Digital Voice and Picture Communication

Prof. S. Sengupta

Department of Electronics and Communication Engineering

Indian Institute of Technology, Kharagpur

Lecture - 32

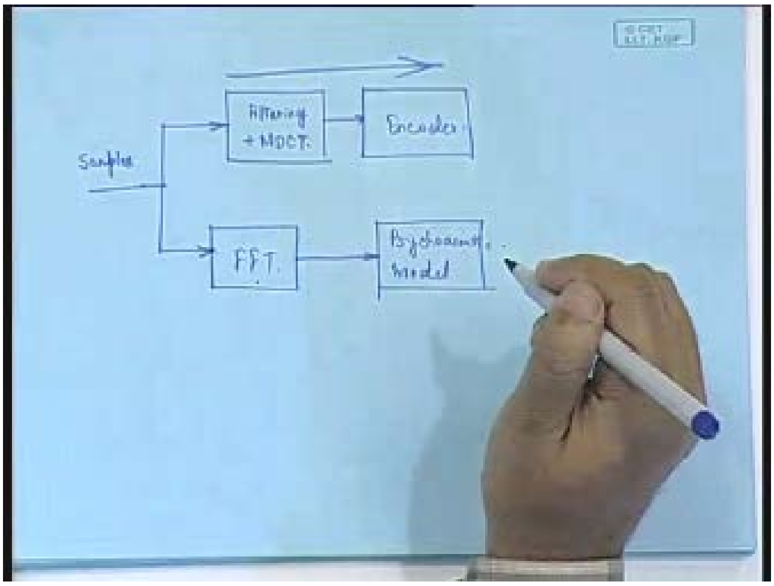
Introduction to VoIP

In the last lecture we were discussing about the MPEG-1 audio coding and especially we were taking up the polyphase filter implementation. Now, after the class I had some interesting discussion with a student where the question that was put is that whenever we are implementing the MDCT there we are essentially having equal spacing in the frequency sample because the way we derived it we are assuming that the center frequencies are pi by m apart and the center frequencies are located at the odd multiples of pi by 2m. So essentially all the spacing that we have the MDCT samples basically correspond to equal spacing in the frequency band range and the question that was asked to me is that in that case how can we relate this to the critical bands that has been talked of in our psychoacoustic model; how do we relate this to?

Well, one thing which you should notice is that this whole process of taking the MDCT and the encoding. So here we can say that whenever we are having the samples, this is where we feed the samples, this is actually going into two paths: in one of the path we are taking the FFT and in the parallel path we are having the polyphase filter implementation so this is filtering plus MDCT and then we are having the encoder where the samples are actually encoded and based on the psychoacoustic model.

So what happens is that psychoacoustic model is not incorporated in this path at all. This path is directly taking this way that all the MDCT samples are equally spaced in frequency whereas from psychoacoustic model it is not so. Actually there what we do is that we take the FFT and from the FFT we identify the tonal components and from the tonal components we find out that corresponding to the critical bands that are defined in the psychoacoustic model, for each of those critical bands what is going to be the masking characteristics so that we get a complete picture about the masking threshold that is present in the entire range of the audio processing that we are doing. Once the masking once the overall masking threshold is obtained the psychoacoustic model will then decide that what should be the bit allocation strategy to those samples.

(Refer Slide Time: 4:23)

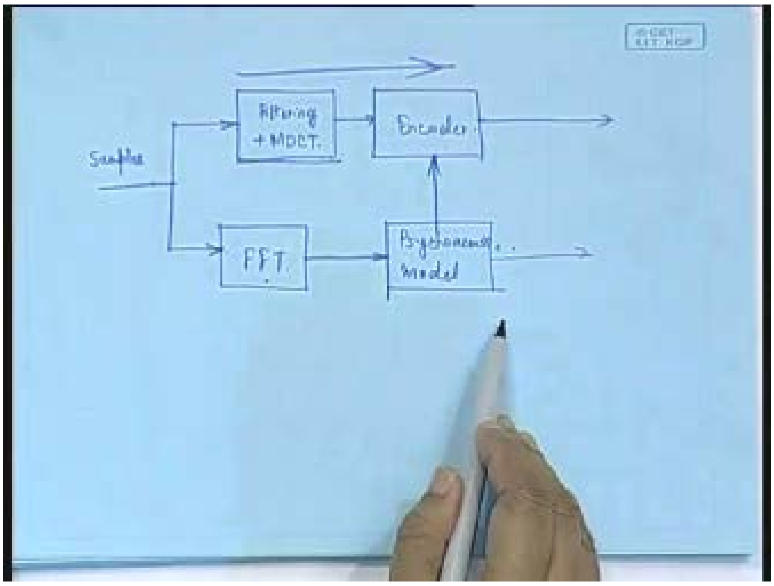


Now what we do, once the psychoacoustic model is derived from this FFT then we encode the MDCT samples; when we observe this MDCT samples we find out that to which critical band this MDCT samples belong and accordingly they are encoded or rather the bit allocation is being done accordingly. So the encoded samples will go out over here and not only that the psychoacoustic model or the parameters which we use for the psychoacoustic model or the bit allocation parameters they also will be put into the channel. This is the basic philosophy so do not think that this MDCT and doing this has got anything to do with the psychoacoustic model because the psychoacoustic model is parallely derived using the FFT. So that was the question which was asked to me and I thought that it is better to share this question with the entire audience because this doubt or this sort of a question can come to you come to your mind.

Before going into the new chapter let me ask that is there any other doubts that you would like to pose at this stage about the audio coding. If there is none then we will go over to the next topic which is the voice over internet protocol and very popularly it is referred to as V then the o is written with a small case o because this is voice over but internets I and protocols P are all capital; only for this over this o is taken as a small one so that it is capital V small o capital I capital P, so VoIP the voice over internet protocol. This will be our next topic of discussion.

Now, why we have chosen to take up this topic in our syllabus?
See, we have so far learnt about the different encoding techniques. We can now apply the encoding techniques on speech, we can apply the encoding techniques on audio, we can apply it on the images, we can apply it on the video and now we are well equipped to see that after applying this technique what is the transmission medium that we are going to consider.

(Refer Slide Time: 5:15)



Let us say for example, take the first aspect that is to say the speech encoding. When the speech is encoded........... we have learnt about different types of speech encoders. We have considered the PCM, the DPCM, ADPCM these techniques we considered; also we considered the linear predictive coding techniques which permits you to have good quality speech over a lower bandwidth application so all these things we did. Then where are we going to use that? The simplest place where we can use it is in our good old telephone system which is called as the plain old telephone systems; I said good old but they call it as Plain Old Telephone System so this is the POTS.

So we can apply the digital voice rather the, I mean, after the encoding of the speech signal we are encoding the speech samples digitally then this could be applied over the POTS. Now POTS has definitely got a bandwidth limitation. But yet POTS is the largest and POTS is very simple to use. But even POTS is also heading for a change. We are now having the encoded speech or encoded digital speech to be transmitted over the wireless medium. Then what happens is that when the speech coding techniques were developed and POTS was thought of as a medium at that time or maybe few years later people started thinking in terms of data transmission. So people thought of transmitting the data from one place to another and they had to think for some data networks and the internet happened to be the best medium for transmitting the data from one place to the other.

Now the thinking came that can we not use the internet medium for our telephonic conversation purpose. Because after all what is going to be the difference between a data network and a POTS. Because if POTS as long as it was analog telephone systems it is okay; but when we are using the digital samples when they are already the digital values the digital values............. I mean, what is the problem if we transmit those digital values to the data networks.

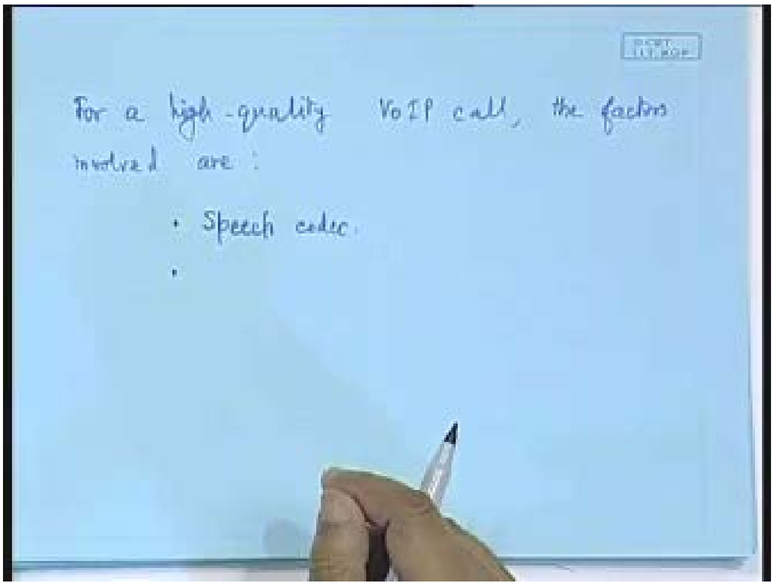
Well, you can argue that the data networks are going to be really congested if we use if for telephone purpose. Well, there we can try to make a judicious use that yes you may not be transmitting data all the time, your land bandwidth may be available, you may be having some spare capacity of your land bandwidth and in such kind of cases why do not you make proper use of your internet bandwidth for the telephonic conversation. This is what the thinking went on so that the internet medium could be effectively used for telephonic purpose and if this is a reality then definitely a very low cost....... I mean, one can make long distant calls at a very low cost because whatever price you are paying for your internet; especially if it is an Institute or University domain like ours where the students are getting free internet access the teachers are getting free internet access we can talk freely over long distance and we can communicate with anybody who is there in US or anybody who is in Europe, or any anybody who is in Australia we can freely communicate with them without incurring any STD or ISD expenditures so that looks really attractive.

Now, in order to have in order to make that into reality; in fact in some formed cities it is already a reality whether we are making wide usage or not is a different question we may not be able to use it very widely because the technology has to develop even further and right now what we are having is that there is a remarkable variation in the quality.

Now what you can do today is that you can have a voice mail chat and that I suppose that many of you might be already accustomed to do that with your friends and others, friends and relatives you must be having a voice mail chat where you are basically entering into a chat room and then you can make use of your voice in order to communicate. But most of the times you face such problems like discontinuities; may be that you are talking something and the other end is unable to follow because there is very long delays. So in order to have a very good quality conversation many factors are important. So, for a high quality VoIP call for a high quality voice over internet protocol call the factors which are involved are: first is that what quality of speech codec are you using. There are wide variety of speech codecs.

Some speech codecs are very low bandwidth speech codecs. Very low bandwidth speech codecs means where the compression factor will be very high and in fact such codecs will be quite costly to implement because, I mean, it will definitely involve a more complicated hardware realization for a very large bandwidth compression or you can alternatively use a lower bandwidth compression but have a better quality of speech output.

(Refer Slide Time: 14:52)



Then in this case, I mean, whenever we are transmitting the voice over internet the digitized samples what we have that has to be sent in the form of packets. Just like the way the data goes from one terminal to another; from the source to the destination the voice packets also travel in exactly the same way. So the scheme that utilize for the packetization that also dictates the quality of the VoIP.

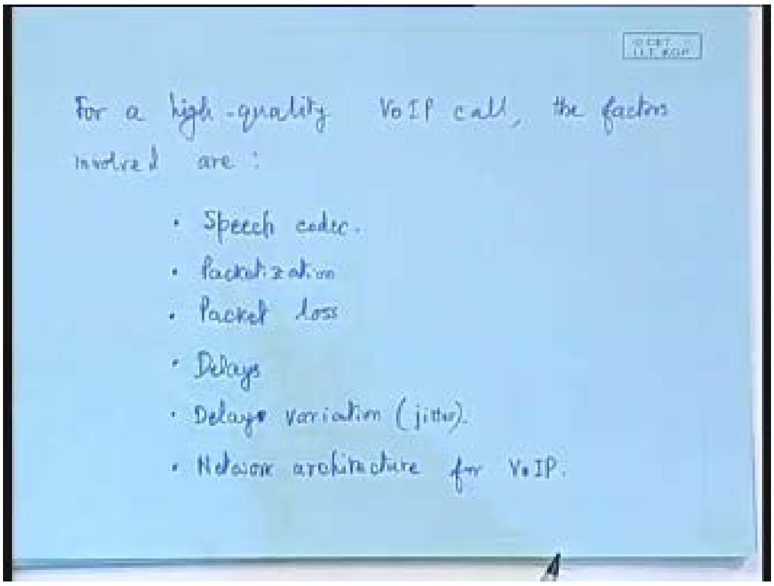
Then, whenever you are packetizing and transmitting you are not restricted to the circuit switched networks because in the case of circuit switched network what you have to do is to establish a connection, establish an end to end connection and then only you can transmit the data or in this case the voice as the data. But whenever you are using the data networks or the packet networks there it is connectionless so the packets can travel through the available roots and there is a possibility of packet loss.

So if your packet loss rate is high then you are definitely going to have discontinuities in the speech that you are listening. So, when you are listening to somebody, when you are talking to somebody who is there overseas you find some discontinuities in the voice. Invariably it happens that there is a good amount of packet loss that is taking place and when packet loss is there then that part of the speech you are missing.

Then the delays. this also is a very commonly observed phenomenon that you ultimately feel really disgusted while talking if you find that the delay is excessively high; because then there is no fun in making a conversation if the delay is inordinately high. And then apart from delays; delays mind you, fixed delays are easy to tackle because fixed delays means you are prepared for that but there are sometimes delay variations or what is called as the delay jitters in the networking terminology.

Delay variations or delay jitter variation also referred to as delay jitters and sometimes because of delay jitters what happens is that at one time the delay is less and at one time the delay is more and in order to have a constant delay what you have to do is to use a buffer so that you have to account for this delay jitter and you have to make use of the delay buffer and again the buffer should be so designed that you do not have an overflow of buffer. If you were having an overflow of the buffer then that again results in the packet loss.

(Refer Slide Time: 18:26)

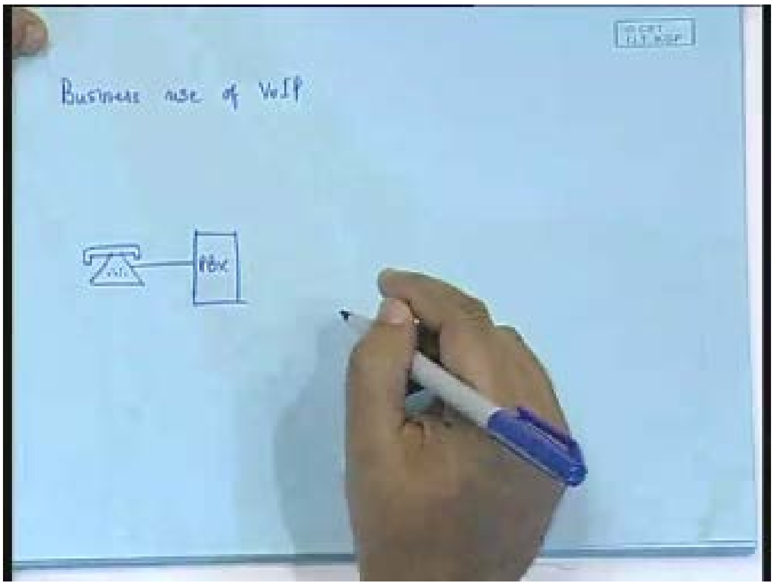


And then we have the network architecture to provide the VoIP. This also is a dictating factor. Now, before going into the technical aspects of the VoIP let us look at some very easy to understand configurations of the ways in which we can adopt a VoIP. You see, we have a choice; supposing a telephone set is available before us and supposing it is a digital telephone or need not be a digital telephone this could be an analog telephone also but we must have some facility. I mean, in the PBX to which we are connected that PBX should have a facility to digitize the voice and then once it is digitized once the digital samples are obtained there should be a flexibility that if the network is available if the network congestion is not too high you can root the voice packets through the IP network.

Alternatively when you find that the network is busy you can make use of the PSTN the POTS in order to make use of the in order to transmit or in order to establish the connection in the usual way. So let us see that how to have both that means to say voice over internet as also using the public switched telephone network.

Let us see that some business use of VoIP some business use of VoIP where both PSTN as well as the voice over IP is used. Say we have here a PBX and here we have a telephone set.

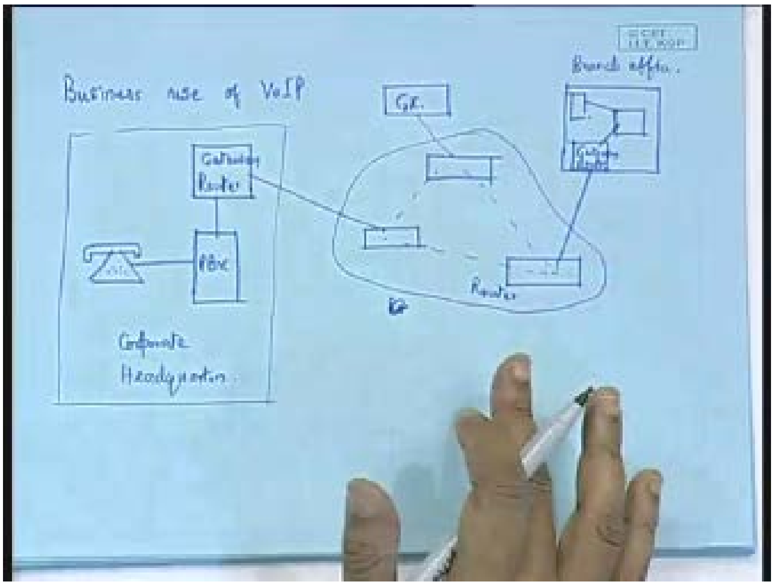
(Refer Slide Time: 20:51)



So this is a telephone set that is available and it is connected to the PBX and this is (Refer Slide Time: 20:59) supposing the corporate headquarters, this is the corporate headquarters in a business; now there must be a way to connect this voice into the internet so what we essentially have to provide is that this PBX should be connected to a VoIP router or what we can call as the gateway router. So this is a gateway router so the gateway router will basically connect the PBX to the network. So these are the systems that we are going to have at the corporate headquarters and then from this gateway router we have a router arrangement so basically we can have several such routers which are connected together.

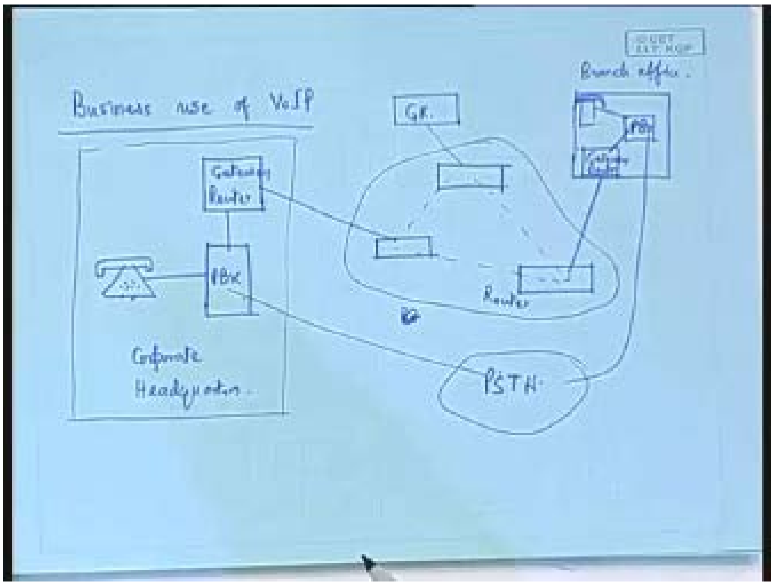
So you can route the packets from the source to destination making use of these routers and there is a gatekeeper; in short form we write as the GK which is the gatekeeper and supposing in one of the destinations; supposing we have a branch office over here (Refer Slide Time: 22:26) supposing this is a branch office where we will be having another gateway router so there will be a gateway router so this router will be linked to these routers so this is a network of routers so this is a so this is the router network which is there in the............. this could be in the wide area network but again in the branch office we must have a router and we must have a PBX, we must have the telephone and also there could be a gateway which is connected to the central office and to that the PBX could be connected. So like this we can realize.

(Refer Slide Time: 23:25)



Now what happens is that whenever we want to make a call from this telephone to this telephone (Refer Slide Time: 23:31); supposing there is another telephone over here then either we go through this data network and then we send the voice packets from the source to the destination and then from the destination we again get back the voice data packets and playback over here, now when it is not possible to route or when there is a network congestion then this PBX can communicate to this PBX through the PSTN. So PSTN also parallelly exists so that you can link this PBX through PSTNs all this PBXs could be link through PSTNs and then the call could be established through the PBX way.

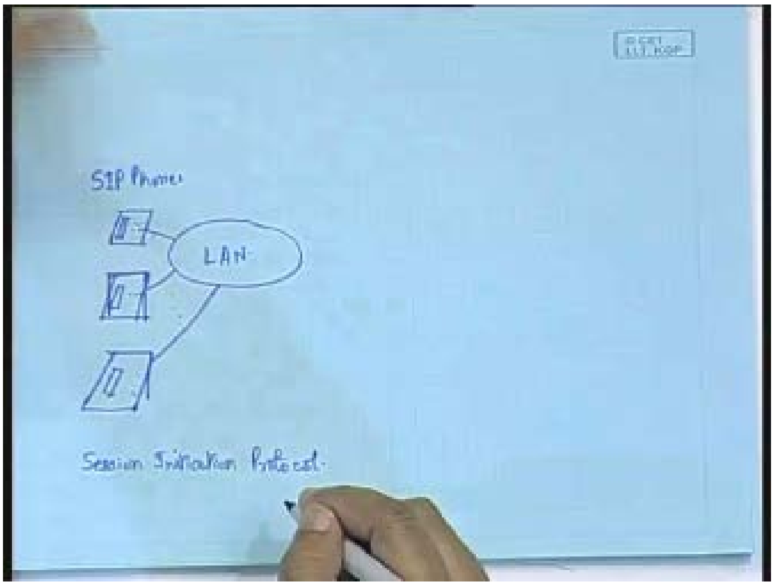
(Refer Slide Time: 24:30)



So here this is one configuration where you are making use of the IP as well as the PSTN based upon the congestion and based on the requirements you can choose either of the two routes. The other alternative way is that instead of connecting the telephones to the PBX why do not you connect the telephones to the LAN itself, those telephones are referred to as IP telephones.

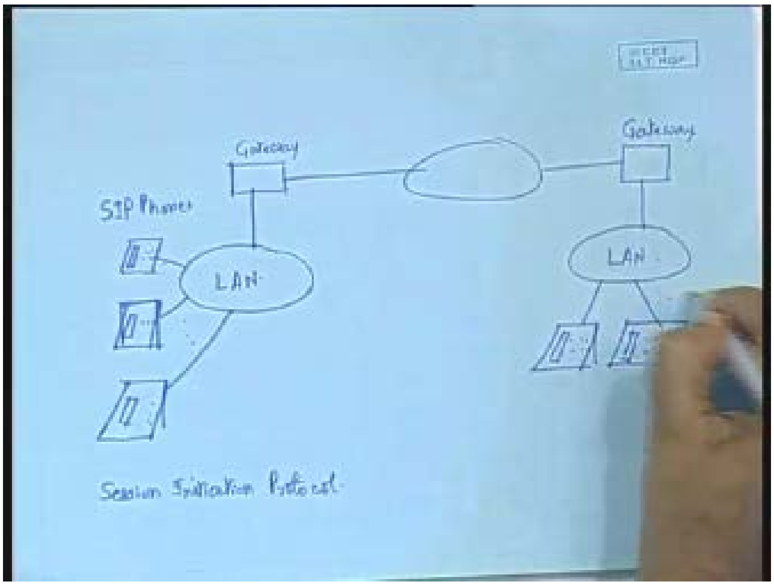
You can have a local area network to which you connect all the IP phones. This is a one such IP phone where there is a set over here then we have.......... it should be drawn like this then...... like this we can keep a large number of such IP phones which are connected to the LAN and then these are actually referred to as the SIP telephones SIP phones and the full form of SIP is the Session Initiation Protocol or in very common language it is referred to as the IP phones. So the IP phones are connected directly to the LANs so definitely these IP phones have to be digital phones.

(Refer Slide Time: 26:17)



So here these SIP phones will necessarily have the digitizing equipment and the digitized samples will be connected to the LAN, I mean, digitized as well as encoded, so the digitizer and encoder will be kept within this IP phones and then this LAN is connected to the gateway (Refer Slide Time: 26:53) and then this could get connected to the wide area network, then you can have a gateway over here, so this is gateway, this is another gateway and then you can have a LAN at this end and to this LAN you can connect the IP phones; another set of IP phones could be connected to this LAN.

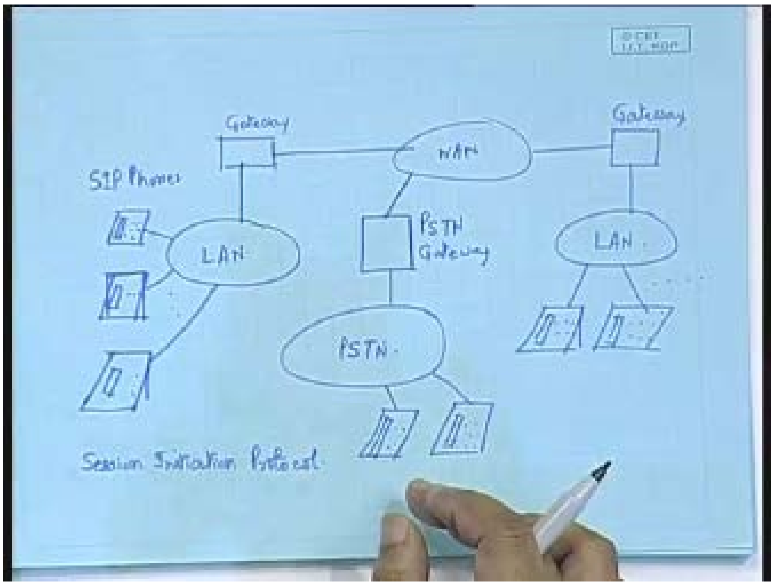
(Refer Slide Time: 27:29)



So now what one can do is that one can communicate between this end and this end. Any telephone from here you can either use it for the internal conversations because when it is connected to the LAN even all the within office conversations you can use this one and whenever you are communicating from one office to the other just simply you go through the network and you can communicate from here to any phone which is kept over here. But the difficulty is that how you get connected to the POTS because not all telephones will be connected as the IP telephones so there will be the normal telephones which will be connected to the PSTN.

So what one can have is that to this network to this wide area network one can have a separate gateway which we call as the PSTN gateway and to the PSTN gateway one can have the PSTN. Then the telephones which are connected to the PSTN........... so these are not the IP phones this could be the normal telephones which are connected the normal LAN lines which could be connected to the PSTN so that it is possible for you to communicate from this internet phones to any phones in this PSTN network; all that you have to do is to go through a PSTN gateway. So there are two ways of two different ways of implementing the voice over internet protocol: one is where you can connect the phones to the PBX and in this approach you are connecting the phones to the LAN.

(Refer Slide Time: 29:45)



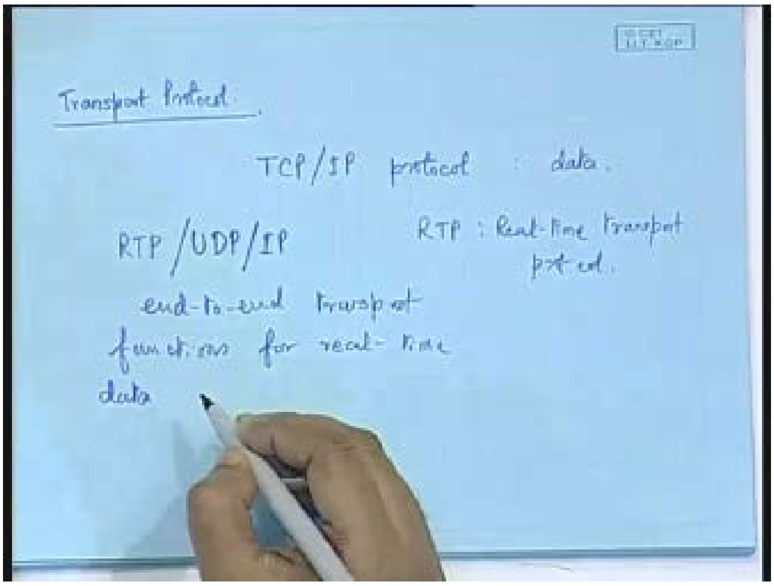
Now I would like to mention something about the transport protocol that we are going to use for this voice over internet protocol. Yes please? PBX is public branch exchange branch exchange yes branch exchange; PBX is the public branch exchange; that is the telephone. I mean, for the PSTN network all the exchanges all the telephone exchanges are connected to the PSTN network so we are referring to that; the public branch exchange; like in IIT Kharagpur you have the (Hej........30:36) branch exchange that is a PBX, IIT is also having a small PBX.

Now, coming to the transport protocol, now the typical internet applications which involves the data that uses the TCP/IP protocol to know that; TCP/IP protocol is used for the data and what is this TCP/IP protocol, the full form is transport control protocol and IP is the internet protocol. Now this is a connectionless protocol. So what happens is that the packets are sent from the source to the destination.

Now there is no guarantee that the packet that is sent from the source to the destination really reaches the destination, there could be congestion, there could be inordinate delays and delays may render the packet as non-delivered or there could be burst errors in the network because of which you may lose a packet so there is no guarantee that once you transmit a packet that packet will have a delivery. But if you evolve the system where you have a system of obtaining a delivery report and then transmit the next packet then that is what is being done using the TCP/IP protocol. So TCP/IP that sends a packet and then it waits for the acknowledgement and then every time there is a process of sending and acknowledgement, sending and acknowledgement like this. So that way it is a very reliable protocol in the sense that if you do not get the acknowledgement or it is a negative acknowledgement that is to say that a there is an error in the packet that was received in that case no acknowledgement or negative acknowledgement means that you have to retransmit that packet. So you retransmit so there is no question of packet loss; at the most what you have to do is to make certain retries but you can deliver the packet.

The whole problem is that for data it is absolutely fine. If we do this for transmitting one email message to another place that is well and fine because even if some packets are lost but it is sent through TCP/IP we know that the packet did not receive and we retransmit the packet but ultimately the packets arrive and the email message can be composed, it is okay. But can we do the same thing for the real time transmission, no. Because it is a continuous speech that we are going to transmit from the source to the destination. So we are having the speech continuously and then the digitized and encoded speech samples which are packetized that is being transmitted and then if that packet gets lost we cannot really wait for the acknowledgement and we cannot................ why we cannot wait for the acknowledgement is because it is a real time data which is coming so the next data has to be sent and even if we try to evolve a system of obtaining the acknowledgement and there is a negative acknowledgement or no acknowledgement and we want to resend there is no question of resending because we are first of all not able to store a large volume of data, data means the voice data and there is no question of retransmission for the real time voice applications. That is why we are unable to use the TCP/IP protocol for the VoIP and in this case what we have to use is to use the UDP. UDP’s full form is the Universal Datagram Protocol and we have to use the UDP/IP protocol for that. So what we do is that we use the UDP/IP in conjunction with the Real-time Transport Protocol RTP. So RTP is the Real-time Transport Protocol so when RTP is applied over UDP/IP then we can use it for the real-time data transmission. So RTP/UDP/IP that provides end-to-end transport functions for real-time data.

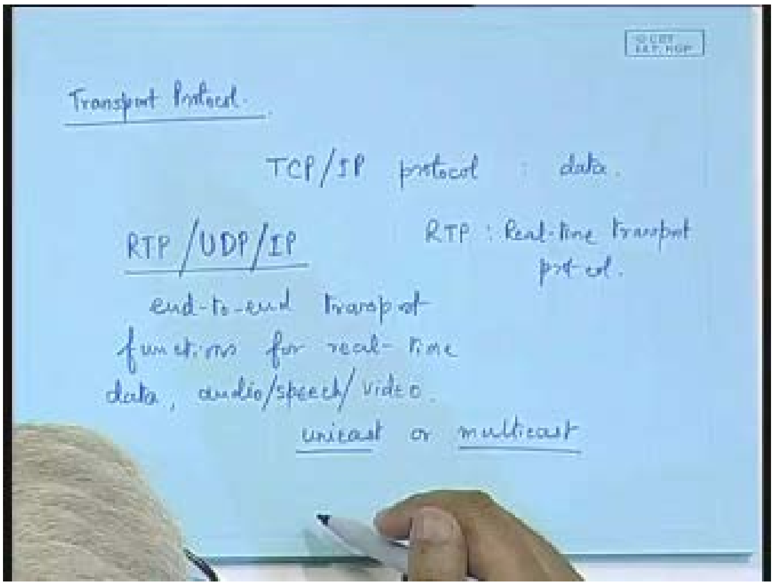
(Refer Slide Time: 36:14)



What are the forms of the real time data?

The audio, audio of course...... audio, speech, video so everything is real time and this RTP/UDP/IP this could be used for the unicast. Unicast means when it is from one point to another point, a one-to-one communication or multicast mode; multicast means when we are transmitting as one to many, when it is being used to many users then it is a multicast configuration. So for unicast or multicast this RDP this RTP/UDP/IP can be used.

(Refer Slide Time: 37:12)



Now there is a million dollar question about the amount of bandwidth which we are going to use for the voice over internet. Now, what should be the what should be our bandwidth requirements.

Well, today you can say that the bandwidth is getting available at cheaper and cheaper rates but it is not cheap everywhere. Especially whenever we are using any private networks or not in all parts of the world the bandwidth is available at a cheaper cost, we have to pay the price for bandwidth. When we have a limited bandwidth scenario; especially say whenever the part of the network is a wireless network; part of the entire network happens to be a wireless network, there we have to work under a bandwidth restricted consideration.

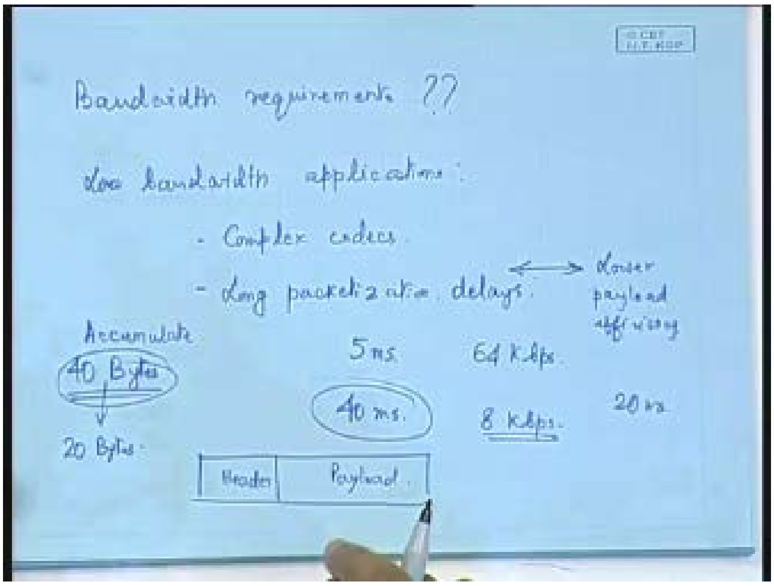
And when we have a limited bandwidth, there some questions that arises that what should be the........ I mean, for low bandwidth applications two factors come in: first of all low bandwidth applications of speech, two factors coming first is that we have to use a we have to use complex codecs because as I was telling you sometime back that the quality of the compression factor that we are going to achieve out of the codec that compression factor will be very high for the low bandwidth application and for low bandwidth application also we have to encounter very long packetization delays. This will be evident from this example that supposing we have let's say supposing we have to accumulate 40 bytes supposing we want to accumulate 40 bytes in a buffer and there are two choices: one is that we can use a 64 kbps bit rate channel or alternatively we can use an 8 kbps channel.

Now, whenever we are using 64 kbps 64 kbps means that it is 8 kilobytes per second so that means to say that for 40 bytes of data we require 5 milliseconds of accumulation time whereas when the bit rate is reduced to 8 kilobits per second there we have to accumulate 40 bytes in 40 milliseconds of time. So the buffer accumulation time is higher and that means to say that your transmission delays also, I mean, your processing delay gets increased because the buffer accumulation time that gets accumulated to the overall transmission delay that you are making.

So reducing the bit rate definitely results in longer delays, long packetization delays that is quite evident and how do you correct for this? There you have to use........... in order to reduce the packetization delay at lower bit rate or low bandwidth application what you have to do is to reduce the size of the packet and if you are reducing the size of the packet or you are reducing what you are reducing the payload that is present in the packet then it leads to inefficiency because anyway for a packet transmission there are two parts; you have the packet header and the packet header is followed by the payload.

Now if I use a payload of 40 bytes and then I incur 40 milliseconds of delay; delay in packetization at a rate of 8 kbps but still if I have to use 8 kbps and somebody tells me that more than 20 milliseconds of packetization delay will not be allowed in that case what choice is there with me; I have to restrict my packet size; that is, instead of 40 bytes I must reduce it to 20 bytes. But may be that I have four bytes of header information so when I had 4 bytes of header information for 40 bytes of payload data so there 40 bytes out of 44 so it is nearly 90 percent of the packet was composed of the payload and only ten percent was the header whereas now if I have 4 bytes of headers and 20 bytes of payload then 20 out of 24 so the payload proportion has come down. So definitely this is an inefficient arrangement but what we have to do is to make a tradeoff that if we want to avoid long packetization delays then the choice for you is to go in for lower payload efficiency. So, either you have to go in for long packetization delay or you have to go in for lower payload efficiency.

(Refer Slide Time: 43:38)



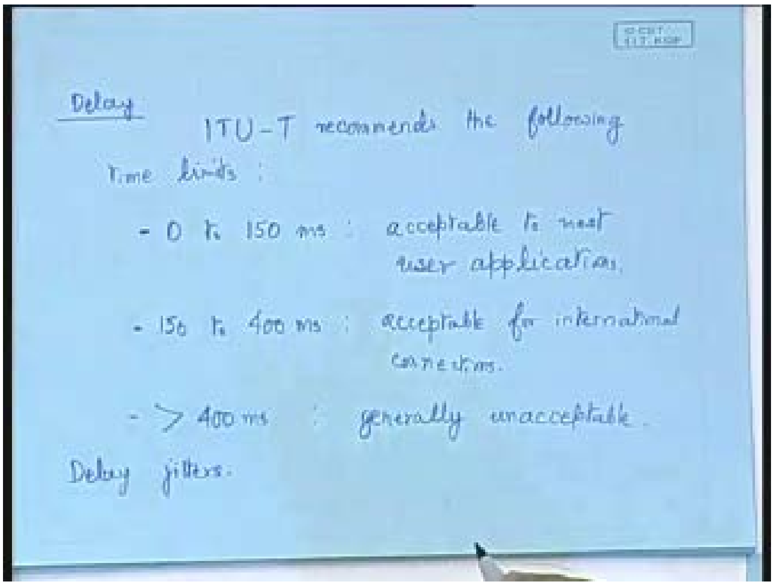
So there you have to go in for judicious engineering tradeoff between the complex codec, long packetization delays and your payload efficiency and in fact there is a relationship that evolves that if the call transmission capacity requirement if the call transmission capacity requirement is given by BW, BW in bits per second then BW can be expressed as the codec output rate R this is because of the payload so this is the codec output rate is the R and this should be R plus H by S where H is the header size, H is the header size in bits and S is the payload sample size and this sample size is measured in time units means sample time rather so this is payload sample time (Refer Slide Time: 45:09) and if we measure this in milliseconds in that case it is bits per milliseconds so effectively it is kilobits per seconds units.

So, if you are expressing the call transmission capacity in terms of the kilobits per second and you use your codec output rate also will be expressed in kilobits per second then this relationship is there, so R is because of the payload and H by S is because of the header that you are obtaining so this is the relationship. And in order to improve the payload efficiency, now for the RTP/UDP/IP for RTP/UDP/IP the header size is 40 bytes so it uses a 40 byte header but header can be compressed to 2 to 7 bytes. So 40 byte header may be compressed to 2 to 7 bytes. But mind you, whenever we are doing a header compression where that header compression has to be done, it is not that if you do the header compression once in the source and then you can forget about the rest of things up to the destination because it has to be done at every router that when it receives the packet it has to decompress, it has to know the packet header and then again from this router to the next one it has to once again compress the header and send it. So the header compression has to be done at every stage. but at least the compression of the header definitely improves your efficiency in the sense that if you are compressing the header from as much as 40 bytes to 2 to 7 bytes typically you can say that 4 bytes then it is a compression factor of nearly 10 is to 1 that you can have and that can improve your efficiency of transmitting the payload.

There is yet another factor that is really important for the voice over IP communication and that is the delay. And delay, we are mostly bothered about the end-to-end delay and the ITU-T the same body which we mentioned sometimes back the International Telecommunication Union they in fact have proposed all the standards pertaining to this and we will be soon coming to the aspects of the standards which are involved in the voice over IP protocol especially the call initiation protocols etc which is given by the H dot 323 that part we will be talking in the next lecture. But ITU-T basically recommends the following time limits. This is formed based on the opinions of several observers. People were asked to evaluate the performance based on a and based on a mean opinions score where the observers were asked to evaluate the performance on a scale of 1 to 5 they were asked to give numbers in the range of 1 to 5 that whether you like the presentation or whether you do not like the presentation at all, so based on the user feedback ITU-T concludes that when the delay is between 0 and 150 milliseconds that is acceptable for most of the users' application.

For delays in the range of 150 to 400 milliseconds there is a problem but there is no choice we have to accept this for the international connections because the international connections the route delays itself are so high, the transmission delay itself is so high that 150 to 400 milliseconds is acceptable but only for international connections and greater than 400 milliseconds this is generally unacceptable. And also they talk about the delay jitters.

(Refer Slide Time: 51:00)



Delay jitter means that where there are delay variations. if it is let us say the network generates a constant delay of 100 milliseconds then there is no problem because then all that we are talking will reach the destination 100 milliseconds later. But if it so that one packet takes 100 milliseconds of time and the next packet takes 70 milliseconds of time and the next packet takes 140 milliseconds of time in that case it results in what is known as the delay jitter and in order to account for the delay jitter we cannot have it this way that some packets are arriving faster some packets are arriving slower so what we do is that we keep a buffer at the receiver and that buffer is referred to as the delay jitter buffer and the delay jitter buffer has to be of such a capacity so that up to the maximum extent of delay jitters the packets can be accumulated and only if it exceeds the allowable delay jitters limit then the packets get dropped or rather it results in packet loss.

Now we will discuss about the more detailed aspects of the voice over internet protocol in the coming lecture; till then thank you.