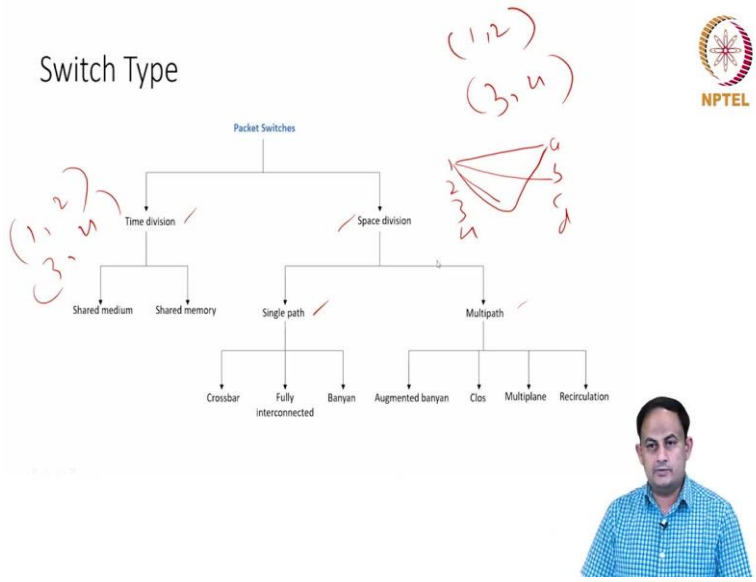
**Advanced Computer Networks**
**Professor Doctor Neminath Hubbali**
**Department of Computer Science and Engineering**
**Indian Institute of Technology, Indore**
**Lecture 22**
**Packet Switching Fabric Design – Part 2**

(Refer Slide Time: 0:20)




After understanding the some terminologies and the issues that can happen while you do the switching and the routing. So, now let us spend some time on understanding what are the different architectures or the design strategies available for designing the switches. So, the packet switches in all the fabric when I say the switch here they just the fabric are broadly classified are of 2 types, the first one is called the time division fabric and the second one is called something called a space division fabric.

In the time division fabric what happens is only 1 pair of the switches sorry port numbers are engaged in the conversation. So, when I say that, if let us say 1 and 2 or input port number 1 and 2 are engaged in the conversations, so 3 and 4 cannot talk some parallelly only one pair at a time. So, that kind of the strategy if your switch fabric is allowing only the communication of that sort then it is called other time division the type of the switch.

And the second category of the switches are called as the space division switch fabrics where multiple parallel connections are possible. So, in this case, the same example 1 and 2 can talk 3 and 4 can talk, these are parallel forwarding that the switch fabric is doing that there is a connection between 1 and 2, there is a connection 2, 3 and 4. And the each one of these, be it either the time division or the space division, they are again multiple strategies are followed.

So, along the time division, you have something called the shared medium and the shared memory, along the space division you got something called a single path and the multi path.

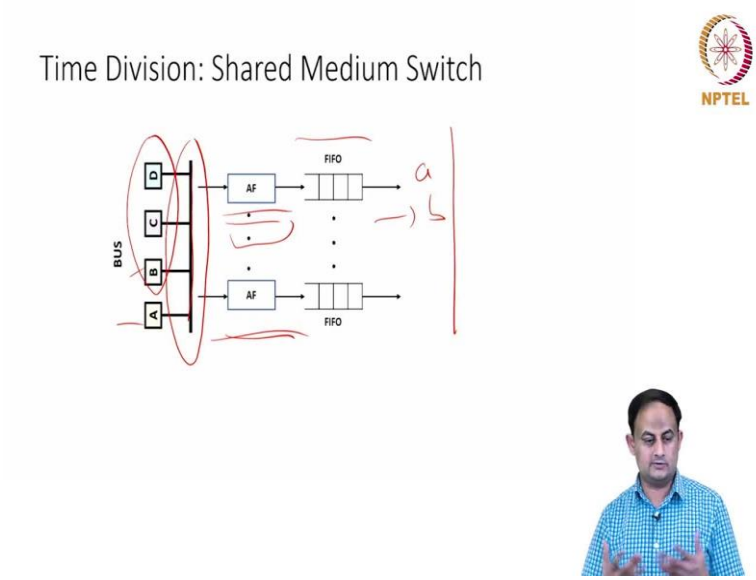
So, we will take a look in detail of what is the shared medium and shared memory, but let us understand what is the single path and the multi path. So, single path kind of the switch fabric has got a unique path, the single path from one input port to the one output port, what I mean by that is. So, let us say I have input port are numbered 1, 2, 3, 4, and output port number a, b, c, there are 4 of them and there is only one and unique path to go to enrich from 1 to a and there is exactly one path from how go to go from 1 to b and same thing applies for others as well.

From 4 input to switch port number 4, there is only one way to go to a, there only one way to go to b and c and d, so as many output ports are there those many unique paths are there. And every time you have to forward a packet, you will need to use only that path to do the establishing the connection between the input and the output port. But in the second category, where the multi path was there, there are multiple paths available between an input port to an output port.

So, meaning maybe the simplest way is in on the input from the input port number 1 to reach the output port number a, you have 2 different paths, one is this one and one through something else also you can reach. So, like this problem every input port to every output port there can be at least 2 number of distinct paths that exist in the in your switch fabric. If your switch fabric is of that type then it is called as the multi path switch fabric, among the space division.

Anyway, parallel connections are possible, when I say the parallel these are independent parallel connection. So, to the same port output port number 2 different people cannot anyway transmit so that is not possible, but the independent transmissions like 1 and 2, 3 and 4, 5 and 6 they are nothing to do with each other's conversation, so they can engage in the transmission parallelly. So, if switch allows that then that is called the space division switch fabric. So, let us go and see some of the examples of these which types and try to understand little more detail how exactly they look like.

(Refer Slide Time: 04:57)



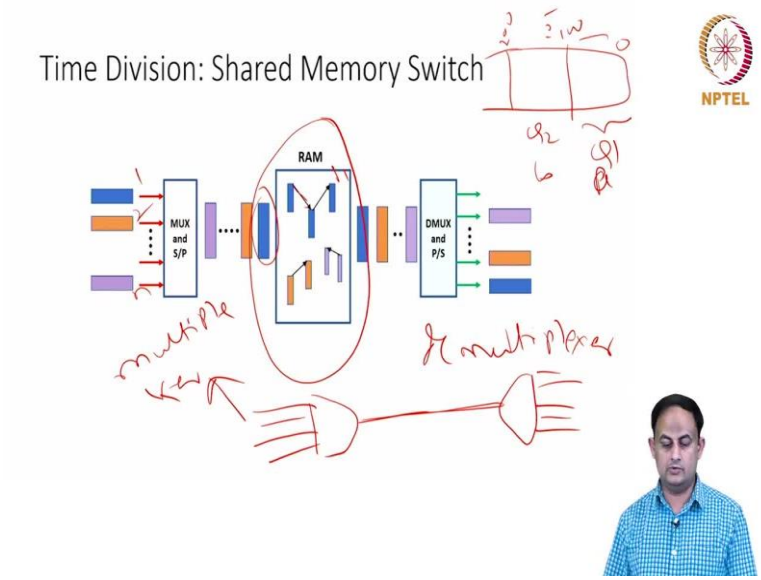

So, here is a picture, it is of the category called time division where what we said only one conversion from one input to the one output port it is transmitting the packet. And this is the time division is here implemented in something called as a shared medium. So, if there is a set medium in case there is a medium here in this case it is a bus, so that bus connect the all the input ports and all the output ports and as I said buses shared medium we can roughly compare this with the Ethernet technology where only one sender and one transmitter receiver is able to receive the only one sender is able to put an Ethernet frame onto the link.

Something similar is happening here. So, A is transmitting none of these D, B and C will not be able to transmit it. If B is transmitting no one else is able to transmit it and transmission here means that B is simply putting the packet onto this bus and there is something called as the address filter and this is again exactly similar to that of the Ethernet. So, when it broadcasts over the bus the interface is going to select the packets or the frames address for itself and then it will transmit it to the upper layer.

So, similar the address filter here is on the broadcast medium on this bus the packets are put. Now, if it is meant for the port number a and this address filter will take that packet and then put it into this queue there is one queue corresponding to every output port. And similarly, if the packet is meant for the second output port maybe B then there is a second address filter and something like. As many number of output ports are there those many queues are there and those many address filters are there and the address filter job here is to understand whether the packet is mean for the port number corresponding to itself, if it is meant then it will pick up the packet and then place that entry on the queue, so that is all is happening.

So, again, only one input port can transmit the packet at a time. So, now although they may not be going to the same output port, even then it is everyone else is blocked when one is transmitting that is the shared medium kind of the switch, it is a time division. So, here you need to take turn once sometimes A transmit, sometimes B transmit, sometimes C transmit, something like that, so that is what is the time division.

(Refer Slide Time: 07:57)




So, the other way to implement a time division kind of the switch fabric is something called as a shared memory switch fabric. So, the conceptual idea here is you got a memory in this case here is the RAM, which is the central portion of the switch fabric. And there are 2 other components one is called as the multiplexer and the second one is called the de-multiplexer. So, de-multiplexer is this on the other side and you can think of something like this. So, the many inputs are coming and one output is going this is the multiplexer and the de-multiplexer is the other way around. So, you receive one input on one line and then you need to segregate into multiple them.

So, the said memory here is the RAM that is where the packets belonging to different output ports are placed in. So, it just can be very well a logical division something like this, I have got a memory structure from let us say the starting address of this memory 0, from 0 to 100 bytes this is the queue where the queue corresponding to the port number 1 or the maybe that port number is A. So, the Q 1 is logical within this address space and from the address number 101 to somewhere around 200. This is the Q 2 corresponding to another port output port number B, something like this.

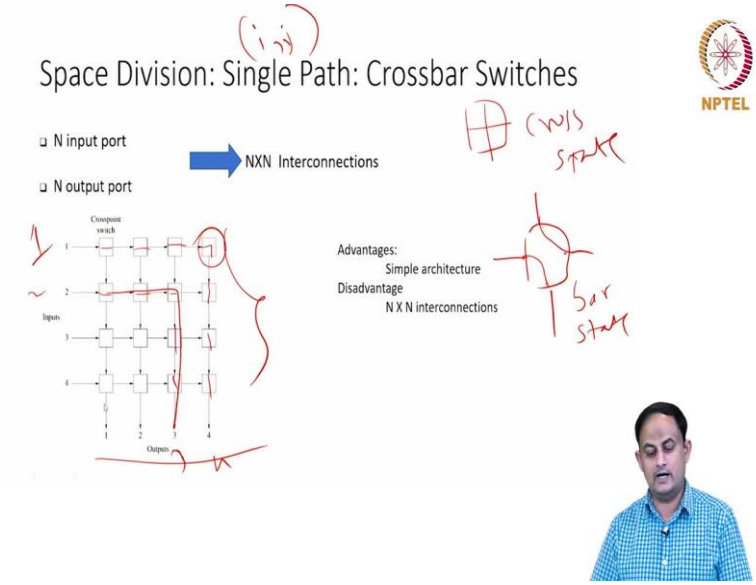
So, whatever Q's are there, for each of the output port they are actually organized in the said memory. So, if something is not occupying probably you can vary the amount of the space given to the other Q and then logically adjust where exactly the Q 1 starts and ends where exactly Q 2 starts and then something like that you can do. So, point is, there are 2 components one is the multiplexer and another one is the de-multiplexer they are sitting at the two endpoints of the switch fabric and inside this the link whatever is shown here, what here is the memory where the packets are actually getting received and then they are placed there inside that memory station.

So, you can very well schedule what is shown here as the with the arrow marks from the this is the color coding of the packet received on this particular port, may be the port number one, this is the coding of the packet received on the port number 2, and this is a color coding of the packets received on the port number n, when I say the received it is the input port number. And then you identify where exactly they need to go and then accordingly you can schedule them or taking the (())(10:52) packets are placed in that Q and then you can mark them and send it to the appropriate output port.

So, again, as only memory is one, so you can read the data from only one location at a time. So, this is a single operation. So, when I am selecting the when I am transmitting the data corresponding to packet. Let us say this is P 1 and I cannot really do or pass the contents from the packet or another packet to group are supposed to go to some other port number.

So, only one operation is happening at this point of time, only one read and one transmission. So, that is why this is a time division operation, so this is implemented with the said memory. So, these are the 2 things, one said memory and the other one is the said medium based implementation, both of them are in either of these 2 cases the transmission is taking turn, one packet from one input port is getting transmitted on to the output port.

(Refer Slide Time: 12:04)




And the second category as we saw is the space division where parallel transmissions are possible. And we understood that there are again 2 variants of the space division technique one is the single path and the multiple path. So, single path from the i-eth input port to the j-eth output port there is a single path exists. Here is an example of that kind of the switch designed called as the crossbar switch, the name of the switch technique the fabric crossbar comes from the 2 different states in which these elements these are called connection elements can be.

So, one in a case it can be in the cross state. So, this is the cross state and the other one can be in the bar state. So, these are the 4 points and on four sides we have a link and it can connect to two of them something like this. So, this state if the connection point is listed, then it is called as a bar state. So, all these connected elements can be either in the cross state or in the bar state. So, that is where the combination of these 2 states is called a crossbar switch. And the way the crossbar switch is organized is the switch fabric corresponding to the crossbar design is something like this.

So, on the left-hand side, what you see is a bunch of input ports and what is shown at the bottom here is the output ports. Here in this case, there are 4 input ports and the 4 output ports. And so, the depending upon the number of the input ports and output ports, the number of such crossbar points that are there inside this design is N cross N. So, you have N number of input ports and N number of output ports you will have N cross N number of the such interconnections.

So, one of the issue with this kind of design is the number of interconnection points that you require inside the switch fabric will grow quadratically. Meaning, so here in this case, you have 4 input ports and 4 output ports you require 16 number of the such interconnection point inside the switch fabric. And if you increase the number of the input and output ports maybe instead of 4 you make it 5 and 5, then you require 25 number of such interconnection points.

And one of the advantage of such a crossbar switch is there is no internal blocking that is happening here. So, for example, if 1 and 4 wants to engage in the conversation. So, all that I need to make is make this connection establishment, this link, this link, this link, can then come to this link and connect this and vertically also you need to add them. So, this will not block any other conversation not involve in port number 1 and 4.

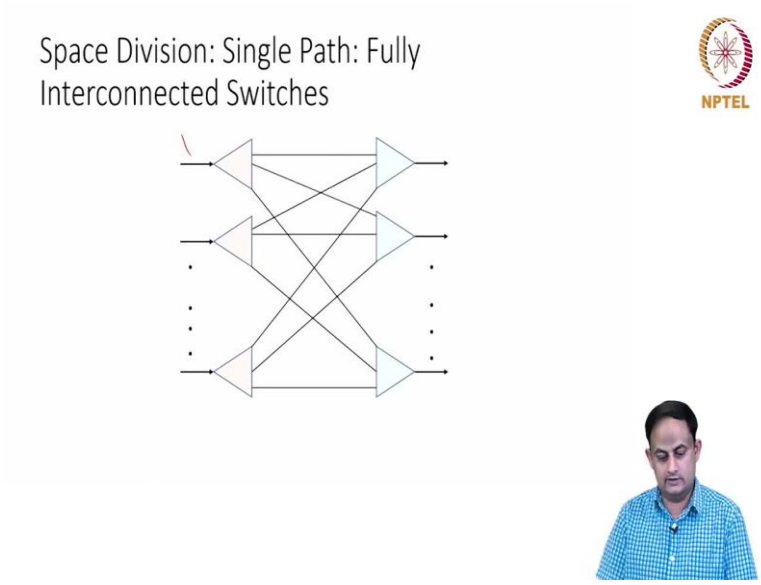
Meaning, if I want to engage in the conversation or 2 and 3, I can still do that, how do I do that? You establish this link, this link, this link, and then this and this vertically, so you come to the port number 3, 2 and 3 can also talk, similarly 3 and 2 and can also talk or 4 and 1 can also talk. And it might also happen that 1 can talk to 1, 1 can send 2 can send it to 4, 3 can send it to 3, something like that. So, as long as the input and output pairs are different, so you can still establish the unique connection between the input port to the output port and transmit the packet.

So, this is but there is only unique path between one pair of the input and output point. So, when I want to transmit from port number input port number 1 to the 4, this is the only link that is available, so that is the link I should take. And if I want to transfer to 2 and 3, this is the path that I should take, so nothing else, so that is how it has been designed. So, there is a unique path between every pair of the port numbers, but it is not internally blocking the other parallel communication as long as the input and output port numbers are distinct, you can still establish the parallel connection, parallel transmissions are possible.

As I said a single disadvantage of the such a switches is basically the number of interconnection points are increasing as the number of the ports are increasing on these (())(16:58). So, it internally it means that internally it means that the size of the switching equipment fabric itself will grow as the number of interconnection points increase. So, you need to package those many number of interconnection points inside your fabric, so size will grow and that will have other implications. So, as the size grows, your overall switch size or router size will increase, it requires larger space, cooling equipment and other things come

into picture. So, that is one major disadvantage, why this crossbar switch is actually not used in practice.

(Refer Slide Time: 17:40)



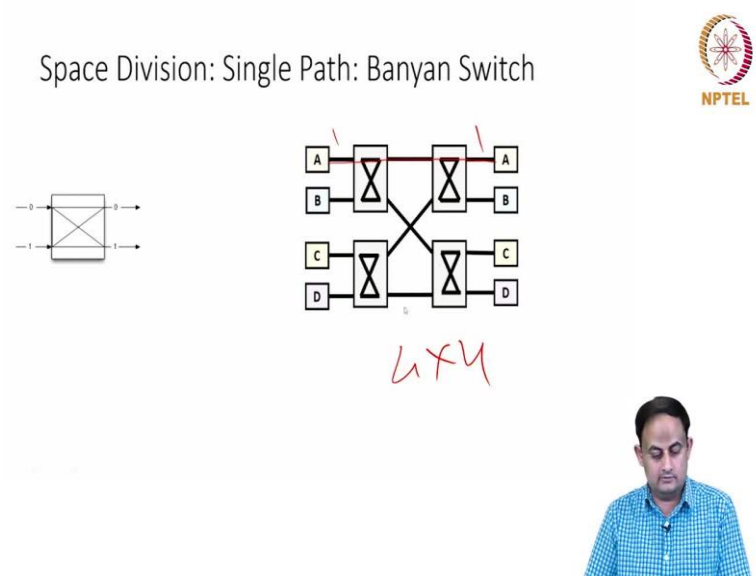

And the other case, second kind of the space division switch again with a single path is something called as the fully interconnected switch. So, the fully interconnected switch, the way to understand it is there is a link between every input port to the every output port and that link is a direct link there is no intermediate stages that are involved, unlike the previous case where when you want to go to from port number 1 to port number 4, you are crossing multiple interconnection points, those are not there.

And one of the simplest way to implement this kind of typically interconnected switch is to maybe have a bus which is actually connecting each of the input to port number to each of the outputs port number. So, that bus is dedicated to that particular input port, but it is connecting to all of the output ports.

So, when input port number 1 wants to transmit it can do an independent decision of transmitting and the probably what can happen is the packets or transmissions from the input port number 1, 2 or some other port number can collide at the output port and then thereby you bring the issue of output blocking that might happen.

But as long as the path that exists, the number of the links that exist between the input port and output mode are unique so that is why this interconnected switch is actually falls under the category of single path. There is a unique path and but the switches the input and output ports are actually directly connected with each other and that is the example.

(Refer Slide Time: 19:26)




The other way of looking at these were in order to bring the scalability to the Switch Design or the fabric number of the elements of the interconnection points that you want to place in the switch fabric. People found out different techniques to enhance how do I modularly build the for example if I take a component or switch fabric designed for connecting 2 different port numbers or 2 cross 2 into input ports or 2 output port and then using that as a building block can I extend modularly the and come up with the higher order designs.

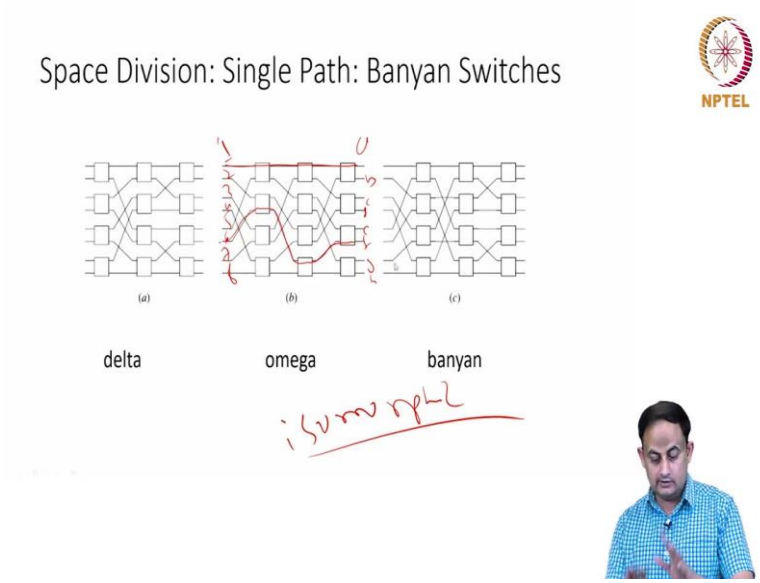
So, using the 2 cross 2 switch components, switch fabric design as a building block, can I build a 4 cross 4 switch fabric or can I build an 8 cross 8 switch fabrics, that is precisely what is done here. So, this kind of the design what is shown on the right-hand side here is something called as the Banyan switch. So, this is an example of a 4 cross 4 switch design with 2 cross 2 switch into the stages. So, here, this is still a single path switch, where the paths that exist from one input port to the one output port are still unique.

So, for example, if I want to go from the input port number 1 to the or A to A, so this is the only path that exists in this particular design, there is no other way you can reach the problem input A to another output A. So, for example, if you take this path particularly from this one and then you can go to the one then you can you will not be able to reach A. So, there exist in a unique path, but this is modularly designed using the 2 cross 2 switch fabric as a building block.

So, this is a you can think of as a way of extending the switches designed to the higher order. So, but there are other designs available as well. So, you can make it to multipath using the

same designs technique, if using the 2 cross 2 switch fabric design as a building block, you can also construct the multiple switches as well. So, we will take a look at one of those examples in the subsequent slides.

(Refer Slide Time: 21:51)




So, this Banyan switch as I was talking is a single path and switch, there are multiple ways of design that is what I said. So, here what you see on this slide, or the 3 different variants of the single path Banyan switches, in fact, these turns out to be isomorphic to each other. So, it means structurally they are similar, but only the way they look like the organisation the link they are actually looking slightly different. So, those designs are called as delta, omega and banyan.

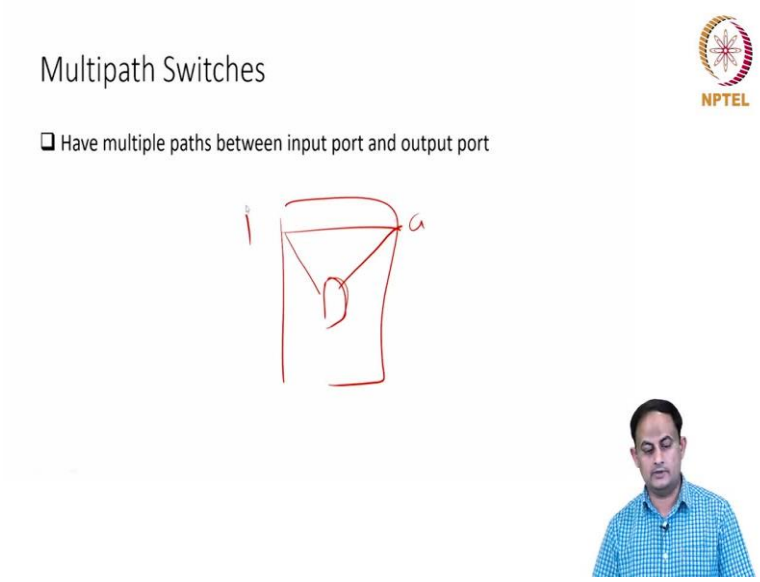
So, these are broadly falling under the category of the banyan switches, but they are structurally isomorphic to each other. If you think this is a graph with the interconnecting points as the nodes and the links between them as the adjust then these are structurally isomorphic to each other, they are exactly same isomorphism means that they are exactly structurally they are same, but the way they look like might be different, but they are similar, exactly the same graph structures, so this is an example.

Again, although the connections looked a little bit haphazard, but you can think of there is only a unique path that exists between the one port to another port. For example, if I want to go from the input port number 1 to output port number A, this is the only link that exists. So, if I take any other path, then I will not be able to reach it. So, similarly, if I want to go from

maybe I will number these as 1, 2, 3, 4, 5 and 6, 7, 8, there are eight input ports, the output port A, B, C, D, E, F, G, and H.

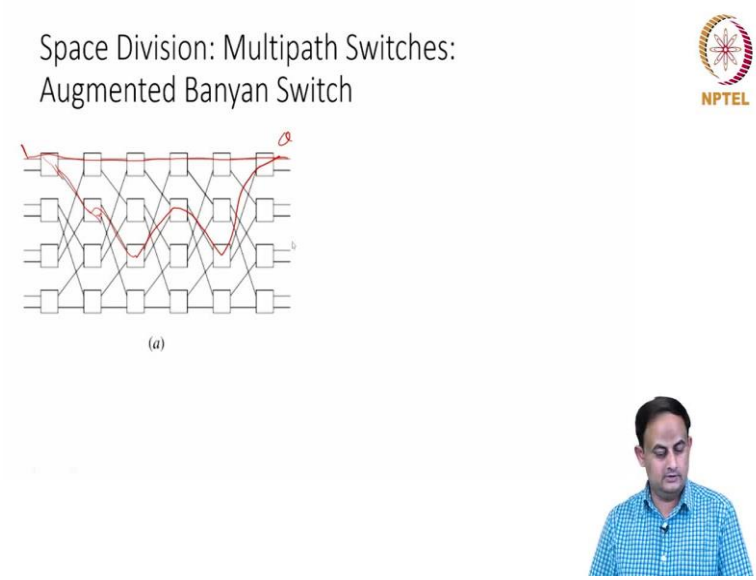
So, if I want to let us say I want to go from the input port number 6 to F, how do I go that? Maybe I should take this link go here and then go here, here and then go here and then go here. So, this is the only path that exit, if you take any other path then you will not be able to reach to the output port number F. So, that is what this single path between any input port to the output port. So, same argument holds good for the other combinations as well in this all the 3 variants.

(Refer Slide Time: 24:15)




So, that is the single path where any 2 pairs of the input and output ports are got the unique ports. Now, let us take a look at the multi paths switches where the from the one-unit input port to the one output port, there are multiple paths available. So, the simple way to look at it is. So, on the input port 1 and the output 1, so one link is this one and the second link is I may not be going directly to this one, I go to the intermediate connection point here from here to this this link is possible. If you have such a design then it is called as the multiple paths switches.

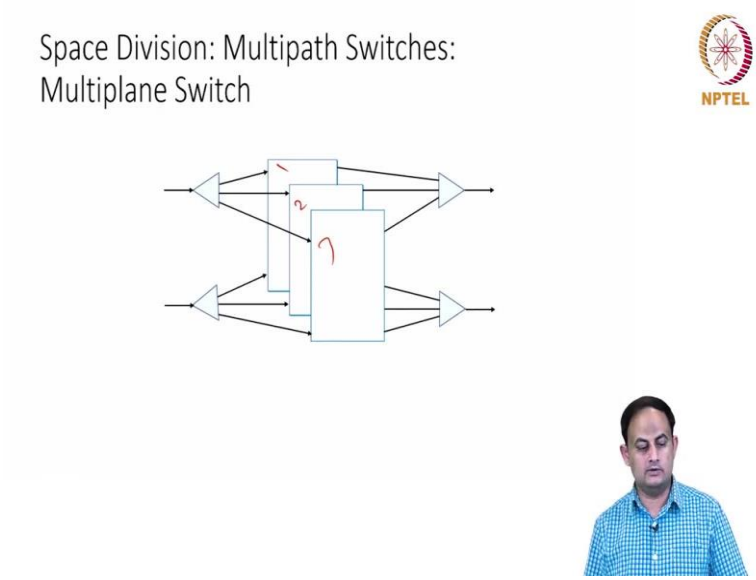
(Refer Slide Time: 24:56)




So, here is an example which is a variant of the banyan switch called as the augmented switch, augmented banyan switch. So, here there are multiple paths available from the input and output pairs. So, for example, if this is the input port number 1 and this is the output port number A, and one path from A 1 to A A is this direct link which is running in the straight line fashion. And the other way to possibly go there is to maybe I can go from here to here and this link and then this link, this link, this link and then go here connect.

So, this is the connection between the input port number 1 and the output port number A. So, there is 2 distinct paths one in the straight-line fashion the other one that I just drew. So, these 2 distinct paths are exists. So, by virtue of that, this is a multiple design. The similar argument actually holds good for the other input and output port pairs as well. So, there are indeed multiple paths from every input port to the every output port. So, such a design is called as the multipath design and that is one example.

(Refer Slide Time: 26:21)

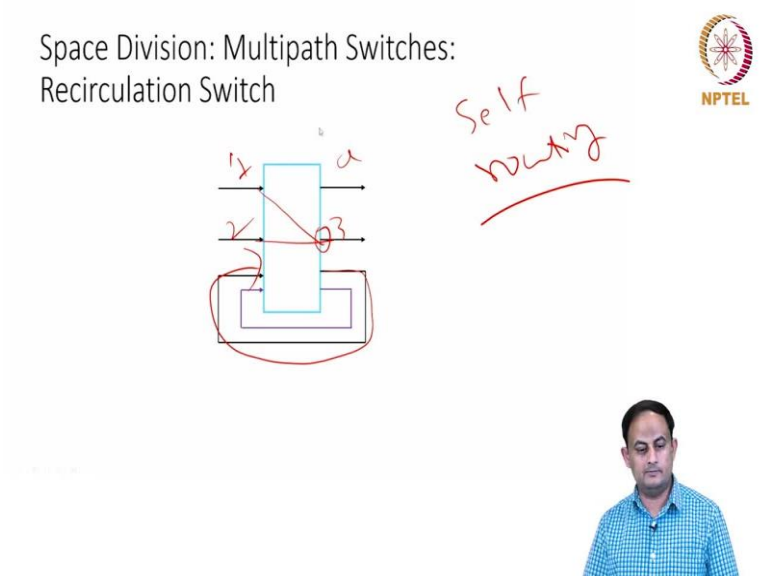



The other example is something called as a multi plane switch, where there is not one the way to understand the multi plane switch is to again there are some input ports and output ports and there are multiple switch fabrics sitting in between which are actually operating independently. So, what is shown in this diagram is there are 3 planes plane number 1, 2 and 3. So, you can very well think of them as replicated versions of this switch fabric, I can very well come up with whatever design I want to do for the design the switch fabric number 1 2 and 3.

So, let us say I go with the crossbar and there is one plane with one crossbar there is a second plane with the second crossbar something like that, by virtue of replicating them I bring the parallelism in the transmission. So, these some input port can to some other port and do I can do the parallel transmission. So, from i and j, a and b something like parallelly they can transmit it.

And again since these every plane is connected to the every input port and every output port there is an inherent to multiple paths that exist in the transmission. So, one plane is connecting to every input and output port. Even though it might be transmitting only doing one time can connect to one of the input and output ports because there are multiple such planes available, so you can think of because, the second plane is also connected to the same input port and the output port you can bring the parallelism in the transmission. So, that is an example of the again multiple path switch.

(Refer Slide Time: 28:01)




So, often what can happen is or you can recollect our earlier discussion, there is a processor which is actually in the control in it, which actually establishes a link between the input and output port. And we understood that there are 2 categories of them one is the centralized processor and the second one is the distributed, the distributed processing mechanism is often also called as the self-routing and what it means is the decision of transmission can be done independently on each of the port.

So, when a packet is available at the input port number 1, that control unit can take the decision for forwarding or establishing a link. And on the second input port maybe on port number 2, there is in a second component sitting that can also independently decide. And once we lose the central control, because the forwarding decisions or establishing the links between the input and output ports are distributedly done. So, the 2 of them can parallelly decide at the same time instant t they can decide I can transmit it to port number 3.

So, for example, 1 also wants to establish a link to port number 3 and 2 also wants to establish the link to port number 3, and here is your port number 3. Then, because these decisions are not centrally coordinated, they are independent what might happen is, you end up with what we described as the output blocking, both of them want to go or access to the same output port. Now, equal two different transmissions from two different ports by virtue of two independent decisions start arriving at the output port number 3, you will have an issue, you will need to arbitrate them and then you are need to ask someone to take it back.

So, this is a switch design which is called the recirculation mechanism where by virtue of the arbitration you are able to decide one of them you are going to pick up one of them. So, either 1 is going to transmit or 2 is going to transmit and whatever you are not transmitting at this point of time the bits corresponding to that packet need to be put it back to the same queue from where it came, so that is what is done here.

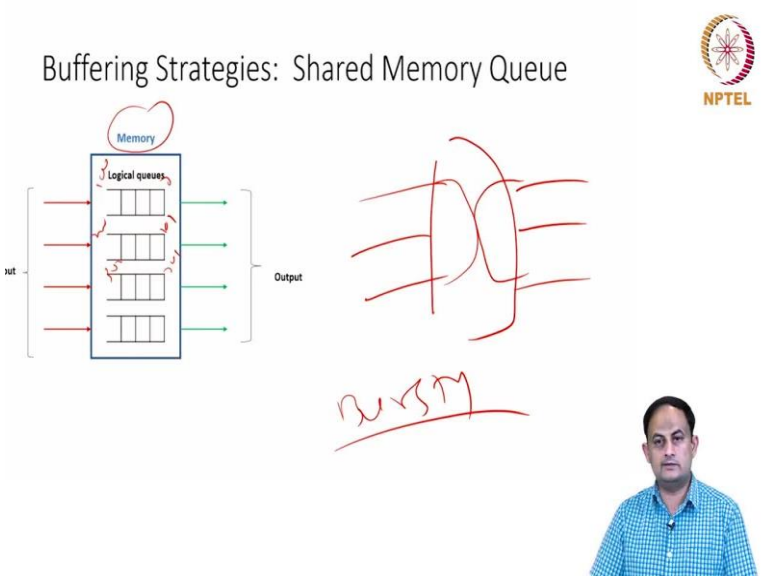
So, because of the output blocking and independent decision of the distributed transmission decision, distributed transmission processing happening in the switch fabric, two different transmission sending up at the same output port and what because of that tree you have the output blocking and then you are taking back one of the packet or the transmissions corresponding to one off the packet to back and put it in putting back into the same queue from where it came from. So, such a design is called recirculation.

So, here you have the multiple paths available between one pair input port to the output port, parallel transmissions are possible, but those parallel transmissions are not like this, as long as they are independent then you can do the revision. So, again there is from the one input mode to the one output mode there are multiple transmissions, but only one path you can take at a single point of time and then do the transmission.

So, you cannot really use the parallel paths that exist for transmission from the same input port number 1 to the same auto port number a and assuming there are 2 different paths exit, it does not necessarily mean that I will be able to transmit 2 different packets from the same input port to the same output port.

From 1 to a there are two different packets are not parallel transmitted, only one packet is transmitted you take either the path 1 or path 2 by virtue of that, you are making use of only one of the existing path in the switch fabric. But in a way you can understand this as the alternative things are given to you, when you are not able to make use of the other path then you go where alternative road so that you maximize the number of the parallel independent transmissions in the switch fabric, so that is the recirculation switch which is of the type multipath switch type.

(Refer Slide Time: 32:50)




So, let us spend some few minutes in understanding the something called as the buffering strategy. What we say it is the end user topic is burst in nature or any transmission in the network is burst in nature. So, burst in nature is you will have a surge of the transmissions to shorter spirit of span of time and then there might be something where you will not be having enough transmission when you have got bursty transmission.

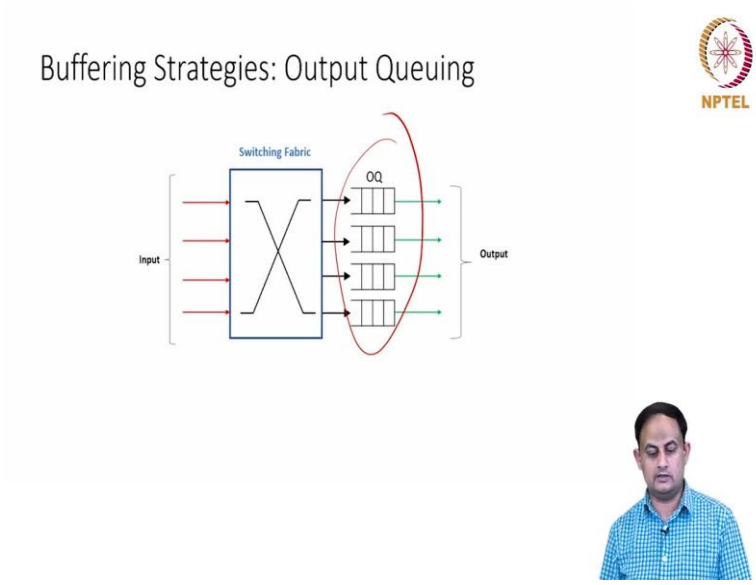
So, the transmission rate, the rate at which the switch fabric can actually transmit from one port to another port or some input ports, the output port is limited and the even the input lines are also got or the output lines or got the finite capacity with which they can do the transmission. And if there is bursty transmission, then probably you require a mechanism to hold those packets inside the switch, so that is what where the buffering comes into picture. This is fundamentally to handle the bursty traffic in the network.

So, now what this buffering strategy about is, where exactly I am going to place this buffer, where exactly I am who are going to hold these packets. So, one mechanism that we earlier touched upon is something called the logical queues, where we got a single memory that could very well be a RAM. And then you earmark this is the portion buffer for the port number 1, this is the portion from 0 to address number 0 to 100, this is the queue for the port number 1 from 101 to 200. You have the second queue and from 201 to 300 third queue something like this.

So, the memories 1 but you logically divide that memory and then you put the when the packet comes an input port, you decide which output port it is supposed to exit and then pick

up that packet and put it into the queue or the memory portion corresponding to that output queue, this is one way of organizing or storing the packets inside your switch.

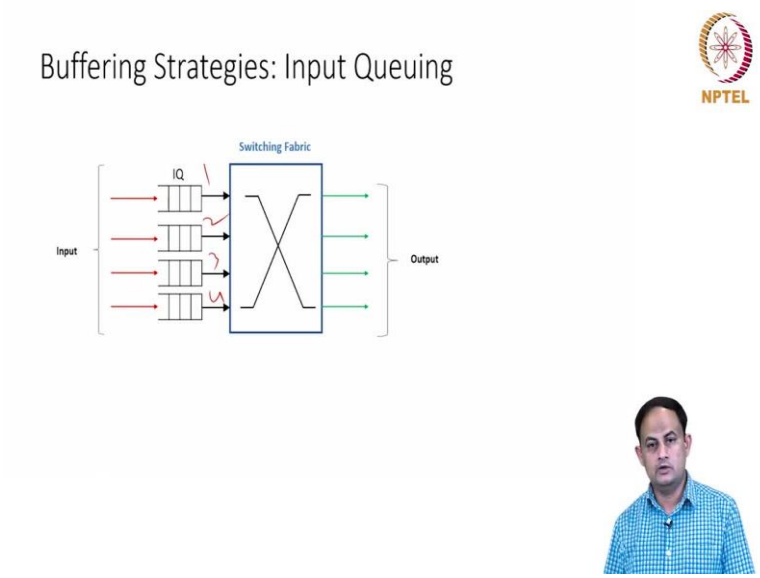
(Refer Slide Time: 35:18)

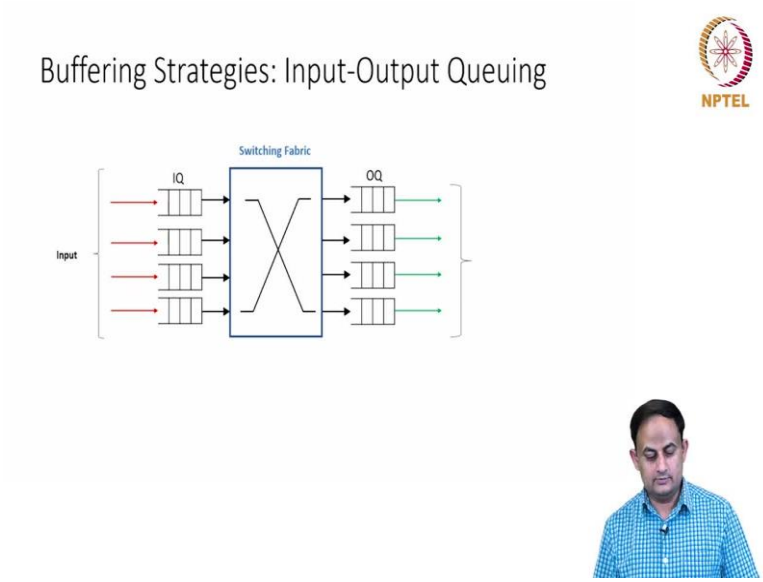



And the second strategy of replacing the queues is you might have very well have young different buffers available at each of the output port and packets are coming on the input port and then you switch fabric is making establishing a connection and then you come back and put them in the respective output ports.

So, n independent memory, n independent queues and each queue correspond to one of the output port then you transmit the you schedule the packets for transmission from that output queue and then do the transmission. And here there is no buffering that is happening at the input queue, so the packet will be able you will be able to transmute that packet only when you are able to establish a connection from the input port to the output port right here otherwise the transmission the packet will be lost.

(Refer Slide Time: 36:18)




The third strategy of the buffering is instead of doing the or buffering at the output port, you do the buffering at the input port as and when the packets arrive, you put them inside the buffer that is the input buffer. So, there is there are as many buffers as many queues as there are port numbers. So, there is one queue for the input port number 1 and there is one queue for the input port number 2, 3 and 4 something like this. And you put the packets there, and then you look at when the switch fabric is able to establish a connection for your transmission that time you pick up the packet from that queue and do the transmission.

Fourth way of looking at the buffering is or the way it is done is we placed the buffers in both places, you do have a queue at the input port and also you have a queue at the output port and in fact, most of the examples that we saw in the previous discussion in previous lectures, we

did say that there is a buffer input buffer and then there is an output buffer. So, here is an example of that.

So, there are n different input memories, n different output memories and then the packets are coming, you place them inside the input queue, then do the FIB consultation, we established the connection through the switch fabric placed them in the output queue and then you can do the scheduling operation. So, scheduling is done both at the input and the output queues as well. So, this is a fourth strategy. So, the all of these can be very well handled the bursty transmissions that usually happen in the network. So, with that, I will stop it here. Thank you.