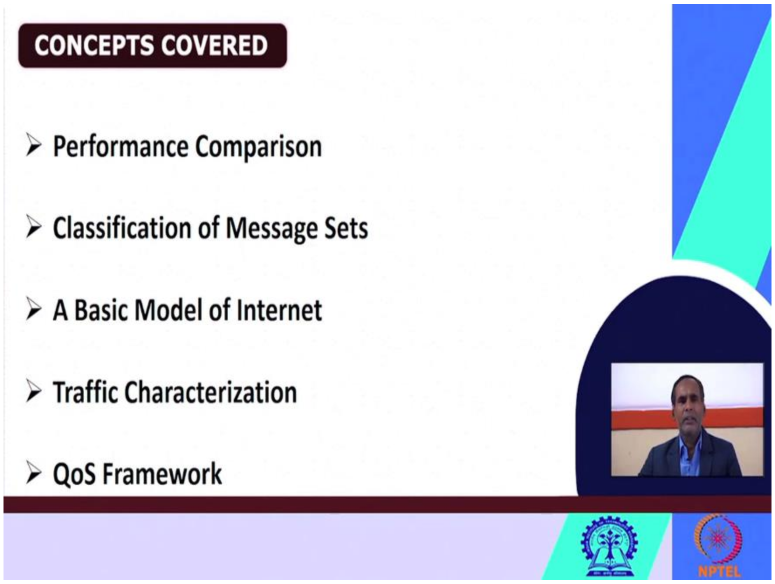
Real Time Systems
Professor Durga Prasad Mohapatra
Department of Computer Science and Engineering
National Institute of Technology Rourkela
Lecture 53
Performance of Comparison and QoS Framework

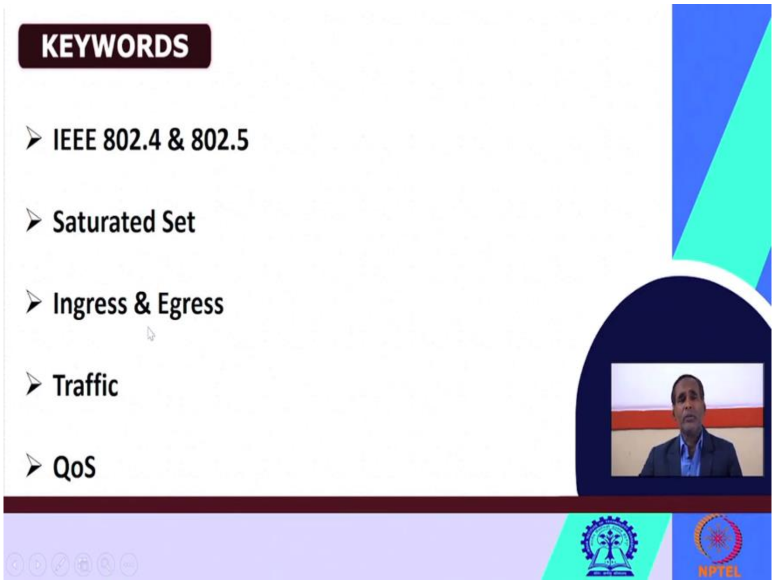
Good afternoon to all of you. Now, today we will discuss about the performance comparison of some of the protocols we have discussed in the last class and the QoS framework.

(Refer Slide Time: 0:30)



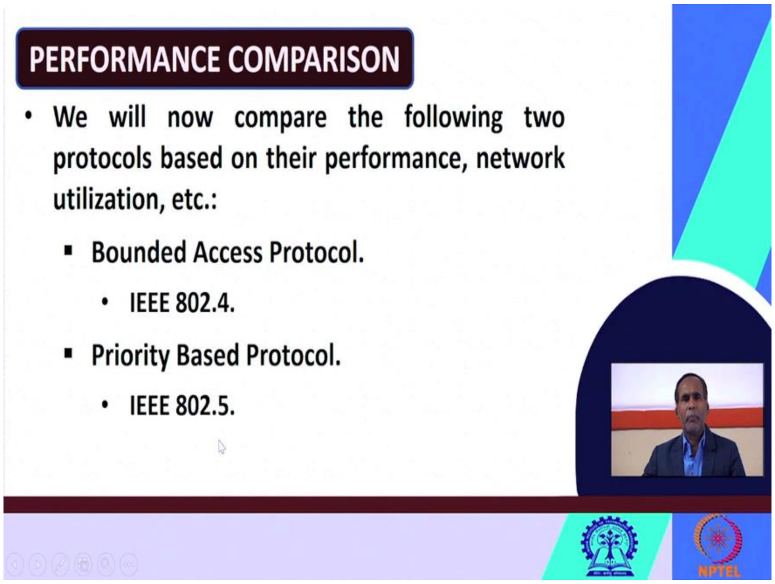
So, we will first cover about performance comparison of some protocols especially IEEE 802.4 versus IEEE 802.5, we will classify the message sets into different categories. We will see a basic model of internet. We will see some of the traffic characterizations. Then a framework for QoS quality of service.

(Refer Slide Time: 0:52)



So, who we will use, we will compare this IEEE 802.4 and IEEE 802.5 protocols? We will see what a saturated set? What is the unsaturated set? Those things you will see, some routers you will see specifically ingress routers, egress routers and what your core routers, we will see what is a traffic and what is a QoS.

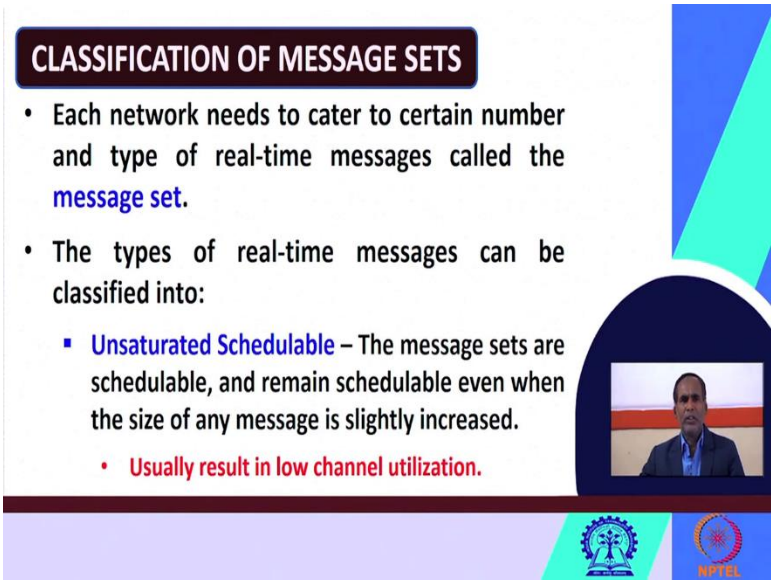
(Refer Slide Time: 1:16)

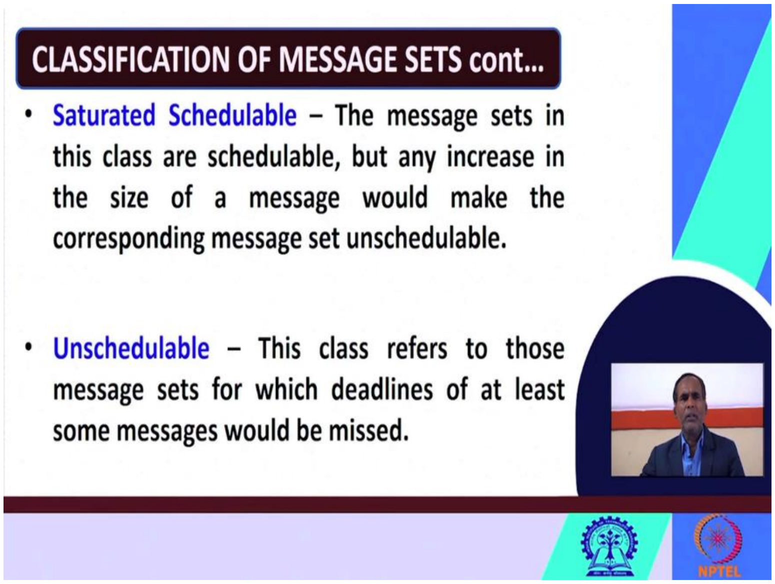


So, first thing about the performance comparison. Today we will compare two important protocols based on their performance network utilization et cetera. First, we had seen many priorities best protocols such as the countdown protocol. And then these virtual time protocol window-based protocol and the IEEE 802.5 protocol last class we have seen two input on bounded assets protocol IEEE 802.4 and RETHER net.

So, today we will try to compare at least one protocol from bounded assets protocol and one protocol from priority-based protocol. So, we have chosen IEEE 802.4 and the IEEE 802.5 for comparison. So, now we will compare IEEE 802.4 with IEEE 802.5 based on that performance network utilization and those things and similar things.

(Refer Slide Time: 2:07)





So, before going to the comparison first let us see about the classification of the different message sets, each network needs to cater to certain number and a certain type of real time messages called as the message set. Each network needs to cater what to certain number and certain type of real time messages which are called the message set. The types of real time messages can be classified into some categories. First one is unsaturated saturated schedulable.

Then saturated schedulable then unschedulable. So, the types of real time messages can be classified into three important categories, let us first see the messages which are on saturated schedulable. The message sets are schedulable that means were first dependent unsaturated schedulable this type of messages says there are all schedulable and remains schedulable even when the size of any message is slightly increased.

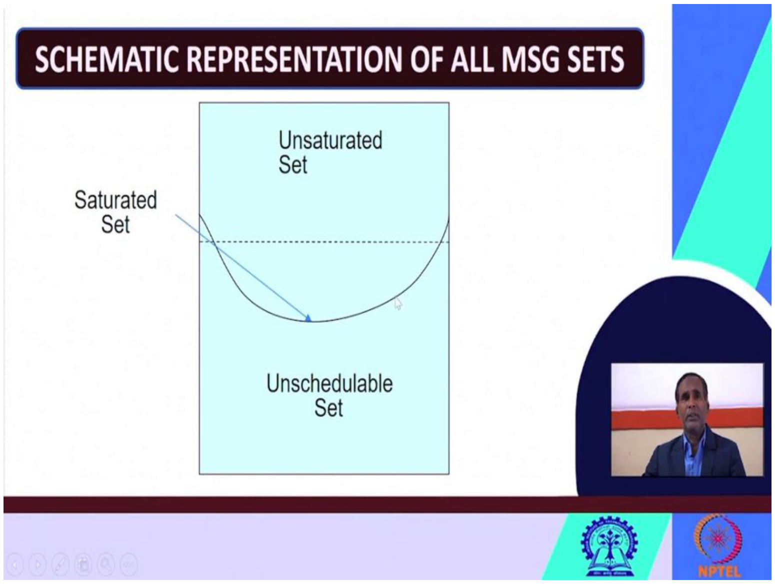
I am defining now unsaturated schedulable messages. So, the message sets which are schedulable and still remain schedulable even when the size of any message is slightly in increased, we call those messages sets are unsaturated schedulable message sets or messages. Usually, these message sets they result in low channel utilization. Because they are what this message sets are schedulable and still, they remain schedulable.

We call them as unsaturated schedulable message sets; they result in low channel utilization. Then you will see saturated schedulable they are slightly different than the unsaturated schedulable. Here the message sets which are schedulable but any increase in the size of the message will make the corresponding message set unschedulable currently they are schedulable but even increasing the size of the message then it would make the corresponding message sets unschedulable.

So, we call them as what saturated schedulable. So, saturated schedulable means this is the set this is the message set in which the messages are schedulable but any increase in the size of the message it will ultimately, it would make the corresponding messages sets unschedulable. Then unschedulable this is very clear cut this class of messages refers to those messages sets for which the lines of at least some messages should be missed.

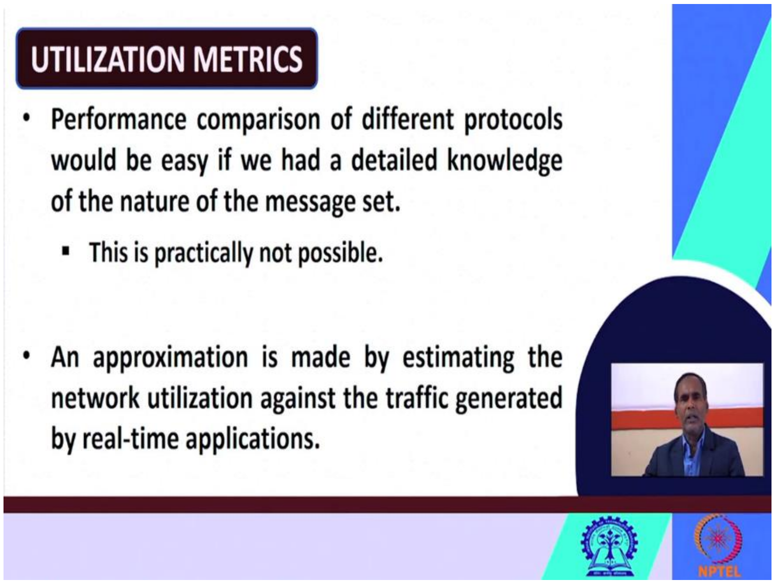
Unschedulable means this is the class of messages which refers to or these class reports to those message sets for which the deadlines of at least some messages would be missed. So, they are not completely schedulable. In this case, the deadlines of at least some messages would be missed. We call them as, this set we call as unschedulable message set.

(Refer Slide Time: 4:55)



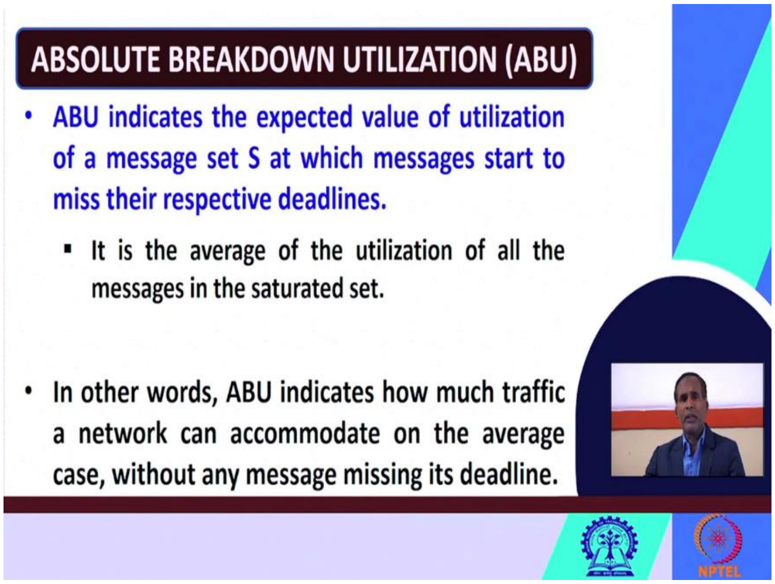
Schematically we can represent all message sets we can represent as follows. We can see here what, here the top level these messages represented unsaturated set. And below this line these are unschedulable set and the saturated set follow the curve like this. So, this is how the schematic representation of all message sets looks like.

(Refer Slide Time: 5:18)



We will now see some of the utilization metrics then based on the utilization metrics we can compare IEEE 802.4 versus IEEE 802.5. So, the performance comparison of different protocols will be easier if we had a detailed knowledge of the nature of the different message sets but this is practically not possible. So, an approximation is made by estimating the network utilization against the traffic generated by the real time application. So, let us see about this network utilization and maybe about the traffic.

(Refer Slide Time: 5:54)

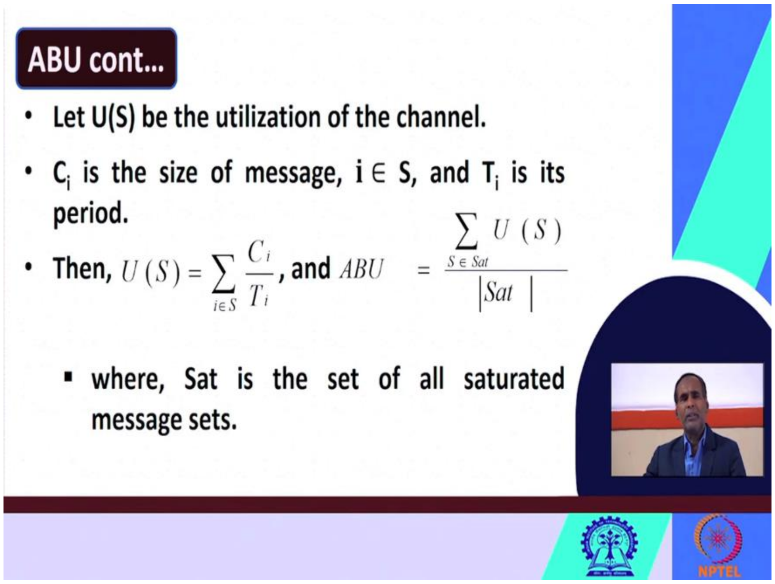


We will see the first metric as absolute breakdown utilization in short, we call as ABU. What does it indicate ABU indicates the expected value of utilization of a message set S at which the messages start to miss their respective deadlines that means after which the messages they will start to miss their respective deadlines? I am repeating again absolute breakdown utilization it indicates the expected value of utilization of a message set S at which the messages will start to miss their respective deadlines.

So, this ABU it is the average of the utilization of all the messages in the saturated set. ABU is defined the average of or the utilization of all the messages present in the saturated set. In other words, ABU it indicates how much traffic in network can accommodate at the based the ABU the network can accommodate what is the what, how much traffic without any message missing its deadline, without missing, without any message missing its deadline how much traffic a network can accommodate on the average case or on an average this is indicated by ABU.

So, in other words, ABU indicates how much traffic a network can accommodate on the average case without any message missing its deadline.

(Refer Slide Time: 7:19)

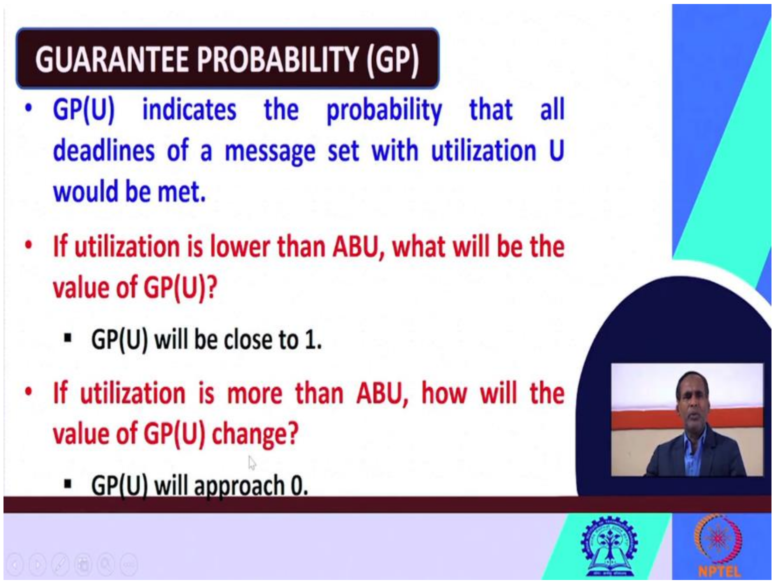


So, let us see an mathematical expression for a view let us assume that the U(S) be the utilization of the channel. Let U(S) represents the utilization of the channel, Ci represents the size of the message where i ϵS and the Ti is its period. Then utilization of the channel U(S) can be defined as $\sum_{i \epsilon S}$Ci / Ti. So, what is Ci / Ti we have already known this in the utilization Ci the size of the message where Ti is its period.

And this is how we can define the utilization of the channel. Similarly, how you can define ABU? ABU can be defined as what is this utilization take the summation S for that means for all the sets in the saturated set. Then divided by cardinality of Sat. What is cardinality of? What is Sat? Sat is the set of all saturated message sets. ABU is defined as it is equal to the summation of U(S) where S belongs to saturated message sets divided by Sat.

Where Sat is the set of all saturated message sets. So, in this way you can compute the U(S) that is utilization and the ABU. So, now we have known two metrics now can compare these IEEE 802.4 and IEEE 802.5.

(Refer Slide Time: 8:44)

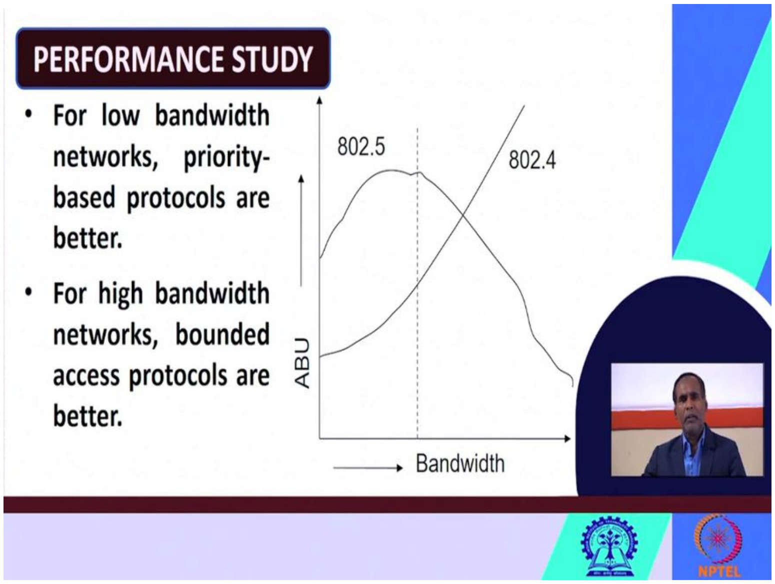


Before going to that another metric, I want to say that is called, before comparing the performance let us see another metric called as guarantee probability in short, we write as a GP. So, guarantee probability of U. So, GP(U) it indicates what? It indicates the probability that all the deadlines of a message set with a utilization U would be made that is why the name is guarantee probability. So, GP(U) indicates what?

The probability, what is the probability that all the deadlines of a message set, all the deadlines of a message set with utilization U they would be made this much guarantee you up to give. So, this can be given by this metric guarantee probability GP or GP(U). If utilization is lower than ABU then what will happen if utilization is lower than ABU what will the value of GP(U), very simple GP(U) will be close to 1.

If utilization is more than ABU, how will the value of GP(U) will change? GP(U) will just approach to 0. So, these are the fundamental things you please remember.

(Refer Slide Time: 9:48)

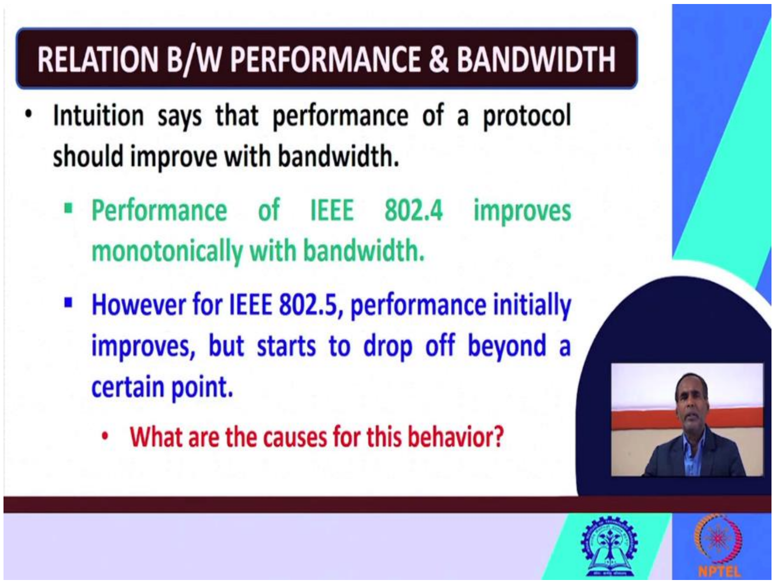


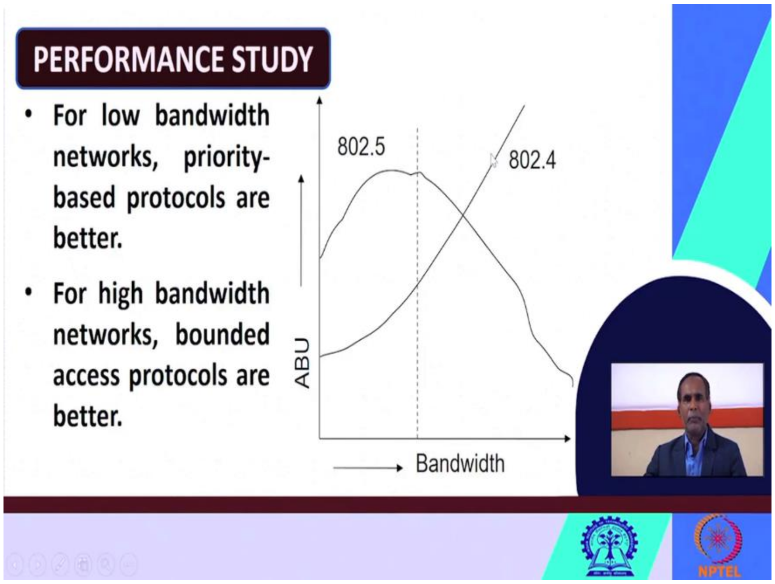
Now we will compare 802 to 802.5 and 802.4. So, you can say that 802.5 we did this what the global priority protocol is an example of global priority protocol whereas 802.4 is an example of bounded SS protocol. You can see that in x axis we have taken bandwidth y axis we have taken ABU for low bandwidth this can see here this low bandwidth for low bandwidth networks you can easily see 802.5 performs better.

That is priority based protocols, the priority based protocols they are better but for high bandwidth like convert if the bandwidth increases for a high bandwidth which will give which one will better. This 802.4 will better that means the bounded access protocols are better I am summarizing again for low bandwidth you can see ABU is better in where case ABU is more in case of 802.5 that means for low bandwidth priority based protocols are better.

And for high bandwidth ABU is better ABU is increasing here. Because 802.5 is falling down. So, ABU is increasing for 802.4. So, for high bandwidth from the graph you can see that 802.4 protocol or the bounded access protocols they are performing better. This is how we can compare 802.4 and 802.5 based on what this ABU.

(Refer Slide Time: 11:06)



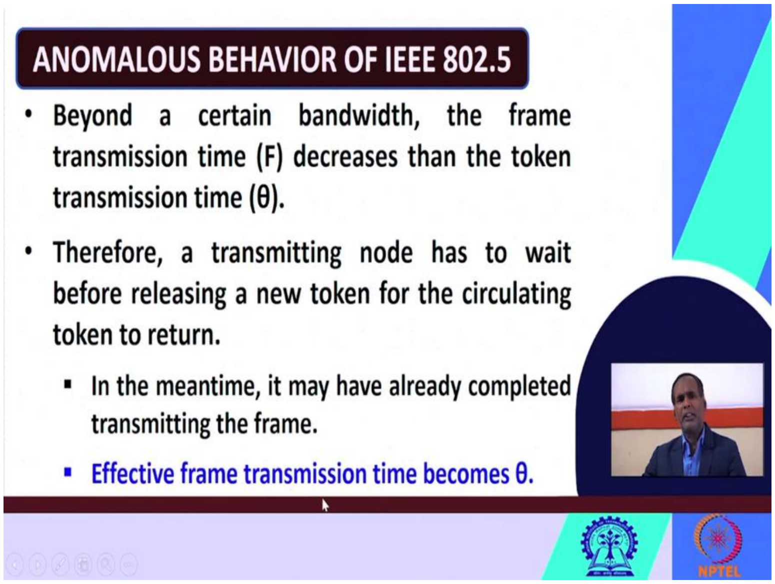


Now let us say the relation between the performance and the bandwidth, normal our intuition says that the performance of a protocol should move improve with the bandwidth, is not it? Our common sense says that the performance of a protocol it should improve with the bandwidth. But you can see the performance of IEEE 802 improve monotonically with bandwidth. No problem 802.4 if bandwidth increases the performance improves, no problem.

But what happens in case of IEEE 802.5 the performance initially improves but starts to drop up beyond a certain point you see. Initially for 802.5 the property performance improves if ABU value improves but at certain point from this point you see the from this point then this what performance

ABU value it is gradually drops it falls down. Why? Why this is happening? So, IEEE 802-point performance initially improves but starts to drop up beyond a certain point. What are the causes for this behavior?

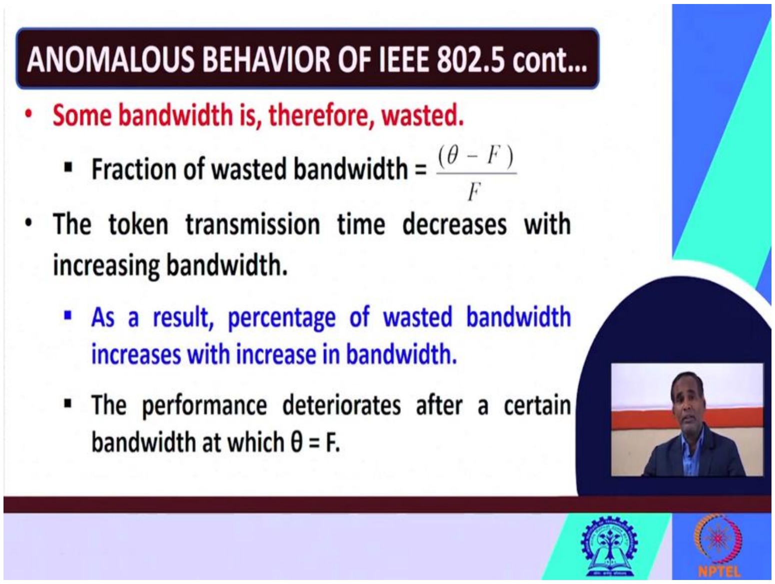
(Refer Slide Time: 12:08)

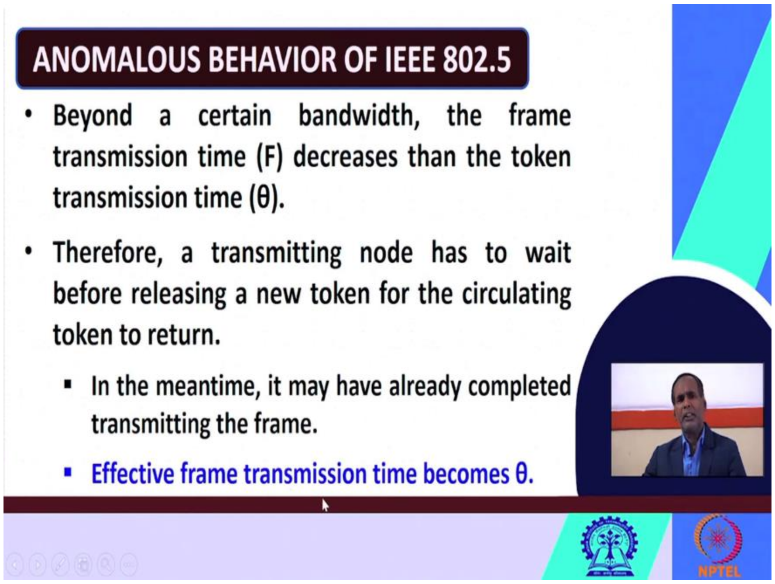


So, let us see what are the causes for this behavior, you know that beyond a certain bandwidth the frame transmission time F it is decreases then the token transmission time theta last class we have already discussed in the what the formulae and why we have known what is frame transmission time, what is a token transmission time you can see that beyond a certain bandwidth the frame transmission time F it is decreases.

Decreases then what? Then the token transmission time theta therefore what will happen it transmitting node it has to wait, is not it? So, since this frame transition time decreases then the token transfer time therefore a transmitting node it has to wait before releasing a new token for the circulating token to return in the meantime what will happen it may have already completed the transmitting the frame. So, the effective frame transmission time it will becomes ultimately theta.

(Refer Slide Time: 12:58)



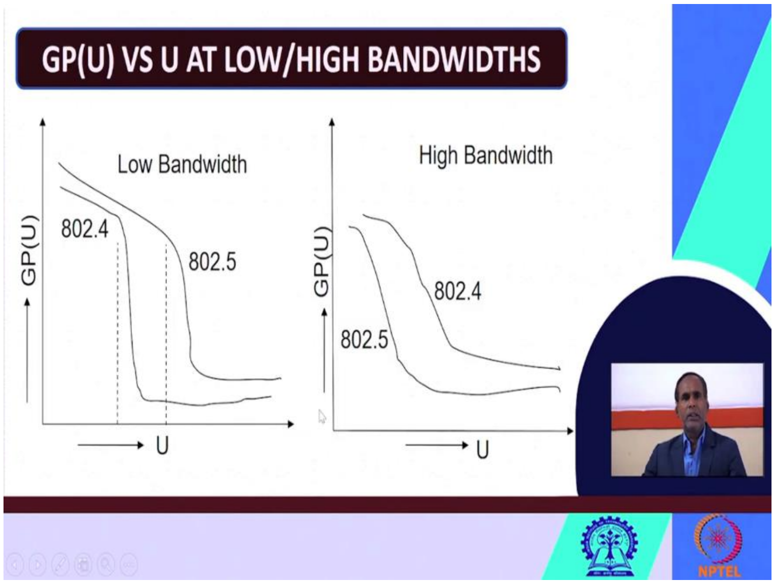


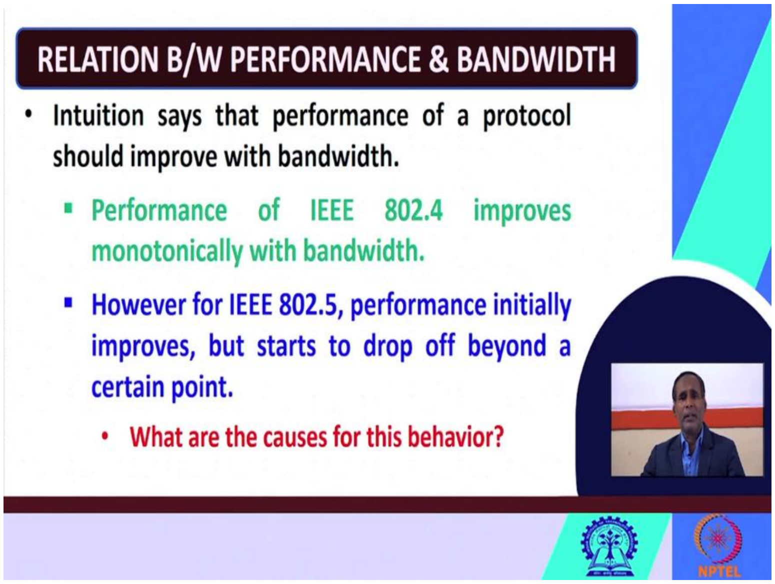
You know that some bandwidth is therefore wasted because initially our what frame transmission time is F now it decreases and, in the meantime, it may have already completed transmitting the frame. So, the effective frame transmission time it becomes theta. So, some bandwidth is therefore it is wasted. So, now let us find out how much bandwidth or what is the fraction of the bandwidth that is wasted.

So, fraction of wasted bandwidth can be computed as what is the value of theta minus F divided by F, the token transmission time decreases with increasing bandwidth. So, since here you see some of the bandwidth is what wasted. So, the token transmission time it decreases with increasing with increase in bandwidth. As a result, what will happen? The percentage of the wasted bandwidth it increases with the increase in bandwidth.

So, the bandwidth increases the percentage of wasted bandwidth is also increases. So, this is the reason here what will happen the performance deteriorates that is why the performance deteriorates after a certain bandwidth at which the theta is equal to the F. So, when theta is equal to F when theta becomes some equal to F then the performance it deteriorates after certain bandwidth. So, after a certain bandwidth at which theta is equal to F the performance starts deteriorating or the performance deteriorates.

(Refer Slide Time: 14:23)

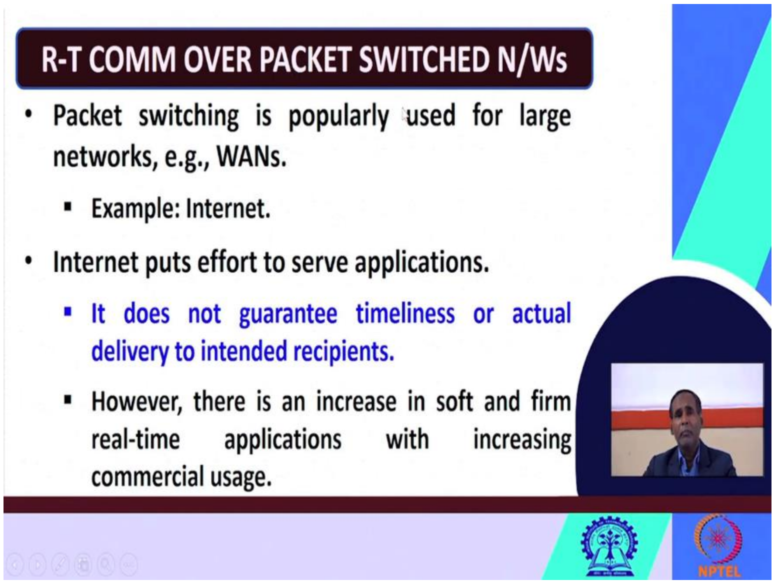




So, now you can say, well let us, so that is why you have seen what is the reason why the I, initially the IEEE 802.5 it initially improves. So, why then after some point it starts to drop of the region I have just explained now. Now we will compare this guaranteed probability versus this utilization at low and high bandwidths. If you will see, at low bandwidth so because the in case you will see this is the lower bandwidth.

And in low bandwidth you can see that is 802.5 it performs better and for high bandwidth if you can see which one is performing better 802.4 it performs better. So, as utilization is increasing for lower bandwidth 802.5 performs better you can see from the graph. And as the utilization increases. So, at higher bandwidth 802.4 it performs better than my 802.5.

(Refer Slide Time: 15:28)



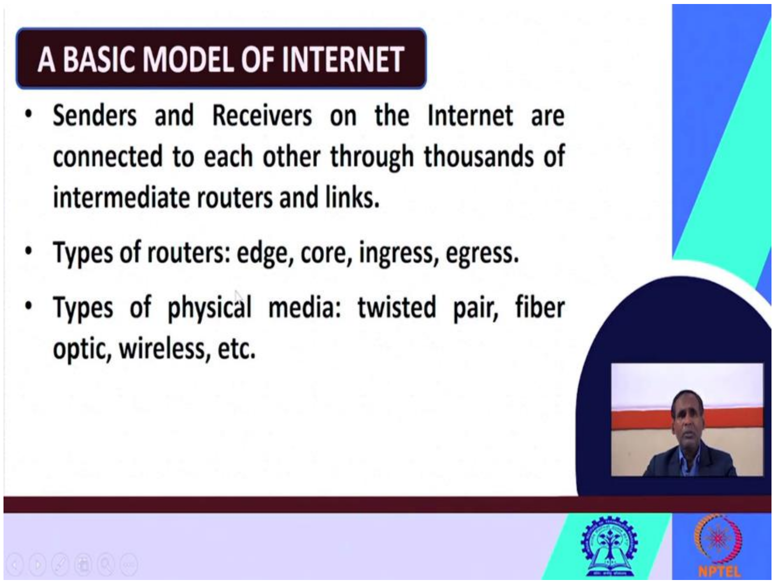
So, this is how we have seen that real time communication for LAN or real time communication for LAN we have seen we have seen the different protocols or charge in this global priority based protocols and this calendar based protocol and bounded access protocol, we have compared some of the protocols for example we have compared 802.2 and 802.5 protocols. Now, we will see a real time communication over packet switch network.

I have already told you one of the examples of packet switch network is what? Internet. So, basically, we will see real time communication over internet. Already we have seen real time communication over LAN now we will say the real time communication over internet which is a very good example of packet switched network. You know that packet switching is very much popularly used for large networks for example wide area networks, example of packet switching network I have told you internet.

So, internet puts effort to solve the applications, is not it. Internet puts the best efforts to solve the applications, it does not guarantee timeliness or actual deliver to internet recipients you are all using internet it does not give guarantee the timeliness. It does not give a guarantee it does not guarantee on the timeliness or actual delivery to the internet recipients. However, there is an increase in soft and firm real time applications with increasing commercial usage.

So, even if this is the drawback it does not guaranteed timeless or actual delivered to internet recipients. So, however there is an increase in soft turned firm real time applications with increasing commercial usages. So, this real time communication over packet switch network mainly over the internet it is becoming popular nowadays.

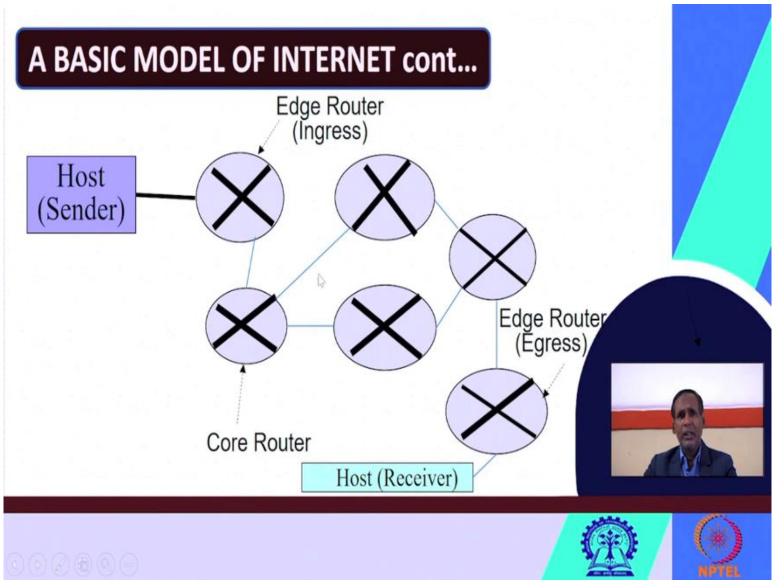
(Refer Slide Time: 17:10)



Now let us quickly look at a basic model of the internet. So, normally the senders and receivers on the internet they are connected to each other through thousands of intermediate routers and links, is not it. In an internet the senders and receivers they are connected to each other through what? Through thousands of internet routers and the links. So, various types of routers you have known like edge routers, core routers, ingress routers, egress routes et cetera.

So, the type of the physical media can be used it can be twisted media, twisted PR, it can be fiber optics even if the medium the medium can be a wireless medium. So, this is how these are the different components of internet.

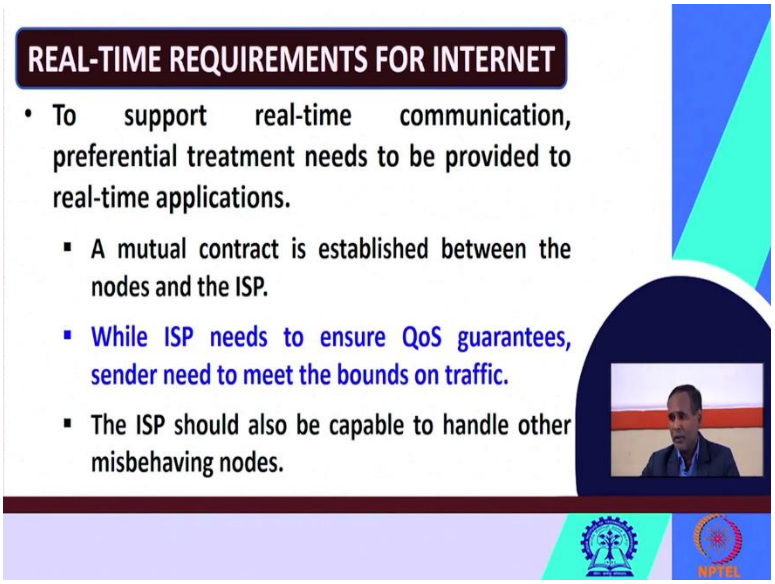
(Refer Slide Time: 17:51)



If you will see in a picture via you can represent like this here normally the edge routers, they are connected to the peripherals that means to the host or to the sorry to the sender to the receiver. So, the router that is or you can see that the sender is connected to a router to an edge router called as the ingress routers. And this receiver is connected that to the edge router called the egress routers.

So, that is why I am saying edge routers, they are connected to the peripherals for example ingress router is connected to the or the host or the sender is connected to the network through ingress router and the receiver is connected to the what a network through the egress router both ingress and egress routers they are examples of edge routers. All those intermediate routers for example this, this and this they are examples of core routers and these are the links. These are what? These are all the links. So, this is how a basic model of internet it looks like.

(Refer Slide Time: 18:52)

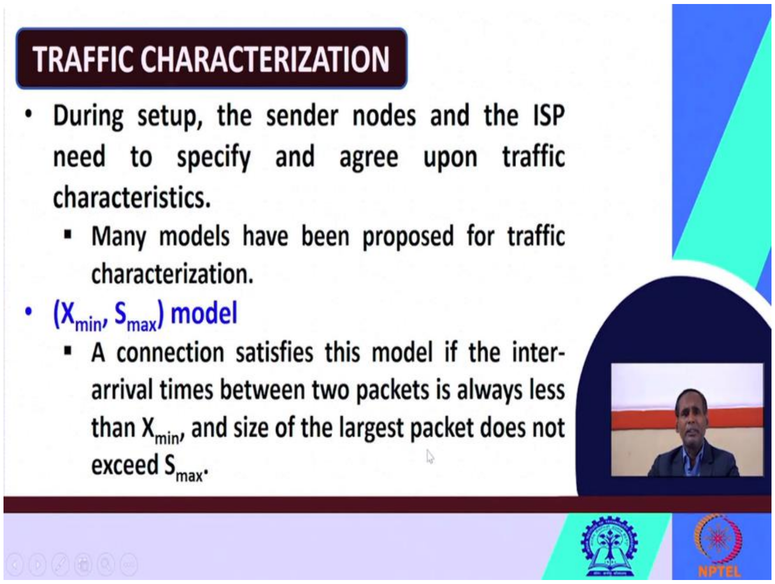


So, the real time requirements for internet let us see what are the basic real time requirements. To support real time communication preferential treatment needs to be provided to the real time application. So, the preferential treatment needs to be provided to the real time applications a mutual contract established between the nodes and the ISP. So, to support real time communication a mutual contract has to be established between the different nodes and the ISP the internet service provider.

While the internet service provider it needs to ensure the what QoS guarantees the sender need to make the bounds and traffic. Now let us see what is the responsibility, while the internet service provider needs to ensure what QoS guarantees the sender need to meet what the bounds on the traffic the ISP should also be capable to handle other misbehaving nodes. So, the you know always out of all the nodes some of the nodes they misbehave they may not be properly.

So, who should take care of this, who should handle this? The ISP should also be capable of handling these other misbehaving nodes.

(Refer Slide Time: 20:01)

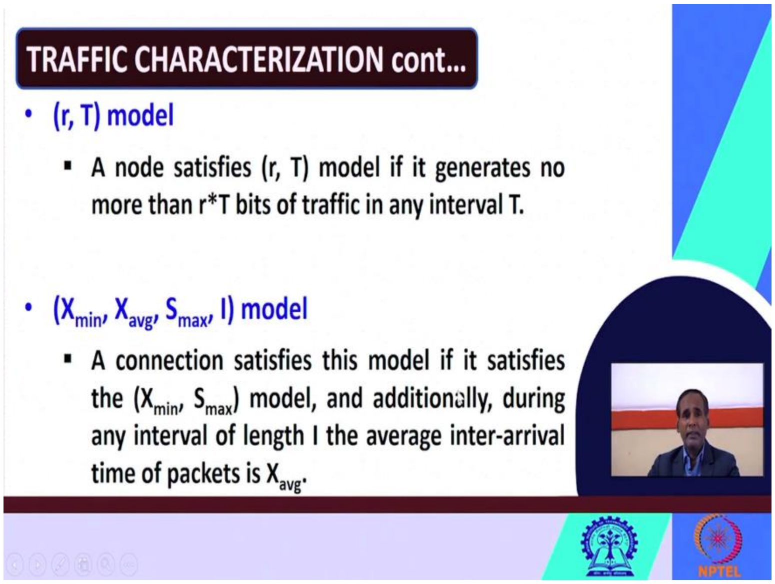


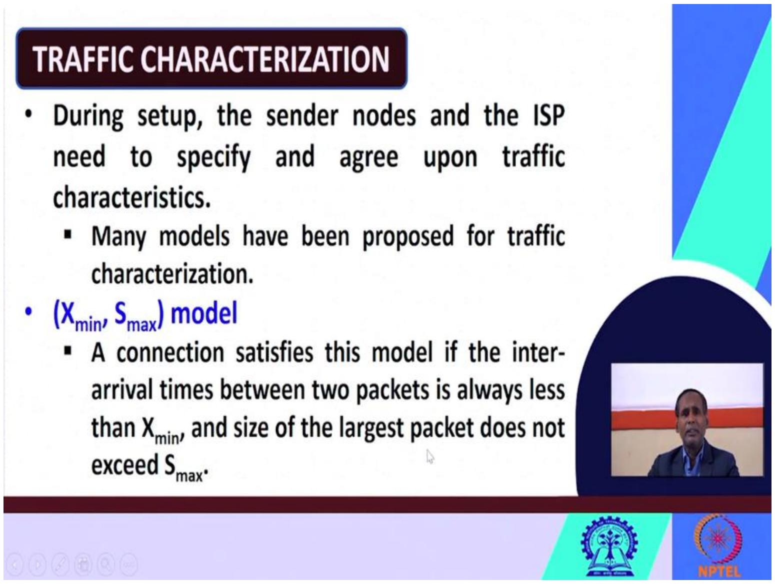
Now let us see a look at the traffic characterization. So, during setup for the sender nodes and the ISP they need to specify and agree upon certain traffic characteristics. Let us see some of the characteristics. During setup the sender nodes as well as the internet service provider they need to specify and then need to agree upon some topic at the stakes. Let us see what are those traffic characteristics we will discuss on these many models have been proposed for traffic characterization.

So, for many models have been proposed for traffic characterization. We will see some of the important models first model we will see X min S max model. So, according to this model a connection satisfies this model a network connection satisfies this model if the inter arrival times between two packets is always less than what the X min and the size of the largest packet it does not exceed S max plus a very simple.

We are discussing about the X min S max model. A model satisfies this model when if the inter arrival times between two packets is always less than what arrival time is less than X min and S is the size, the size of the largest packet it does not exceed this maximum value S max.

(Refer Slide Time: 21:27)



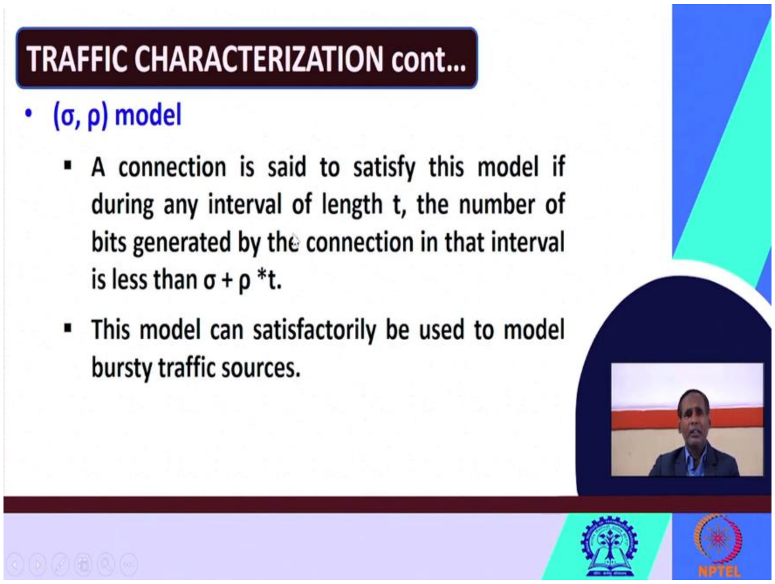


Then you will see the next model called as r T model, a node satisfies this model r T model if this generates no more than r into T number of bits of traffic in any interval t. So, any node will say that it satisfies r T model if it does not generate more than r into T number of bits of traffic during or any interval T. Next model is the X min, X average, S max and I model. A network connection satisfies this model if it is satisfying the following.

What are the followings? If it is satisfied the X min, S max model, I have already told you what is the X min S max model. Additionally, it has to satisfy some few things let us see what it has to satisfy. Additionally, during any interval of length I the average inter arrival time of packets is X average. I am again repeating in X min, X average, S max I model. So, two things you already we have seen X minimum and X average S maximum earlier model we have seen your X minimum as S maximum model.

So, a network connection satisfied this model first if it is satisfies the earlier model we have seen X minimum S max model and additionally during any interval of length I the average inter arrival time of the packets it is X average then we say that the network connection satisfies this model.

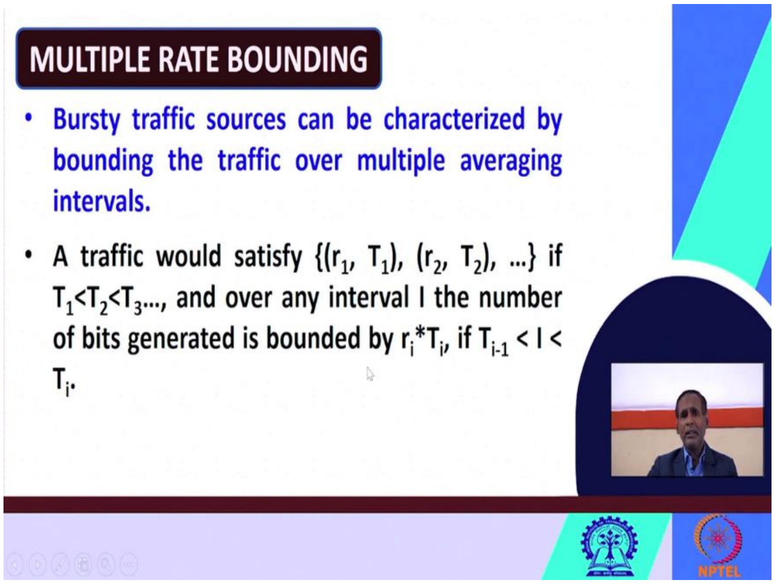
(Refer Slide Time: 23:02)

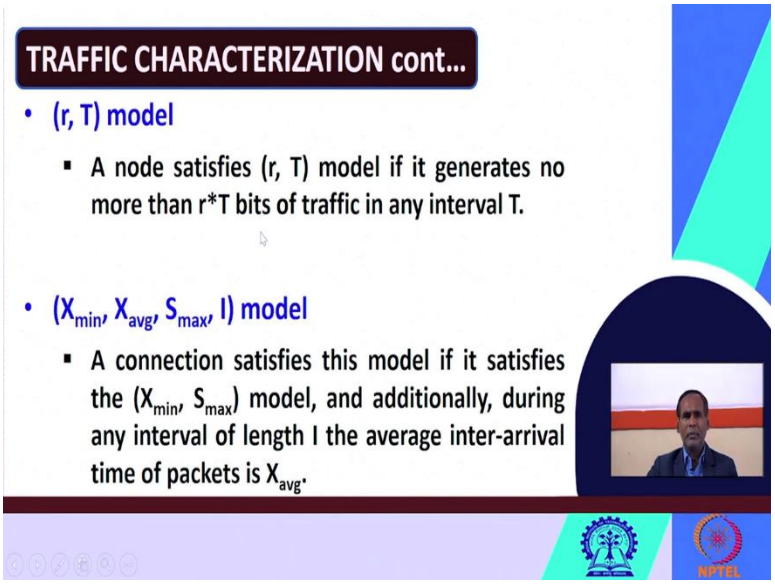


Another model we will see rho and what sigma model a connection is said to satisfy this model if during any interval of length t the number of bits generated by the connection in that interval t is less than the rho plus sigma into t. I am repeating again very simple a network connection is said to satisfy this model if during any interval of length t the number of bits generated by the connection in that interval t is less than this value is less than sigma plus rho into t.

So, this model can satisfactorily be used to model bursty traffic net sources, so you know bursty traffic sources. So, this model can satisfactorily be used to model the bursty traffic sources.

(Refer Slide Time: 23:56)





Now, let us see this multiple rates bounding see the bursty traffic sources can be characterized by bounding the traffic over multiple averaging intervals. The bursty traffic sources how they can be characterized? They can be characterized by bounding the traffic over multiple average intervals. Actually a relation has to be satisfy let us see what relation has to be satisfied a traffic would satisfy r1 T1, r2 T2 et cetera. If T1 is < T2 is <T3.

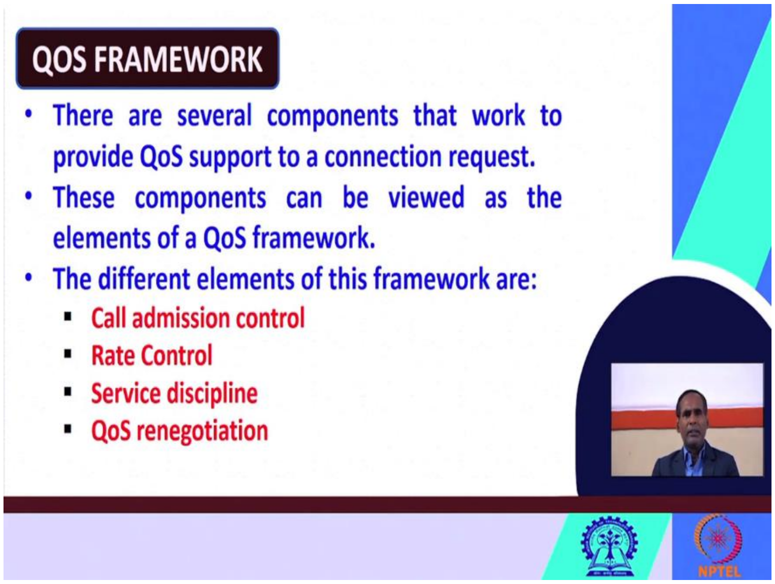
And over any interval I the number of bits generated it is bounded by ri into Ti. If i is lying between Ti - 1 and Ti. So, I think r T we have already seen. Yes, r T model I have already discussed here that a node satisfies r T model if it generates no more than how much bits no more than r into T bits of traffic in any interval T.

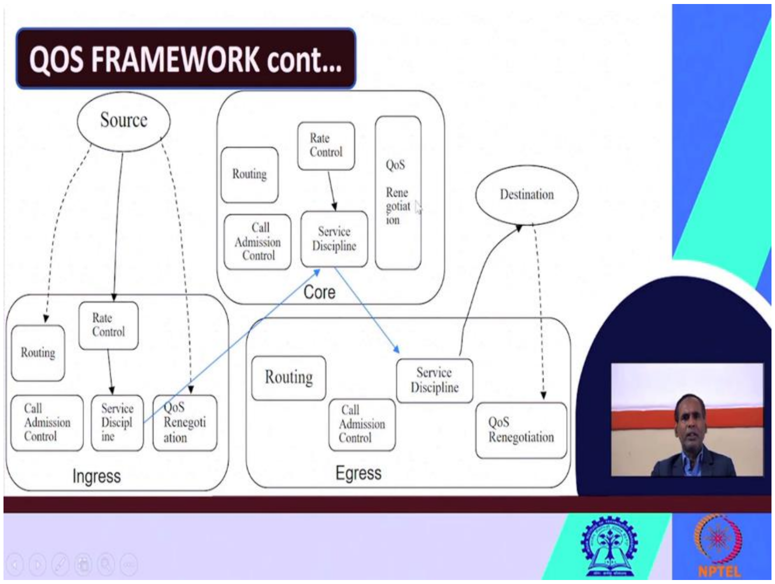
Now, when we are discussing about multiple rate bounding bursty traffic sources they can be characterized by bounding the traffic over multiple operating intervals. A traffic would satisfy this model may be r1 T1 r2 T2 et cetera. If T1 <T2 <T3 and over any interval I. The number of bits generated is bounded by what? ri * Ti. So there we have seen also that because this is a maybe for this is for what multiple rate bounding.

Here it is maybe a single a node satisfy r T model if it generates no more than r into T number of bits of traffic in the interval T. This model maybe we extended to this multiple rate bounding a traffic could satisfy this model maybe r1 T1 r2 T2 in this we are saying this is multiple this is multiple rate if what this T1 <T2 <T3 and over any interval I the number of bits generated is bounded by what is bounded by ri into Ti. If I is greater than Ti - 1 and it is less than Ti.

And what is T you already have known T is an interval T. So, here are multiple intervals we have taken T1 T2 T3 such that T1 is less than T2 and it is less than T3 and the traffic would satisfy this r1 T1 r2 T2 this multiple models if T1 <T2 less than T3 and over any interval i the total number of bits generated is bounded by ri * Ti. And if I >Ti - 1 and <Ti.

(Refer Slide Time: 26:49)





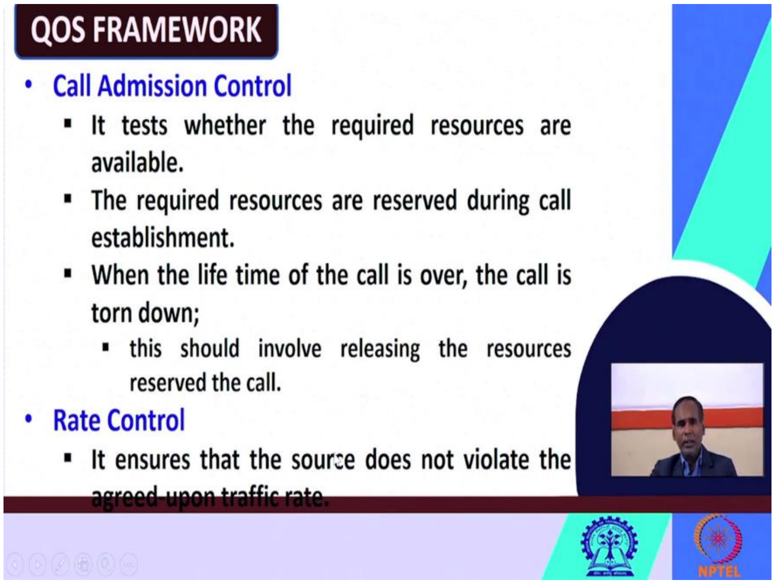
Now, let us quickly look what the QoS framework, there are several components which work to provide QoS support to a connection request. So, these components can be treated as the elements of QoS framework. You can see the QoS framework can look like this. So, this is the source and this is the destination you see the major components are what call admission control, rate control, service discipline, QoS renegotiation.

So, similarly also when you will go to the other end that is this is the ingress part. Because I have already told you source connected to ingress and destination is connected to egress. In egress also similar things routing, call admission control, service discipline, and QoS renegotiation. And in between source and destination there are several core routers are there.

In core routers also you will see this routing, call admission control, rate control, service discipline, and QoS renegotiation. So, let us broadly then classify. What are the components? So, the components that work to provide QoS support to a connection request are the followings. And those components can be treated as the elements of this framework. So, what are these components are elements.

From this figure we have seen that these are the different elements or components present in this framework. Those are call admission control, rate control, service discipline, and QoS renegotiation all those components are present in all types of routers. Now let us say a little bit the basics of this.

(Refer Slide Time: 28:30)

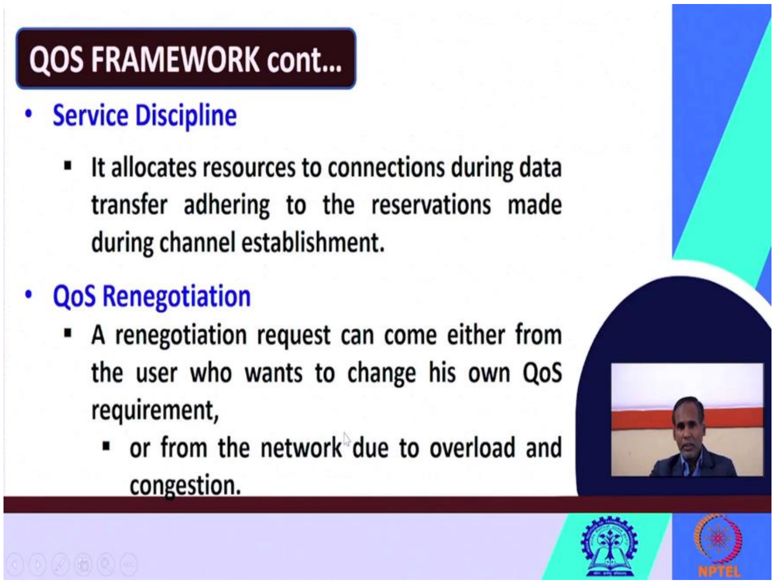


First, we will see call admission control. What does this component represent should this component call admission control it tests whether the required resources are available or not. So, this component call admission control. It tests whether the required resources are available or not, the required resources are reserved during call establishment. All the required resources they are reserved during on the call establishment.

When the lifetime of the call is over the call is torn down. So, whenever the lifetime of the call is over then the call is torn down. And this should involve the releasing the resources reserved the call. So, when on the lifetime of the call is over then the call is torn down. This should involve what? This should involve releasing the resources reserved the call that means whatever resources are reserved by the call now they must be released.

Then what is rate control? This is another component. So, this component ensures that the source does not violate the agreed upon traffic rate. So, what is the agreed upon traffic rate, what traffic rate is agreed. So, what is the traffic rate that is agreed upon maybe by both then that this component ensure that the source it does not violate this agreed upon traffic rate. So, rate control it ensures that the source it does not violate the what agreed upon traffic rate.

(Refer Slide Time: 29:59)

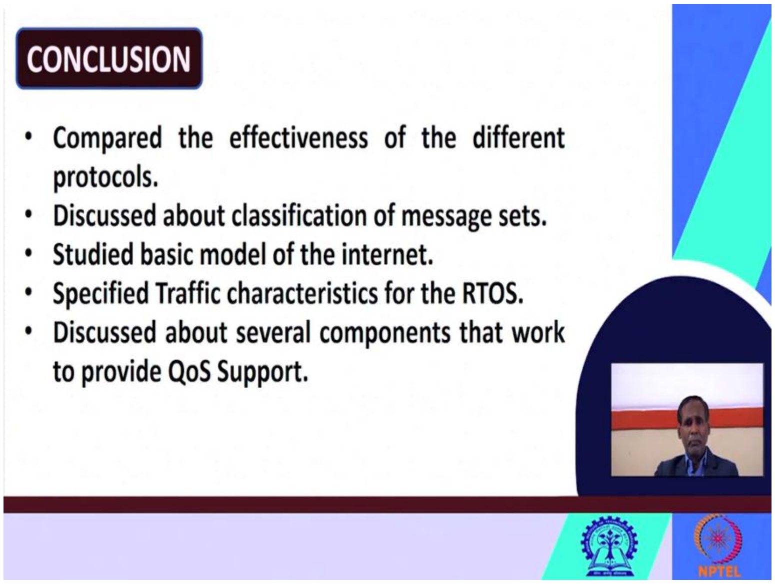


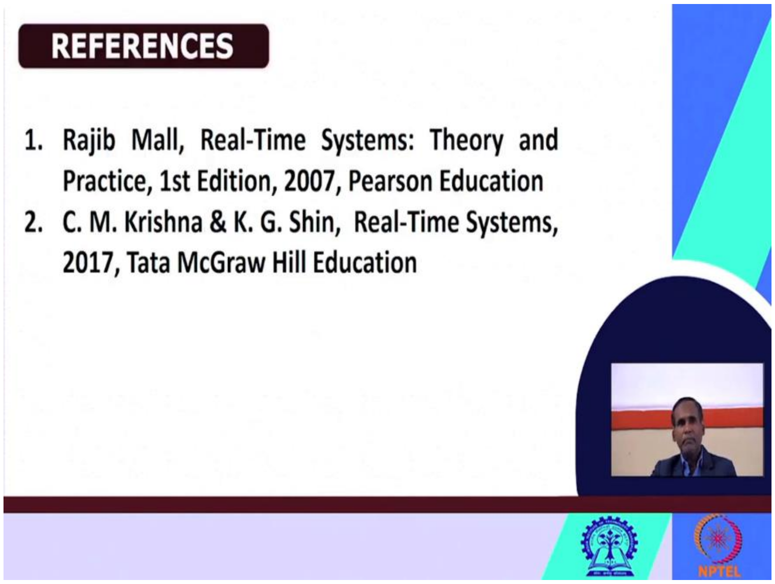
Then the next one is service discipline? What does this component do? This component allocates resources to the connections during data transfer adhering to the reservations made during channel establishment. So, as its name suggests service discipline, it allocates what? It allocates the resources, to whom? To the connections, when? During the data transfer adhering to the reservations which are made during the channel establishment.

The next component is the QoS negotiation, a renegotiation request can come it may come either from the user who wants to change his own QoS requirement or from a network due to overload and congestion. Let us see, we can make this renegotiation request a renegotiation request it make on either of the two parties that means either from the user who may want to change his own QoS requirement or it may come from the network due to overload and the congestion.

Due to what much overload and congestion the renegotiation requests may also come from the network. So, these are some of the components of the QoS framework.

(Refer Slide Time: 31:13)





So, today we have compared the effectiveness of the different protocols, we have discussed about the classification of message sets such as characterization and RTS et cetera or what you can say schedulable set unschedulable set and the saturated sets, unsaturated sets we have also studied the basic model of the internet, how the different routers they are connected or the peripheral we have seen this edge routers are connected and other routers are connected in between the core outdoors are connected in between them.

We have specified the different model traffic characteristics for that way several models we have seen in this regard. We have also discussed about several components which work together to provide the QoS support. So, we have taken from these two books. Thank you very much.