Performance evaluation of computer systems

Prof. Krishna Moorthy Sivalingam

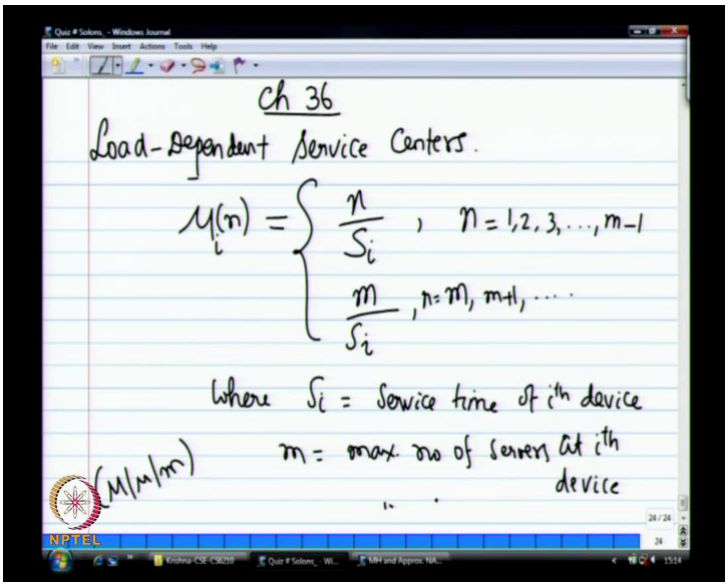
Department of Computer Science and Engineering

Indian Institute of Technology,Madras

Lecture No. # 30

Load-Dependent Service Centres

(Refer Slide Time: 00:10)



So far we have only seen m m 1 systems. So, how do we extend that analysis to systems were there are multiple servers in a given q, serving a given q.

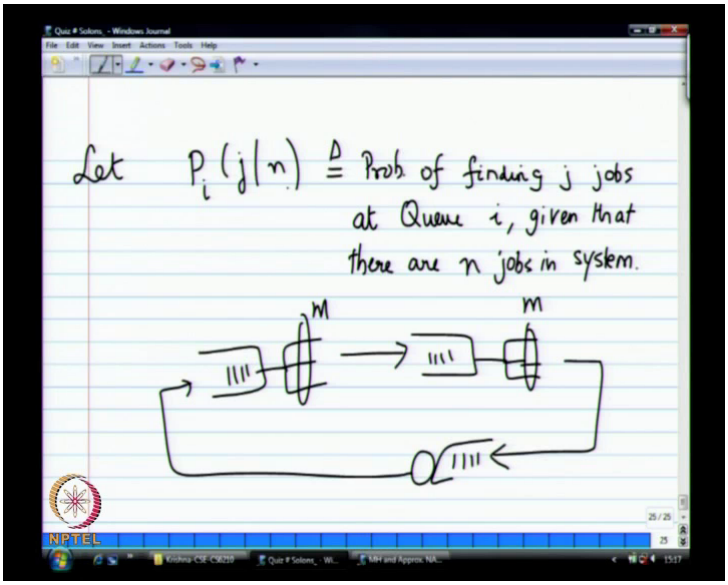
(No audio available: 00:27-00:44)

So, in a low dependent service center, previously we had the mu or Si was simply just one over mu. Now, the service rate is going to depend upon the number of customers in the queue. So, if there is only one customer, the service rate is simply mu. We saw the m/m/1 system or m/m/m system. If there are two customers, service rate is 2mu,3 mu and so on. Therefore, there effective capacity of the system changes with the number of customers in the queue. That is the only change which is not new to us. We know that.

So, the effective capacity of the system when there are n customers in the queue is defined as follows.So, if s is the service time, so it is simply one over s, right.The capacity is one over S

i.This case just ass. You could let say, this is(()). So, when the ithcapacities when there are n customers in its queue is simply n by S i for n upto m minus 1 m, where m is the number of servers, where S i is the service time of the ithdevice andm is the maximum number of servers at ith device. Basically, your m/m/m server.So, this isother wayof defining about the service of the queue varies with the number of customers waiting for service.

(Refer Slide Time: 03:05)



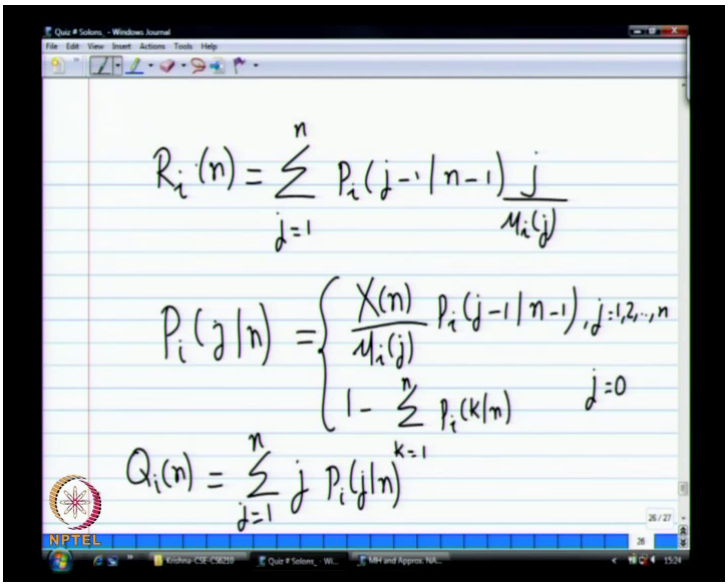
(No audio available: 2:48-3:10)

Now,couple of definitions Pi j of n.So, this is defined as the probability of finding j jobs at queue.So,if it m m 3 or 3 servicein the ith device, so every device is now so what we arelooking okI will just come back.Now, I am defining a queue of networks where there are multiple servers available.Previously, only one server was there and then, collectively this feeds into another queue which could also be a multi server device and then, this feeds into another queue and so on which might be a single server device.So, you might have a system like this.So, there are m such servers.This is a collection of m/m/m queues.

(No audio available: 4:16-4:37)

So, this is p i j of n is probability of finding j jobs at queue i given that there are n jobs in the system.Again this is for a closed queuing network.So, with that definition, then we will try todefine something.

(Refer Slide Time: 05:17)



So, Ri which is the response time at server i, when there are n jobs in the system, it is given as follows.

(No audio available: 5:30-6:09)

Does it make sense? So, if there are j jobsat a given point in a timethat is simply a probability of this j jobs,finding j minus 1 jobs when there aren minus 1 jobs in the system.This is p j minus 1.N minus 1 is the probability of having j minus 1 jobs in the ith queue when the n minus 1 job in the system.Then, your jth type job comes along.Therefore, the total service time is simply j divided by mu of j.

Makes sense? Some of you are saying yes, others are like what I(( )).Yeah, this is the average responstive of ith devicedepending upon the number of customers in the device and there is only one customer or let us say when there are only two customers, the response time is given in this manner.So, the numberof customers in that particular queue can be from one to endsummatingsumming out all those probabilities of 1 to n.In the system,yes.Out of which they could all be sitting in this queue or they could be sitting in some other queues.So, when there are j customers in the queue, the delay is given by p of i minus 1 or j minus 1 n minus 1into j divided byj is the number of customers because each customers required by mu of j service time.J service time.

So, that is not that hard if you spend other couple of days on itif you want to.Now, this magical probability j of n.So, how do we compute this?Thus, once we compute this, everything else falls into place.So, this is actually were not deriving this (( )).This is defined asx of n by ((No audio available: 8:07-8:14).

So, this is defined recursively.

(No audio available: 8:18-8:50)

So, this is well,I haveto go back and dig up some of the other queuing network textbooks.Where is the connection to x by mu?X is the number of x trooped of the system, right.So, this is the average number of customers completing the service at this queue.I do not have the intuitive explanation for that.

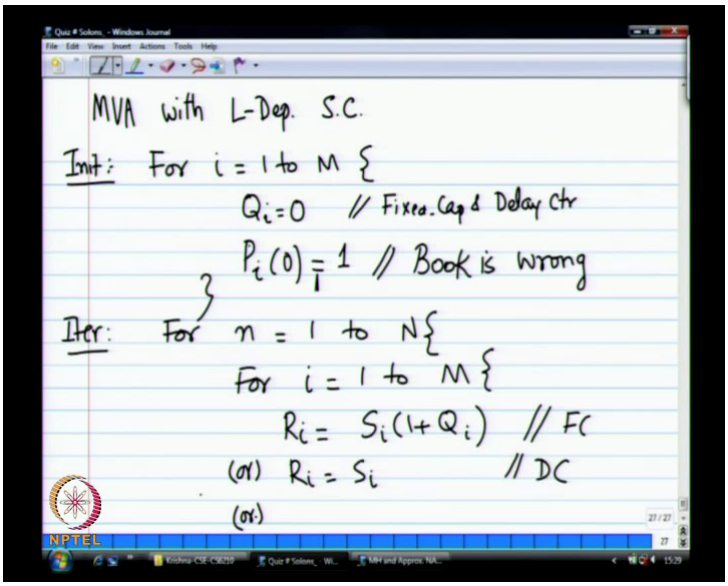
No, all it says is thatif there is only one customer, let say the queue is there are 800mbps links, only one pocket is in the queue,the capacity of the server is 100 mbps.When the two pockets are simultaneously being served, the capacity is 400 or 200 mbps and so on.So, when all four servers are active, the capacity of the system is 400 mbps.No,it simply says that at the ith job,ith queue rather if there are n jobs, then the capacity of that server is n by S i or m by S i. If n is greater than or equal to m, then it becomes m by S i.So, if I have 100 mbps links and there are four such links,right. If there is one pocket in the queue, then the capacity of the system is how much?100 mbps and that instant that is all it serving 100 m b p s.When there are two pockets simultaneously being served, it goes to 200 mbps.When there are five pockets in the queue, it is still 400 mbps because the others are getting queued.The effective capacity of the server is dependent upon the number of customers in the queue and so on.

So,you know it is8,so it is going to be 400 mbps and then, Q i of n.So, what will be Q i of n? You write them or can just tell me.What is the probability of i customers in this queue or when there are n customers in the system, what is the average number of customers at Q i? Sigma j.So, j goes from 1 to n and then, p f.That is all.So, p i j tells me that what is the probability of the j customers of the nresidingin Q i and that into j will give me the overall Q i.Now, we have expression for r and for q based on this definition of p.

Now, again we will have this (( )) vary of enumerate for the different value of j as see go along.(()).Theseare the expressions that are going to be changed with respective to the old MVA algorithm thatwe saw, exact MVA algorithmwhere we also compute. We compute at q,

we compute at then r based on the q. Remember what we did?We said r equals s into one plus Q i and then, from the r, we found R i we found out the r, from that we found the x and from that Q is n did when repeated.So, in this MVA version, the same computation isbeing done.The only difference is that your computation of R is being donedifferently and computation of Q is being done differently.That is all.Otherwise, the same model that we saw, but the exact MVA approach.So, the algorithmitself is kind of you know similarly repetitive.

(Refer Slide Time: 13:00)



(No audio available: 12:49-12:58)

So, this is your MVAwith so-called low dependent service content.Once I write the algorithm will so for (( )). So, this is the exactMVA.We started by saying that Q i equal to 0 andthen, we built up from Q i 1, Q i 2, Q i 3 and so on.So, this is for the fixed capacity and delay center.So, I have now a set of queues.Some are fixed capacity, some are delay center and some are this low dependent service center m/m/m.So, we love to run MVA for the entire systems and find out the trooped in other things.Now, we have to set this to one.

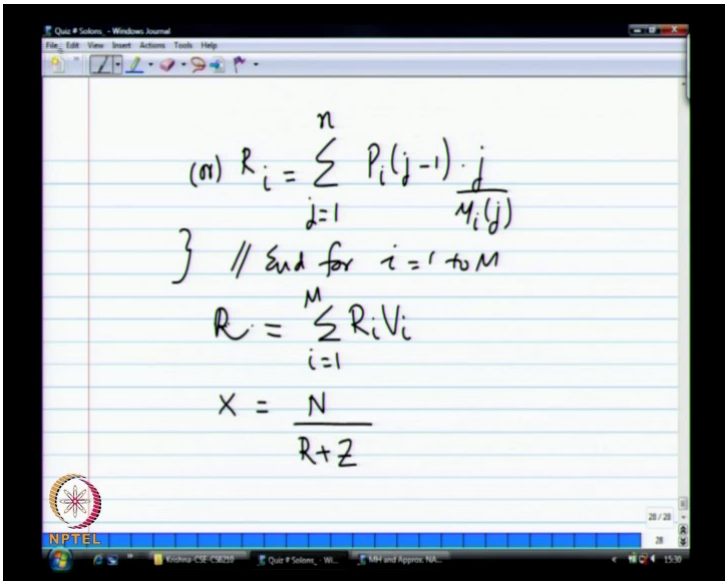
What is a probability oftheir being 0 customers?When there are 0 customers in the system, probability of 0 customersis Q i.When there are 0 customers in the system, it is simply one.That is the initialization.By the way the book is wrong.The book uses as 0 and it is incorrect.Start with0;everything will be 0 if you look at the rest of thecomputation.So,that isan error that there the errata, but not in the chapter itself.So, make sure that you note this

down.Do not make this mistake when you come to the exams, you start with 0. You never will get 0 for everything else.This is your initialization.Then, your iteration.

(No audio available: 14:51-15:03)

So far when it is known for given a value, input value n, we compute Q.The input values for n equals 1 based on the n values equals 2 and so on. That is for exact MVA. So, this is now for every queue.So, what is R i?If it is fixed capacity center, it was S i into 1 plus Q i.When there are n minus 1 customer, what is a queue length?Then, plus by self into the service time that so we calculated this.So, if it is for any i that is the fixed capacity center, this is R i.So, look at this switch statement or k statement and R i equals S i if it is a delay center.So, depending on what iis representing, you would change Buff formula.

(Refer Slide Time: 16:18)

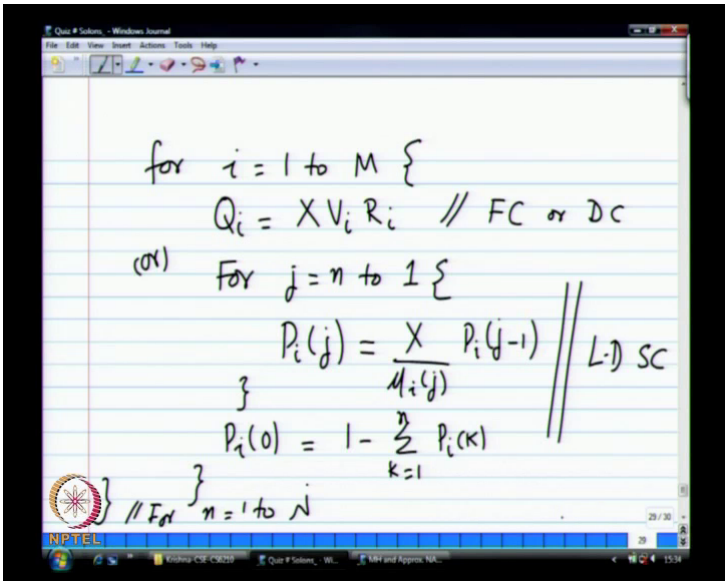


(No audio available: 16:10-16:46)

So, n is your current value of the index, the outer book index.This is what the R i equalswhen there are n customers side where they had the extra j minus 1 and then, if n minus 1 that if n minus 1 is what I removed.So, that is simply said I just need this 1, arrive p i to do that. I do not need the two-dimension.Arrive it simply if there are j minus 1 customer in the system,then what that is into j.That is why they start with n equals 1, I need p i of 0.P of 0 better be 1 that is 0.Then,I have everything else will be0.So,that is why that book (( )).So, this is it. So, this is my end for look.So, computed the R i, then I compute my x.So, this is what

we did last time.Now, the next step is when we compute, there we recompute the Q, then will recompute the p i's based on the previous values.

(Refer Slide Time: 18:14)



So,remember how we computed Q i is equal to x into x i into R i.That is basically x into v i into R i.So,we recently calculated R i, x is also recently calculated.So, we update the values of Q i.This is basically Q i of 1.This is if it is a fixed capacity server or also delay center.Does not matter.So, here we have little bit of (()),sofor j equals one to n.Sorry, there is one mistake in the book and which also has to be corrected.This is not because for n customers, this is again book will say book is in error.If you use that formula,that isyou will be getting the wrong values of x where n is the original input, n is 20.That should not be used here.The current running value of the n variable is what should be used.Any other place where n appears.This hopes for i equals1 to m.So, for a given value of n and computing all the R i's, computing the r,then the xand then,I am recomputing the Q i's to go to the next iteration of n.

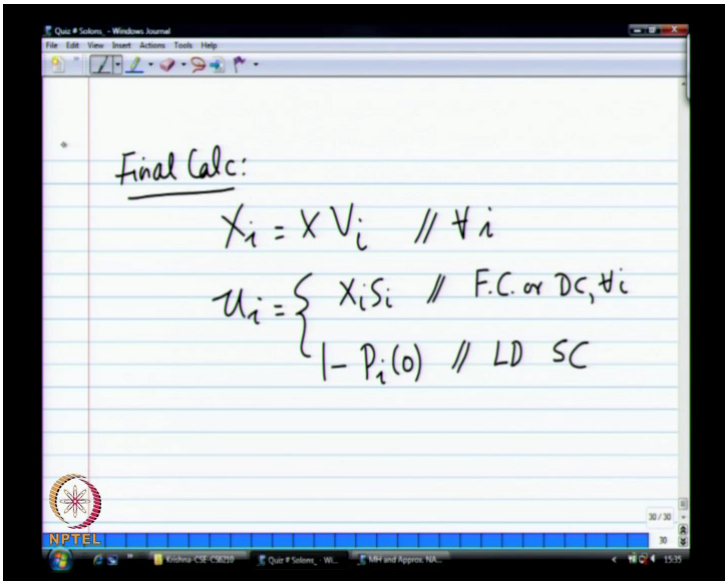
Smaller will go from 1 to n began is 20 will go from accurateMVA.We compute n equals onewhat are the q lines, n equals 2 what are the q lines and so on.So, this is for a given value of n within the, so for this we will again use that expression that we saw.This is the definition of p i of j, this is your x of nthat we just computed divided by m u i of j into p i of j minus 1.Then, p i of 0 is simply 1 minus (No audio available:20:41-20:51).

So, this is for low dependent service.

(No audio available: 20:54-21:19)

So, there is no other place where upper case can appeargood.So,that isone.This is one,this is one,this is as it is before.This is the only changed place on the p i that we saw beforeand then, that loop is closed.Yeah,this closes the, yeah this is for n equals 1to n.So, the end of n nitration you finally have the value (( )).So, what we have is the x of 20 or r f 20 and so on.

(Refer Slide Time: 22:01)



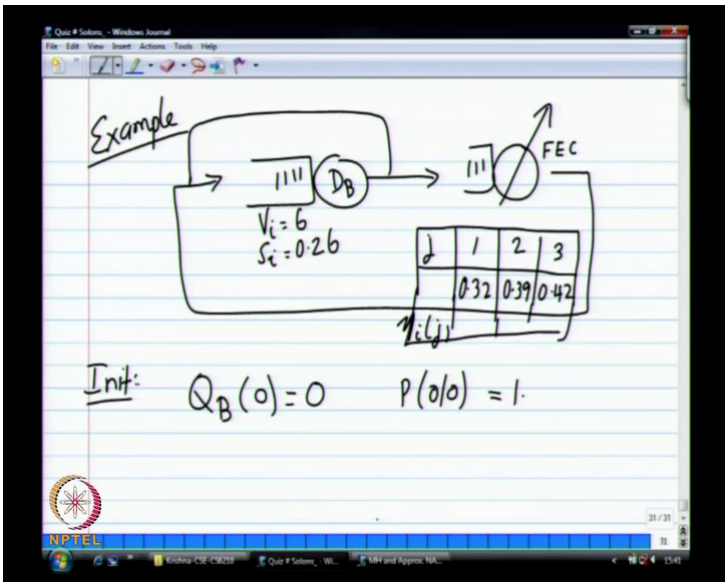
So, then we can do the final calculations.We will do x i equals x into V i.This is for all i and U i is simply XS.So, x i S i this is if it is a fixed capacity center or delay center for whatever values of i that match it and this is case of this one,it is simply (No audio available: 22:34-22:41).

So, p i of 0 is the probability that there are no customers in the Q i at the end of all the calculations and 1 minus that is simply going to be your utilizationparticular queue.Beginning of the equation are computing the R i's depending on whatever it is a service and fixed capacity or delay center or low dependent service center, right.That isthis loopier.So, once I compute the R I, Icompute my r and x like you have done before.Then, once you compute my x,I go back andrecompute the Q i's for all the devices.That is what this loop is all about.

Why it is a Q i compare, yeah that this in matrix that is just for the sake of let see where we will use that.It is only to compute the final,yeah Q i will be simply the number of customers that are in the thinking state, number of terminals that are in the thinking state.That is what a

Q i is.It is still Q i be simply U i.The Q i in that case will be equal to U iwhich is(()).Of course, we do not really have W ifor that the delay center. Yeah, then delay center is where that R i W i is coming.Then, we haveenough time to think for the numerical examples and will call it (( )).

(Refer Slide Time: 24:31)



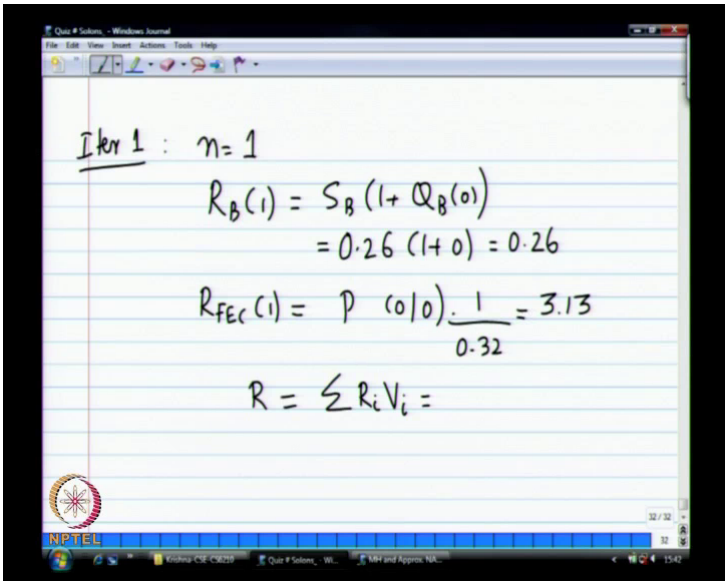
So, this is a shade of funny looking example.The reason for this will be to look at the next section.So, this is actually to serve the next section.So, we have this.This is your symbol forand we call this define this device FEC. So, jobs come from here to this disc and then, from the disc, it comes back here or it goes here.This is here.Only two queues.One is a fixed capacity queue.This is your fixed capacity center. This is your variable capacity center and definitions for this is here V i is equal to 6S i equals 0.26.We saw the same fill up before, this disc b 0.6 and 0.26 and 6.What is happening here is the disc A and the CPU have been so-calledcollapsed, this variable capacity center.How we will do that?Wewill see later. Now, the definition for this queue,this is follows.

So, this is j and this is mu i of j.So, when there is one customer, the service rate is 0.32.When it is 2, it is 0.39 and when it is 3, it is0.4.It is little bit different from what we saw so far where it is always a multiple of the number of customers.Here it is not a multiple, but this is true for any value of mu.You can define what ever combinations.When expectedvalueof number of customers increases, your capacity also increases.That is all.That is only real relationship.

In an m/m/1 system, m/m/m system as m, it is proportional to m and also note that everything defined for m greater than 4 which is also fine.So, this case cannot have more than four customers forwhatever reason.Utmost three customers can be (( )).So, this is our input.So, only two queues to worry about.This is a variable capacity server depending upon the number of customers in that queue.

So, the initialization parameters.So, remember QB, right.QB of 0 equals0.When there are no customers,there are no, in the system itself there are no customers here and p of 0,0 equals 1.So, for a low dependent system, we use the probabilities p j n.This is 0customers in this queue.When there are 0 customers in the system that is always equal to 1.This is our starting point.

(Refer Slide Time: 27:57)

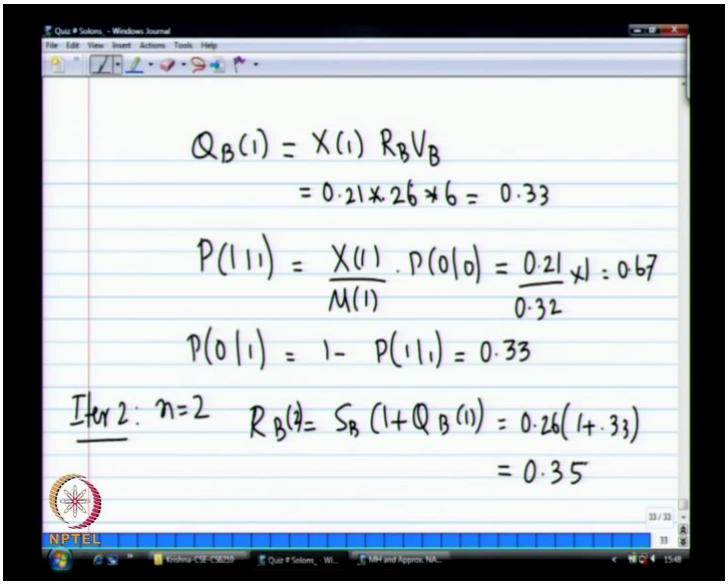


So, we now go to the iteration one.So, n will take the value one.So, now there is only one customer inthe system that is circulate.So, R B of 1 is simply SB into 1 plus Q B0.In the algorithm we did not have this parenthesis of 0 and so on, but this is what it is.So, this is 0.26 into 1 plus 0.Therefore, this is 0.26and RFEC of 1. Remember this is summation going from one to n.N this case happens to be one.So, this is simply PFECof p 0.This FEC also will drop because this is the only device which has this problem.So, it is simply P00 into j.So, j is 1 divided by 0.32.So, that is the first iteration.So, then you recomputeR.

(No audio available: 29:22-29:37)

Oh yeah and neither input is that your, I said V.This is VV.Here your VFEC equals 1.Only 1basically.

(Refer Slide Time: 30:49)

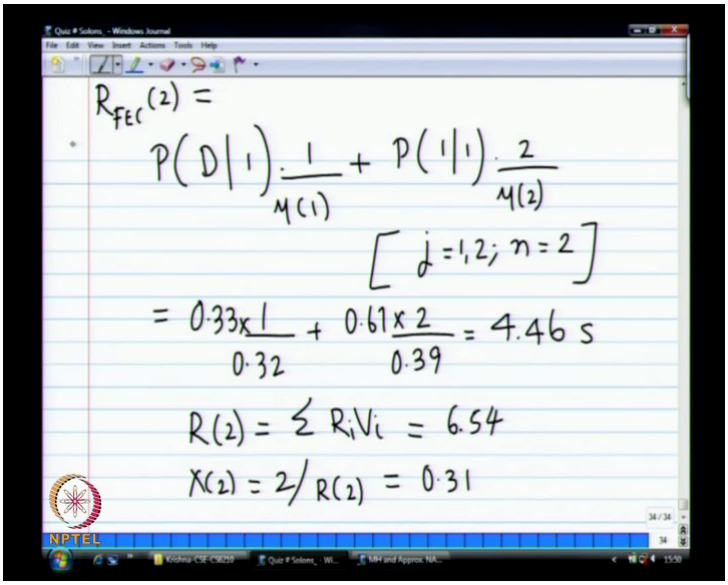


So,Q B of 1, the updated version for the Q length is now based on the new x that we calculated.So,X into R B into V B.So, that is now 0.33.So, from Q B of0, it is now updated0.33.Likewise p of 00 was 1.P of 11 is what based on p 00.So, that is now updated to 0.67 and if youlook at the expression, we compute everything from one to n and p b of 0 or p of that 0, that number of customers being 0 in the ith queue.It is simply 1 minus the probabilityof all the other problems, sum of all the other problems.So, this is the probability of one customer in that variable capacity center when there is only one customer in the system is 0.67.So, probability of no customer in this queue is when there is only one customer.The entire system is 0.3.This is time to compute.This is the first step of the iteration.

Then, we go to the second step.Now, my n gets the value 2.Then, we repeat this whole thing again.So,Q B equals S B into (Refer Slide Time: 30:56-30:59)

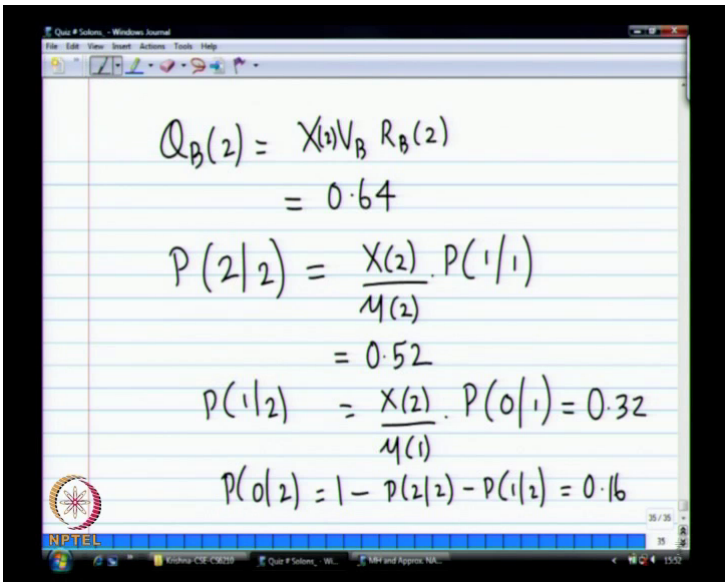
So, now Q B with two customers.This is with Q B with one customer.So, that is 0.26 into 1 plus 0.33.Oh sorry.Yeah, so this is R B,butthis is Q B.Yes, nowthereQR and SRare at properplace.Now, this is 0.35.So,this was the standard update forthe last time.So, only now the probabilities are going to be different.Now, the probability of (()).

(Refer Slide Time: 31:47)



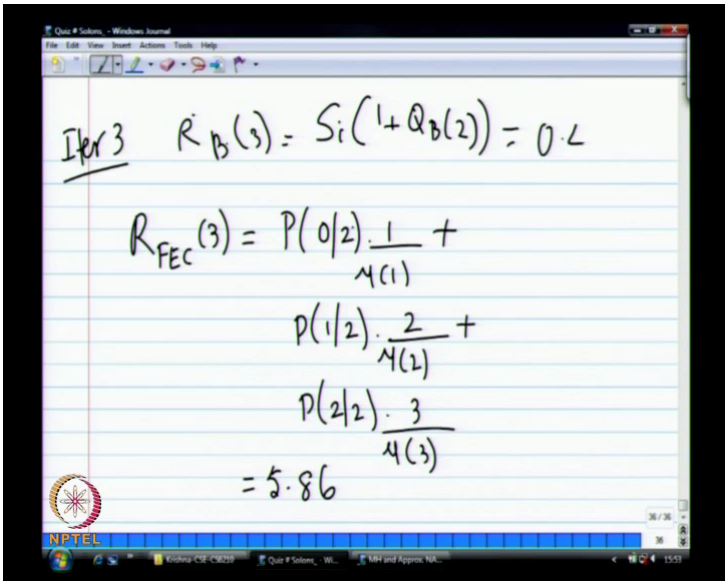
So, this is now RFEC with two customers equal to the probability of one customer or no customers in this queue with overall one customer into j by,so this is j bymu of j.Then, j takes the value.So, j is 1.This is j minus 1;this is n minus 1,j minus 1.So, in this case j is 2.Sorry, j goes from 1.J is either 1 or 2 and n equals 2.So, this is 0.33,thisis 0.32, this was0.67into 2 divided by this was0.39.So, now your mu r value is to be 4.46 seconds.So, it is kind of repetitive.That is the basic idea, but this helps youfind whatyou are looking for.If you have (()). Now,RF2 is now computed like before R i V i and that is nowupdated to 6.54, then x of 2 is now 2 divided by R of2 that is 0.31 and so on.Ok, that followed.

(Refer Slide Time: 34:09)



Then, we do all the updates.So,Q B of 2 is x into V B into R B of 2 or x f 2 and that is now x updated to 0.64.So, the queuelength increases to 0.64.Now, we have to compute thep 221 to and then,0 of 2.So, p 2 of 2 is given by x of 2 by mu of 2 p of,1. So, this is 0.52.Then, p of 1,2 again.So, p of j, nis given by x of n which is that by mu of 1.So, no more that.Then, it is j minus 1 n minus 1.So, this is p 0,1 and that is nothing, but 0.32.Now, what you find is the probability of 2.Both the customers being in this FEC when there are only two customers in the system, it is fairly high or0.52 and then, one customer is 0.32 and therefore, no customers is simply 1 minus p f 2,2 minus p of 1,2.So,that should be 0.16.So, this is end of our second iteration.

(Refer Slide Time: 36:09)



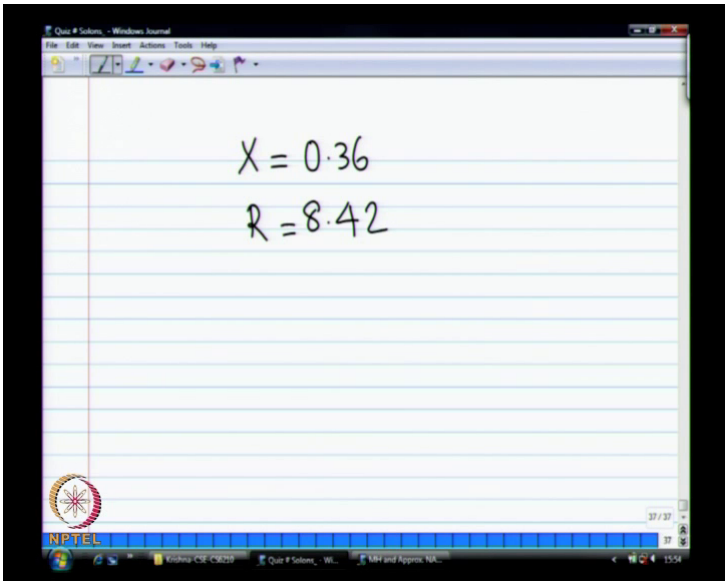
So, third iteration. I will let you figure it out.Now,RFEC of 3 equals 1.So, again j goes from 1 to 3.Therefore, it is j divided by mu ofj and then, p of j minus 1 n minus 1.So, if when j is 1, so this is 0,2 into1 by mu of 1.When j goes to 2, then this is 2 by mu of2.When j goes to 3, this is (Refer Slide Time: 36:47-36:51).

Yeah, this then finally computes to 5.86.Now,Q B we know that issimple.Sorry,R B, this is and so on.(Refer Slide Time: 37:08-37:26).

Once you get the r, you get the r.To get the R i is get the r and then your X and then, you repeat this.We have one more iteration, but that we do not have to worry about, but inthe end

youwill be able to get the trooped of the system and so on.So, the final result we will just writethat down.

(Refer Slide Time: 37:54)



Now, your X is finally going to be0.36 and your final R is going to be 8.42.So, we will stop here.Questions before we (( )).So, this is the (())thatyou canbring in there variable capacity servers also into the system.So, programmingassignment three requires due to implement a bunch of m/m/m server and solve that simulation and then, solve this very simple program and then, compare the two results.