back to itself and then repeat that is all details.

And to we need some other queue somewhere.

That is correct. That is correct, so…

How can we have a self loop in the if it would have a network.

Sir, may be the book did some other partition I do not know. May be, did it take the same designated net.

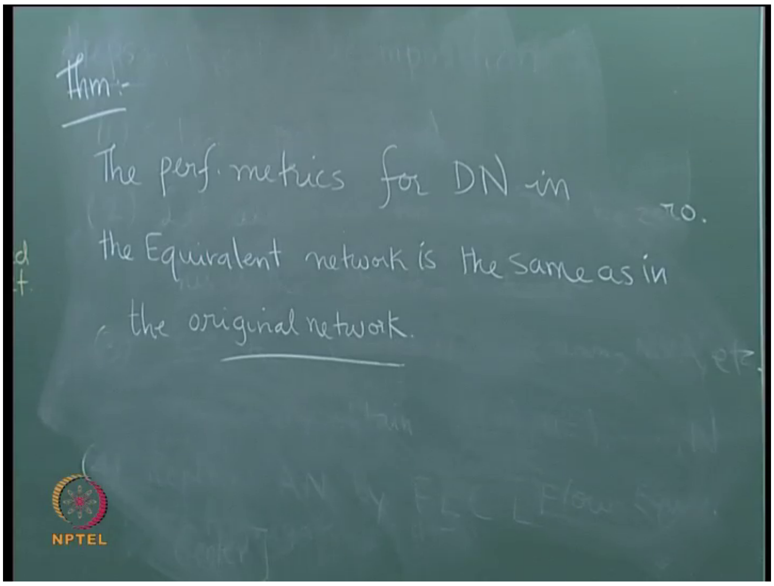
The book as a same designate, infact the book has an etcetera.

Infact, in the book example, if you look at the queue directionitself is different. You what we had is here there was an etcetera queue that actually went this way not even this direction.

The books the the queue direction was slightly was different. But still there saying that this is way thatyou only have two queues, one is thisdesignated network sub network and your FEC that is connection to the aggregate network. Some of this is, it is a just giving of the results as such it is not inequitably showing through.

Give a time, I will go back and look at that. See, if I can explain why this some of this. So, with this now so what is the next what is the big deal so.

(Refer Slide Time: 19:24)



So, this your theorem from so that the performance metrics for the designated network in the equivalent networkis the same as in theso by solving for DN in the equivalent network we get the same results as we go to got by solvingfor DN in the original network, that is the theorem that that is we proved. And based on that,I can simply analyze this equivalent networkto get there is metrics for the DN components.

In the aggregate network.

In the aggregate, if you want the performance for that. So, either, we can still get from the originalwhat you get in the aggregate network will not be the original network. So, than you let that take make that the designated network and then solve for that.

If you have a particular network that you will looking at that you want. So, either if solve the large system, I just solve this equivalent network to you get the result which you want. Then you have to make there to designate network and then do will for the rest of.

Take the submission so let some designated network will some policyare something.

No, no that is up to you.

That I can choose any thing…

You basically, choose what use of this what is of interest to this particular application. As a performance analysis to will say, I want to look at this particular or if you just want to see way can a break this in an efficient way. So, in this particular example as look at that way because I had a 6 or 7 queue network, I am just picking one queue as a designated network but if you want you can also combine the sub network as such take three queues as retain in a say in this part of the network.

Let say you have a this queuing sub system in this CPU queuing sub system and some IO some other IO system you want to focus on the discIO system. And you then you make there as your sub network and then say what is the performance of this alone,without having try to solve the entire system I want to use,that is the advantage of this.

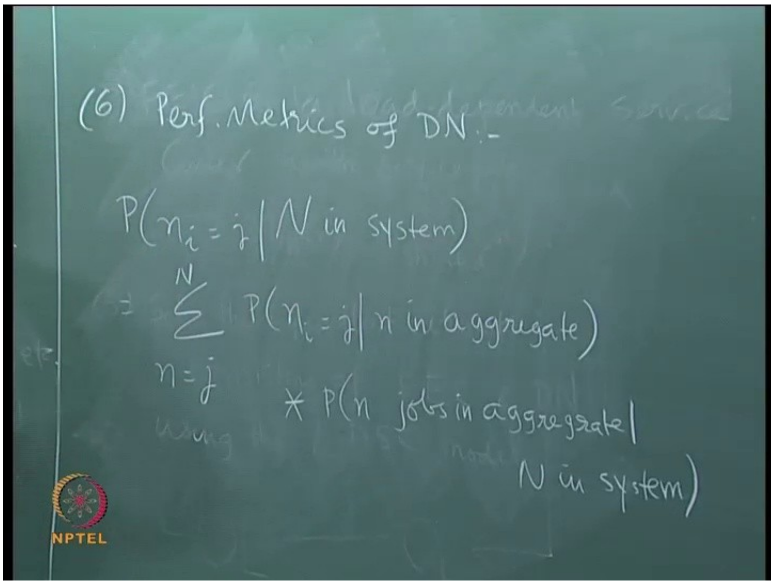
Particular end of service distribution on that…

Those where…

As required.

Those where whether it is to for other things also have to verify, but this one is if I exponential that certainly true so that. So, with these than the, so we had five here.

(Refer Slide Time: 22:11)

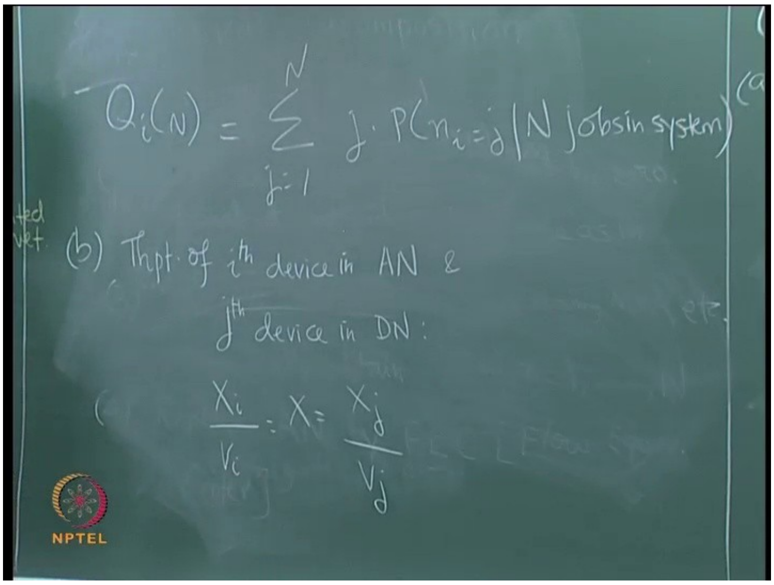


Now, that the theorem tells us we can press this.

So, we can now look for the performance metrics of the,ofdesignated network as follows. So, one is probability that so the probability that there are j elements in i th device as queue and remember that this could be other queues also in designated network. So, for each of these queues you will get the answering because you are solving for n v a, for those queues. So, p of n i equals j with N in the system equals (no audio from 23:00 to 23:43)

Again, we have this taking this at phase value and where to go back and check whether the error at has any other errors of this one that of the calculations wants things to have any this inconsistency see with the formulates of fight. So, the probability of havingj jobin this queues given by this expression where there are n i equals j, where there are n jobs in the aggregate times, n jobs in the aggregate with there are N in the system. So, this is you will get from the aggregate analysis from the shorted network analysis, you will get this spot here. And this one you would get with the equivalent network analysis. So,that is our first analysis than our first metric.

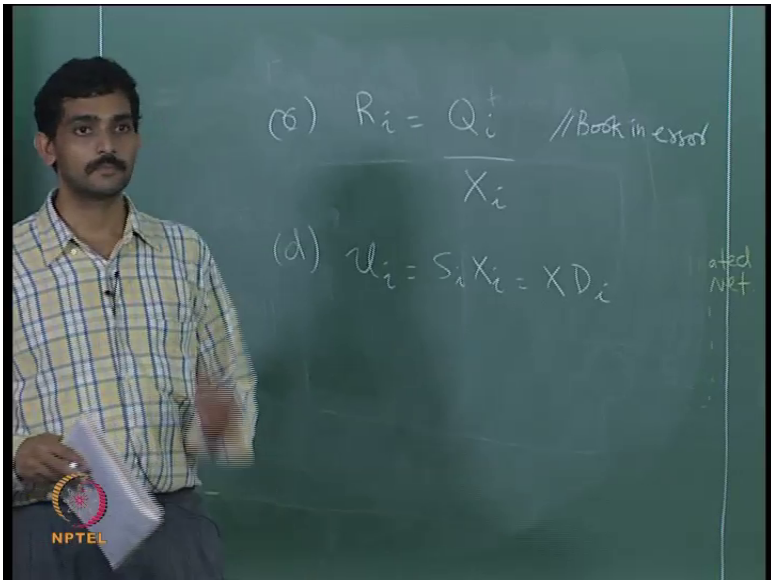
(Refer Slide Time: 24:27)



Then, I can simply compute than average number of jobs in Q i when there are N jobs in the system as such. So, there are N jobs in the system what is the average number of jobs in the i th queue, that is simply given by this that we j into.

So, once I get that p of n i equals j can simply plug in this one to get by average queue occupancy when there are total of N jobs upper case n jobs in the system as such. So, that is your first analysis. Now, I can derive the queue length distribution, so I can get the queue length distributions as well as the average queue occupancyin the throughput of the the i eth device in the equivalent network. So, this we know from before that x equals x j by v j, we look at the example to see how this could be, how this is actually calculated.

(Refer Slide Time: 26:29)



And then of course,R i Q i by X i again the book as in error, say its Q i into X i here. It is Q i by X i. So, the book is in error. This is our little slot once I know Q, once I know X, I can compute R. Then rest of it is a check over then u of i utilization of this S i X i which is also X into D ithen from thisof course, we can compute the overall response time R and so on.

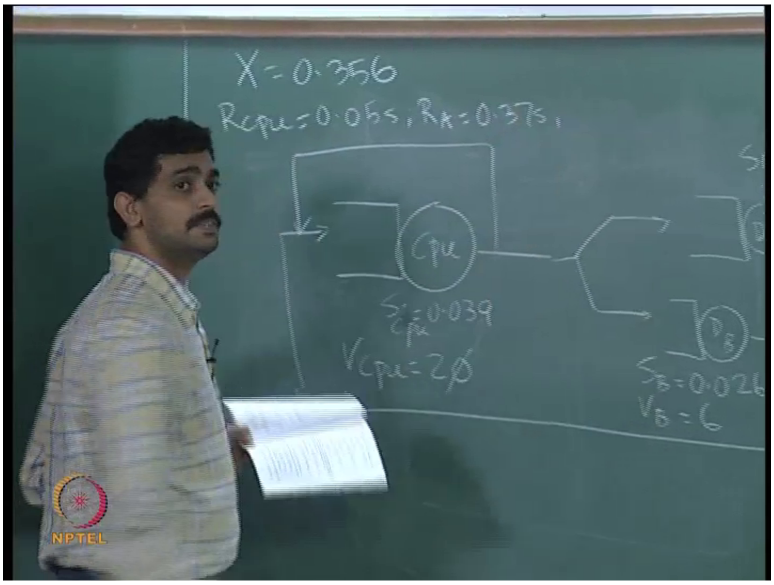
First, our device in a n or b n.

I refers to device in the d N.

In the d n.

I know here that this switch is there. Actually, this is the aggregate networkI is again in the aggregate network. Now, this is actually that this A N. The D N we get from ourequivalent network directly, it is A N that we. So, that is the process to follow.

(Refer Slide Time: 28:24)



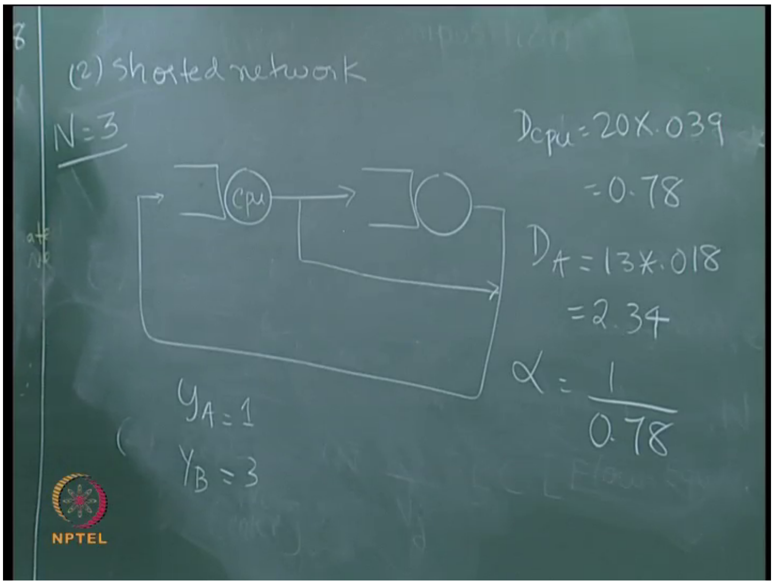
So,let us look at one example to see if this any clear..

Sometime back, you have solve this (no audio from 28:25 to 28:55). So, we have the same cpu this get is free and the S here was point. And just say S,that is the point is to same S cpu0.039, this we solved in the previous chapter with the convolution and service time is 0.2no, we did not solve this particular, and then s v is 0.026. And V A equals thirteen V B equals 6. So, that is the network that we want to solve, so this is we can solve using n v a and then we get the numbers for that.

Let us see where we solve this. No, we solve this is the previous chapterhere.This was our convolution algorithm that we solve so that was so what we got in terms of throughput see throughput numbers of. So,x for this system what you are computed earlier was x was 0.356, this is what we got before andlet us just write a couple of thing. So, R cpu was0.05 seconds R A was 0.37 and soon.

So, this is what we had solved with the previous convolution algorithm, now will try to see, if we get the same numbers with this other approach.

(Refer Slide Time: 31:26)



So, let disk B be the designated network. Now, the shorted network isgoing to like this. So, we have one cpu. So, this is this link is there because we set Si to 0 for disk B. And so the parameters of the same as before sowe want. So far the shorted network than D cpu is 20 into 0.039, so that is 0.78, the D A is 13 into 0.018, it was 2.34.

No.

If there something in which will be think k v should take,so disk A or disk Bforrandomly and using we should.

No, like I said we depends on which sub system you want to focus of on. So, when go into this hierarchical, you are breaking on the system into smaller andsmaller sub systems.

Look at one subsystem and then make that your physical network then look at the rest of the network as they aggregate network.

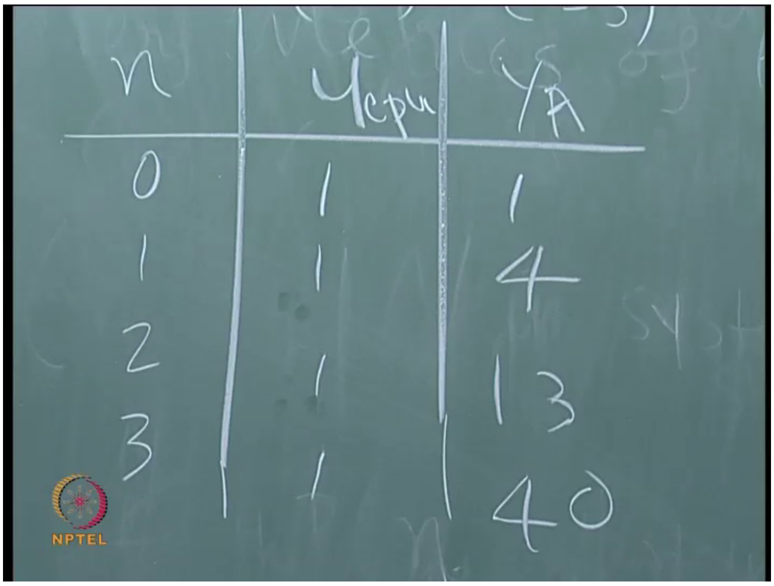
Looking at the complete system.

At the overall, interms of complexity whether one will give better solution there interms of solve a solution times. That I do not have I do not know how you go know to probably.

In this example, it does notmatter that which one that you choose you want be having much of saving computations savings time.In the case of larger network than that probably just comes by looking at that some sort of a, some knowledge as to say that. Because I can look at this system then solve this, because the smaller than and then solve the larger system based on that. But I have that whether there is a clear cut rulefor choosing one over the other.

So, we let say we choose all for will this before. And thereforemy y A equals one and then y B equals three. And we are assuming that N equals, so this is all for N equals three. So, we can we leaves the convolution technique to solve this.

(Refer Slide Time: 34:51)



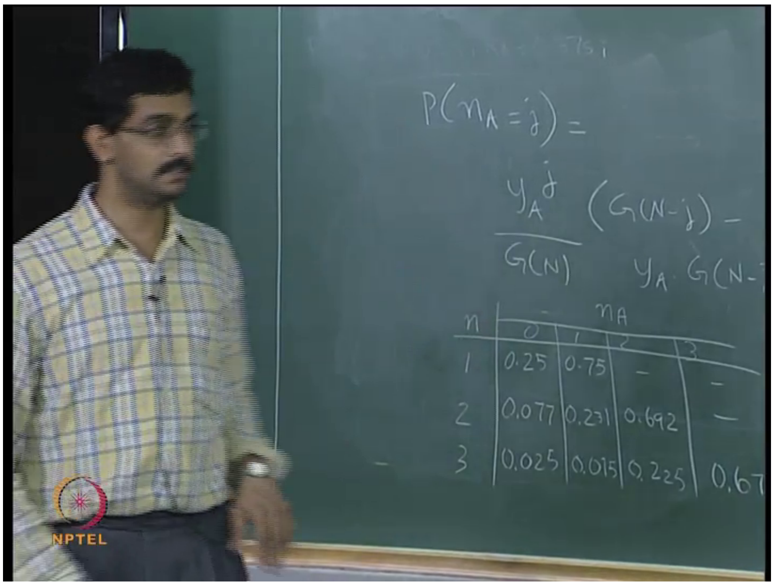
So, our G of N table which we can work out. So, remember thatthe first row in simply one and the first column is y i to the power n a.

In this case, y i equal to n therefore, all of this becomes simply one and Y A equals one and Y B equals three. So, how will this table now grow, so this Y n minus 1 and the the previous column are Y j minus 1. So, 1 plus 3 into 1; therefore, that is 4. One plus 4 into 3, that is13 and then this is 40, this is your G of N values, the immediate column to the left then the column to the go to the top, so we this away from yours.

So, with this table, I can now compute see what I want to find out is basically the X of 1 throughput when there is one customer, two customer, three customer and so on. Because that becomes, my service rate for the FEC. So, X of 1 equals, so what is X of 1 what is x of n. Alpha into g n minus 1 divided by g of n.So, this is in the shorted network, this is the throughputs for n equals 1,2 and 3, so x of 2 is alpha by g 1 to g 2 and so on.

So, we have thatthat is of n a v yeah, by using y equal to (no audio from 37:35 to 38:05) guess, we can take this of.

(Refer Slide Time: 38:22)



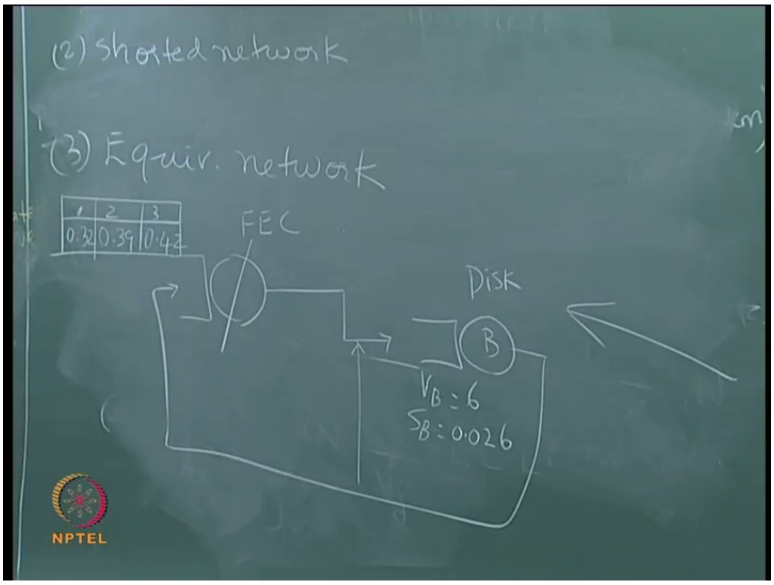
And then the probability that n A equals j A remember is y A j divided by G of N Gof.If this, thefind out the probability that our j customers in thethis network, this in the shorted network. (no audio from 39:09 to 39:39)

So, we will compute this table for now and just leave it that side, we use when you go to the next step, So, n A can take when n equals since the n equals this is your value of n to the largest value of n thatyou are looking at. So, here this n will get replaced by the corresponding small n for each of these table. So, in there is only one customer in the system than n A can we that zero or one. The probability is which we will not look at computing for now will just note this term, these not exist.

Question, so far…

So, I computed the x is and this is again available from the convolution algorithm, so will just list out this table, for now which we leaves later on. The distribution of the number of jobs in the i th in the a th queue a as queue. Now, I have may x so my equivalent circuit is know.

(Refer Slide Time: 41:48)



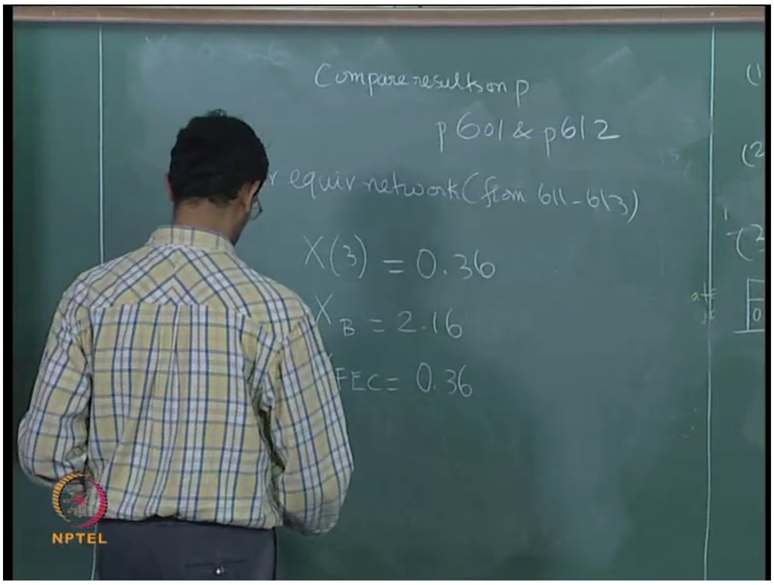
So, we can draw the equivalent network. So, we havethis aggregate network replaced by the FEC and this now feeds simply into the our disk B which feeds back here which interm feeds back here. This our my network now looks like. The disk B is backed at the picture and then the rest of the network is become your FEC and for FEC we know that the so FEC we specify by this table.

So, when there is one customer in this queue the service rate is given by the x one that we calculate there, which is 0.32 and then there are 2 customers, the service rate is given by 0.3 and then this is I am just rounding at all. So, this is directly from this table here, that is where that is nowI have inanother network we simply two queues, one FEC, one discrete.

And the discrete parameters are like before, so this is your V B equals 6, S B equalsso that is the equivalent network that we are going to solve next (no audio from 43:36 to 44:09). So, does this look familiar, we did this we solve this in the last class when it is the load dependent service center, we actually solve this system this set of heared service rates. So, what are the results that we got last time.

What did we solve for x of the throughput was 0.36, which is 0.356, what we calculated originally.And R A, for this system actually we do not havethat is look at R B R B is, R B is 0.43. And it original calculation R B whose 0.43.

(Refer Slide Time: 45:17)

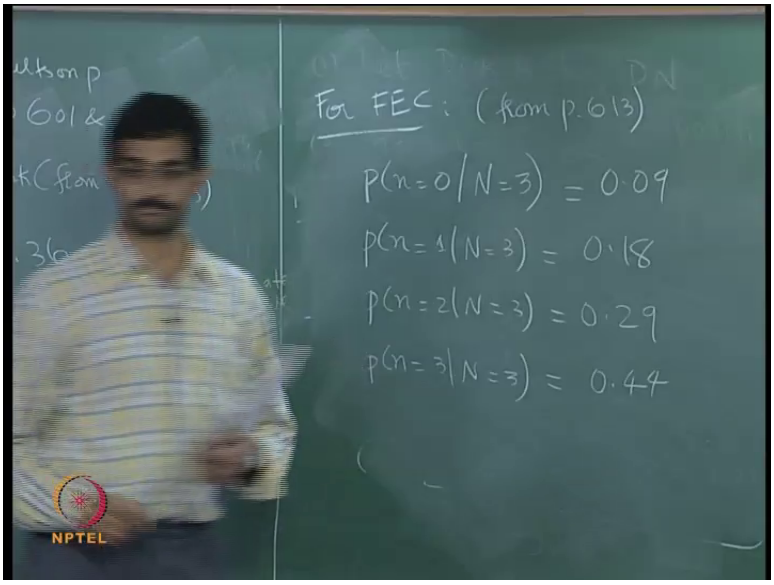


So, you will have to compare results on, so compare the results on.I guess page 601which is solving the with convolution algorithm.

And than the one that is solve with this FEC technique. The FEC one is part of the. This is an page 611 to 35, 612 and so on. So, what we had solved for thisequivalent networks for the equivalent network. And this is from 611 to 613. So, we are computed our front x of 3. So, since our n equals three basically this is what we are looking for X of 3 was 0.36 and X of B and X of FEC equals also 0.36, FEC was taken as we set count equals 1.

So, this is what we already had computed last class, then we also computed several other thing. We are computed at the probability of zero customers being in this particular queue, and there are two customers in the system and so on, that we are computed there, so this too much to flip back and forth page.

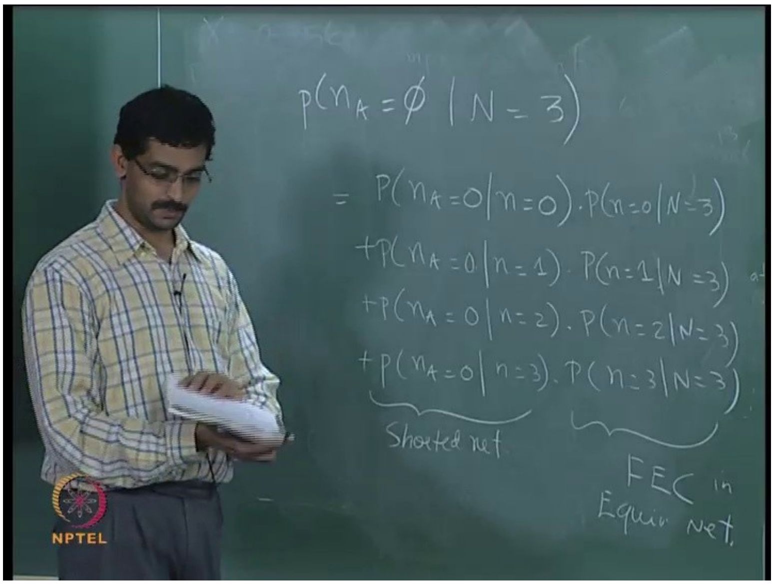
(Refer Slide Time: 47:52)



So, for the FECfor the shorted network your computed some probabilities of zero customers being in FEC and so on. So, the probability of zero customers in the FEC,when there are three customers in the system was given as 0.09, probability of one customer in this FEC is 0.18. This we computed from the FEC analysis last time when we did those p 0 1, 0 2 and so on. We did this for p zero three, if we look at the last.

We actually skip this calculation but this is the, this is from page 630 as such. So, what nextso, with this now for the designated network we have the results which is the same as d equivalent network which is original as well as equivalent or the same networks. Now, we if we want get back to aggregate network. Then let us look at some of the expression.

(Refer Slide Time: 50:06)



So, now to look atthe parameters for the the aggregate network queue, so let say we want to know the probability that in this queue A, there are no customer are no jobs when the number of jobs in the system is three.

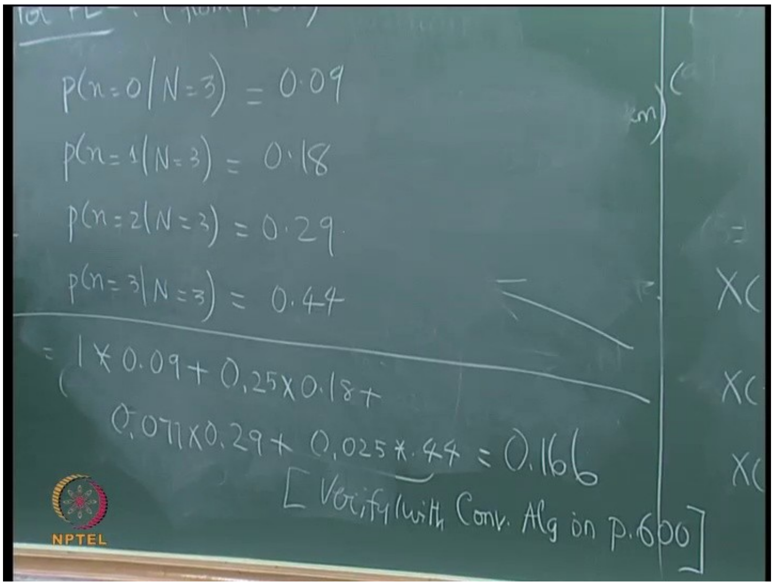
So, what is this given by that is when we go go to the other of set of formula which helps us all for the aggregate network. So, this is given by probability of n A equal to 0, when n equal to 0, probability of n equal to 0 when there are the three customers (no audio from 50:48 to 51:34)

To much on oneone pagewhat that is the formula that we wrote for the aggregate network. Now, this so where do we get this from. This is from the FEC from FEC we know the distribution that there are zero jobs in the FEC, when there are three jobs in the system. What is the probability that there are three jobs in the FEC, when there are three jobs in the entire system. So, that we got from ourequivalent network. So, this we got from the equivalent network, so all this this right side.

This probability we got from the shorted network. We did that derived this from the shorted network remember, it will from this table. So, this comes from the equivalent network FEC, in the equivalent network. So, this is when it done by computed than by hand but to get an understanding about squaring on. So, this what we derived the table it for it will table for n equals 0123. So, what is the probability so how do we what how do we proceed next, so we have to we can compute this probability.

Because, what I am try to do, so I am try to do compute the, they are parameters or the metrics in theaggregate network queues. So, what is P of n i equals 0 given n equals 0 and there is no simply 1.

(Refer Slide Time: 53:12)

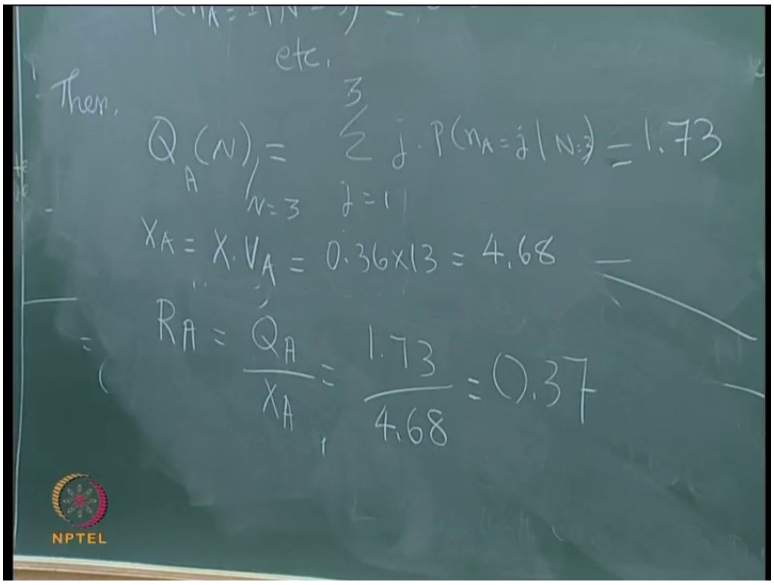


When there is no customer in the system therefore, this is 1 into this is 0.09 thanwhat is the probability that n a equals0 when n equals 1. We go to work table, from of the table that we true so the n i equals 0given that n i equals 1 is 0.25. So, this plus 0.25 into 0.18 plus, you go back to n equal to 0, so 0.077 and then 0.25 into 0.44. Now, I have the probability of zero customer in this k then there are three customers in the entire system, and this works out to 0.166.

And I think we have already somewhere computed this. So, wait how do you verify this, so we computed this, withthis approach. We can compute this some time before to and that is on page 600, then it will the convolution. If we look on page 600, which you might not have but if you go back to your notes for the convolution algorithm example that we did, if you look at P of n i equals zero for n i equals three, that is 0.166.

Do you have that somewhere in your notes? If you not refer page, so verify this. So, there this are two ways to. And like is you can do P of n i equals one, for n equals three then you can get your e of Q A.

(Refer Slide Time: 55:47)



So, likewise we can compute P of n A equals one,n equals threewhich this going to be 0.233 and so on etcetera, So, then I can computeQ I or QA. So, what is the average queue length for this k when there are three customers in the system, given that n equals three. This is for n equals three once I know the P ofn i equals 0, P of n i equals 1, n i equals 2 than I can simply take j into P of n A equals j you can do that. So, this is P of n A equals j given that our n customers in the system. So,that is so you can now systematically compute the other network queues also.

Once,I get my Q A is thenI can get the R. So I can compute R of A. So, R of A is now and again we can compare this. So, with this calculations, we get this to be 1.73. And thenR Aequals, so then we need to find out XA. So, X is X into V A. So, the Xwe already computed from the equivalent to be this 0.36.This is 4.68 (no audio from 58:24 to 58:57) And which we go back to page 600. I hope that this is correct R A is 0.37. So we can check with page 300.

Questions.

So, another exam, so another reason for using this decomposition is also when you have queue is where some of the network of queue is where some of the queue is do not satisfy our product form conditions. It we know we have some of the set conditions for product from it of the service time, so number of classes and types of classes and so on.

So, when you have a network which has partly hmm, product form conditions satisfy queue is an others that do not satisfy, then you make all the non product form solution a queue is an as your designated network. Because when you short all thoseout than the equal the shorted network, simply has all the product form satisfying queues.

So, it then you can solve that with m a or whateverone of this techniques than you have that FEC, and then you bring back your non-product form queues back to the picture than the resultant queue. You probably cannot either you have approximate solution for non product situations are our non product queue is or you can solve that with simulation.

Basically, you are taking away a large number of queues making that the FEC that is another FEC what is the computation have advantages that. You can take away are all the product form queue is out of the picture have a single queue, and then bring in all your non product from queue is than you can one simulation on that.

So, yournumber of queues in the simulation is much less intherefore, you can get the results faster. And then you get the equivalent network with the, which you can get equivalent network. Let us an approximate way of trying to solve when you have a combination of both product form and non product form queues in a given method, in a given network.

Questions.

So, will have this for your last tutorial, so that we try it out.

I mean, sometimes you just likes the routine number crunching which you can easily dowith a help of a computer we do not have to go through. This a routine of trying to compute all this, but on the other hand in some case which useful to know some of this techniques whether use this,in your particular system or not this. Not sure but may be or some pointing time. This is is a just tool get such you get to know that is exist, so that you can solve use them, if it all there is the need for that.

So, this the end of chapter 36, the last section of chapter 36 talks about some of the limitations of queuing theory, but this is you know written 20 years ago. I think it still wholes true but I am not really checked have. I asked Rajan why did not have a update on this later edition of this book and can we never applied to that. But you talks about all the difficulties of using queuing theory doesas such want is exponential times. There are several restriction that we all know.

And also when we look at systems where you have fork where you have process that fork and how do you solve process as or queue is, so if you have a cpu that is running a job and than the job forks, you get a new job that is created in the system how do you modal that.There are several disadvantages of the this models. I mean, there is this is an evolving field like any other field andmap.

AndI am sure there are advances but whether they fully capture in text books are not.I can not know, but for I thinkour purpose is we will basically say good bye to queuing networks, expect for one topic that said, I will cover at the n which is the balanced job forms that what you will finish that today. And then the next few classes will be more on how to conduct experiments applications comparing systems and so on which will beeasier on the brain compare to what your says on.