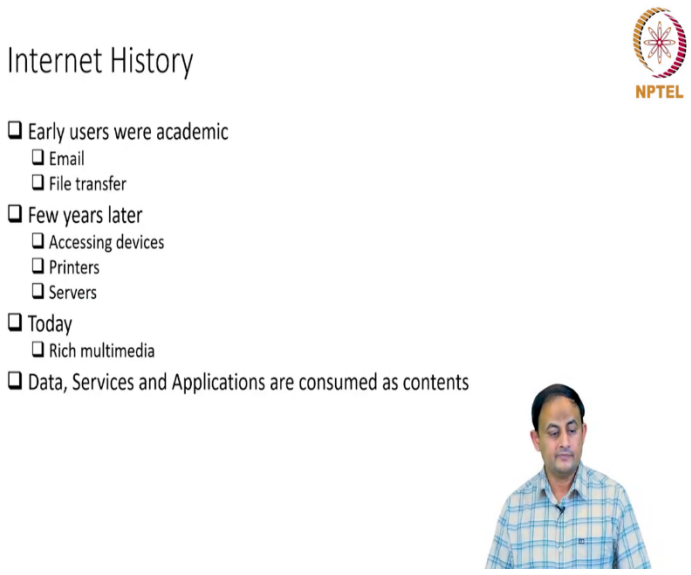
Advanced Computer Networks
Professor Doctor Neminath Hubballi
Department of Computer Science Engineering
Indian Institute of Technology, Indore
Lecture 61
Content Distribution in IP Networks Part I

Welcome back. Today, we will start with the fourth model of this course which is Information Centric Networking. So, what we will do in today's lecture is to basically understand the motivation for the Information-Centric Networking, and why it is required. So, we will spend some time understanding, how Content Distribution actually happens in today's IP network and what are the issues, and then finally why Information-Centric Networking is required and then how it works, probably in the subsequent lectures.

(Refer Slide Time: 0:54)



So, in order to learn that let us begin looking at the history of the internet; and how it evolved over a period of time. So, the early usages of the Computer Networks or the Internet were very limited. So, as I said only two applications were primarily there back then in the early days of the Computer Networks, which is the email and file transfer, people wanted to send emails from one person to another person and then the files which were put in one computer were supposed to be accessed by someone else using the other computer, these were the two primary applications.

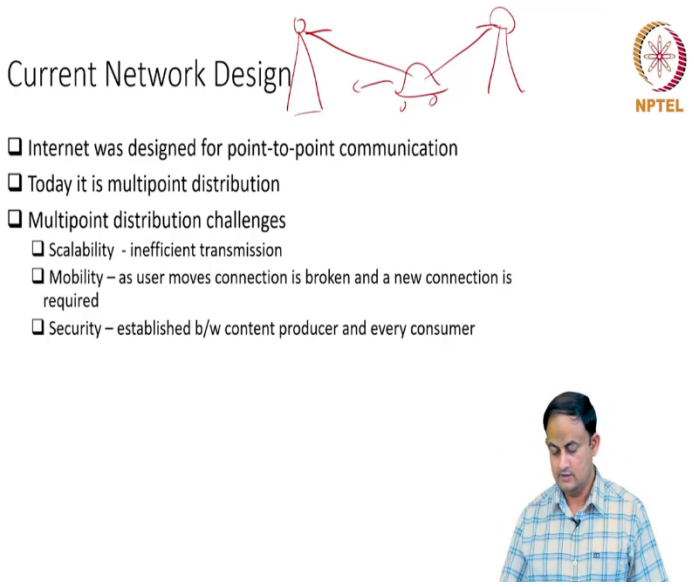
And keeping those applications in mind, their usage in mind, the designers came up with certain design that would work well, and it served well for quite a good amount of the time. And as the days passed by, we came up with new applications, we started talking about remote access, and remote login, several other devices like printers were connected to the network, and then we had remote access for printing the stuff from a remote computer and stuff like that came into picture the file system and like that.

But today, as we speak much of the internet, of a network is actually transmitting multimedia content. So, we do have applications like YouTube and several OTT platforms which are actually rich in multimedia content with audio and video content, which actually constitute a major chunk of the total data that gets transmitted on today's internet. So, this media content back then, was not actually the use case; the internet was not designed for delivering such multimedia content.

So, what exactly is the issue with the multimedia content, why it is actually I mean an issue? So, we will try to learn that. So, in a nutshell, what we see is a network today, there are several data sources available on the Internet, or the network, there are several services that are provided through the web and the network, and there are applications which actually generate a lot of the data.

So, from an end user's perspective whether it is the data service or the application, all of them are consumed as the content. So, there is nothing like this is specifically for data there is this specific for the service, everything is integrated. So, that is what the network of today's IP Network looks like. So, how exactly this pose a challenge for designers and distribution, that is the question.

(Refer Slide Time: 03:57)



So, internet was, as keeping those two applications in mind, it primarily internet or the network computer network is designed for point-to-point communication. So, if X and Y want to engage in the conversation, they want to transfer some data. So, be it the bitwise data, be the packet data, be it the file data, it does not matter what is the data, you know who are the two endpoints which are actually engaged in the communication.

So, if X gives, the data, the network knows how to deliver it to Y. So, whether the same data is repeatedly transmitted, if let us say the same file is downloaded 100 times by my peer, then 100 times the network blindly transmits that data from location X to Y. So, this is actually kind of network does not understand what exactly going on in the network. So, that is the point-to-point communication.

So, if let us say the website is there may be a News website or like timesofindia.com or something like that and whether I access, you access as many users are there those many unique users are there those many times the same content is actually getting transmitted over the network to multiple people. But when you come to the multi-point distribution kind of the environment, the scenario changes, when I say multi-point distribution, what I mean by that is the same content if multiple consumers are consuming the content, if let us say I am downloading the timesofindia.com website and my neighbour is also downloading or reading the same newspaper and the server is somewhere let us say in Mumbai and the same content is

getting downloaded two times to my computer and also to his computer. So, there is a lot of redundant work that is going on. Is there a way we can actually minimize this transmission?

So, that is the question, and the process of downloading the content multiple times is called as the multipoint distribution. So, the content distribution is an issue. In the multi-point distribution there are many challenges. So, in the current IP network, one of them is the scalability. So, as I said because the internet or the network works using the point-to-point communication.

So, even if I want to do multipoint distribution as well, I still use the same semantics of point-to-point communication. So, if n number of users are there, or consumers are there, n number of the times the same content is getting repeatedly transmitted over the network that becomes very inefficient. So, as the number of the users increases, then it poses scalability challenges.

If let us say that today thousand users are interested in the same content, and tomorrow if one million people are interested in the same content, then transmitting one million times the same content over the network is is not only challenging but also it is wastage of lot of Network bandwidth. So, that way it is actually inefficient and the second point is the users are mobile, I might be downloading the content on my mobile phone while I walk or when I am driving somebody sitting in my car might be downloading some content.

So, when you are either walking or driving you are actually connected to one of the mobile base stations. So, and that has got the antenna, and the users are somewhere here, and as I move from one coverage, I might be connected to this base station or the mobile tower and as I drive, I am in the vicinity of some other Mobile station and as I approach this coverage, there is something that happens called as the Handover. So, the user is actually connected to the next available mobile tower which is nearest which will give you better signal strength.

So, when this Mobility happens, when the user moves from one coverage area, one base station to another base station, the existing connectivity to the server breaks down. Let us see I am watching a video, I am connected to the YouTube server, and as I drive. So, I move from one base station to another one. So, now, the previous connection break. So, when I see the connection, this is typically the TCP/IP connection that I am referring to. So, the TCP connection

breaks down, and then you need to re-establish a new TCP connection to the same server and then continue to download the content.

So, in a way, what is exactly the ongoing transmission through one channel, one set of the routers intermediate routers is actually broken, and the new one is actually established, this is actually a challenge. And the third point is the current security mechanism. So, most of the time when we browse the internet or even watch YouTube, you might see this https version or the bar on the screen, so that stands for the secured.

(Refer Slide Time: 09:36)

The security is actually given to, if let us say, this is the computer X and another computer Y and the TLS, the way security works in today's network is through TLS or the SSL connection. So, it establishes a kind of tunnel between two entities or two computers which want to engage in the conversation.

So, it says that I am going to give you this notion of this tunnel, whatever data you put on this channel would be delivered to the other end securely, that is what the tunnel is. So, that does not mean that data is actually not read by or not analysed by the intermediate routers, that is a different story, the content itself is actually secure. I am going to give you this notion, so you put the content on one end and it will exactly come out at the other end that is what the TCP actual security tells you, or https or TLS for that matter.

So, in that sense, the notion of the security that is existing today is not for the content itself. So, we use a mechanism which is actually giving you an illusion of the security, a secure connection to the other party, and then on top of that we are actually putting the plain content. So, that is what the current security mechanism looks like.

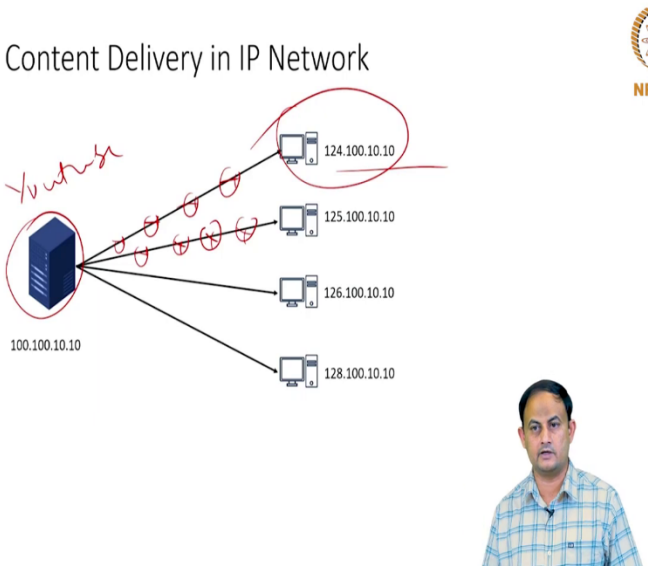
So, now an obvious question that comes to mind is, is there any alternative way of looking at security? So, for example, when I transmit one chunk of the data, can I put the security for that chunk itself? So, that is a different notion which is actually not there in today's Network. So, this point-to-point communication and today's usage of the multi-point distribution and the way security is provided to the content itself, is actually kind of a very inefficient way of operating or distributing the content.

So, the more popular the content becomes many many people access the same content because I am transmitting many many times for individual users over the network, it becomes less efficient. So, if only one user is using it, then probably it is as it may not be that inefficient, but more the popular the content becomes the larger the number of users accessing it becomes less efficient for the distribution.

So, but today consumption, multimedia applications are operating in this format multipoint distribution. So, that is the challenge. So, with that background in mind, let us try to understand what are the mechanisms that we have built. So, because these applications are operating in this fashion and the TCP/IP network has been designed with certain things in mind. So, some of the things we have invented or we came up with in order to minimize this overhead of transmission. So, to make it a little bit efficient.

There are a number of techniques that are retrofitted into the TCP/IP model specifically to deal with the content distribution problem bring some sort of efficiency for the content distribution. So, let us go over them one by one and try to understand what are those mechanisms and then we will be using that as a background let us motivate ourselves for what the Information Centric Networking actually brings on to the table.

(Refer Slide Time: 13:31)

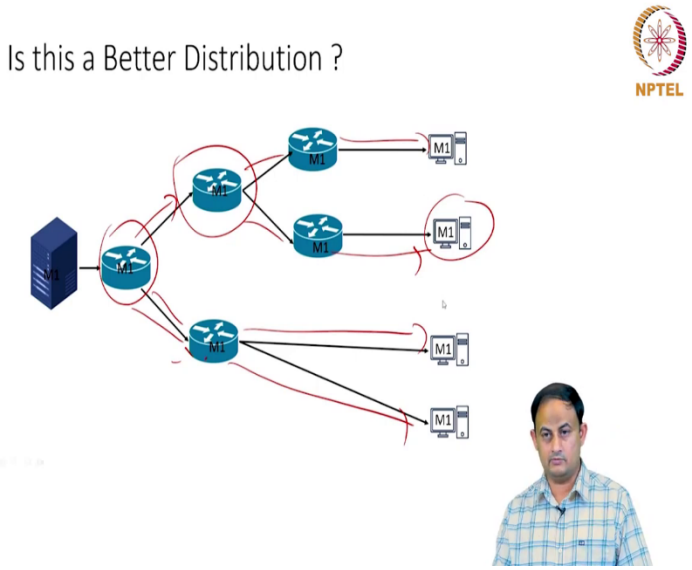


So, here is a picture that actually shows on the left-hand side we have a server and a bunch of computers on the right-hand side, and all these computers have got an IP address assigned to them. And let us assume this is the server which is hosting some content, maybe in YouTube server or any other popular web server or any which is actually having some content. Upon the other side, you have a bunch of consumers, and they are accessing the content.

So, every time this guy accesses, the content actually flows through a bunch of intermediate routers, and it actually comes and gets delivered to this point. And every time the second guy accesses again, it is close to the packet that actually passes through again intermediate routers and then gets delivered like this the content distribution is actually happening.

So, this is when this guy accesses the content, and when this server is actually responding to that, this is a point-to-point communication. Only these two endpoints are now engaged in the conversation, and the intermediate routers' job is to just facilitate the transmission of the packet corresponding to that content over the network over the multiple hops which actally gets delivered to the right user. So, we can see that as the number of the users actually increases then the number of times the same content is transmitted will also increase as the number of the users increases. So, that is the inefficiency.

(Refer Slide Time: 15:08)



So, this is when this guy accesses the content and when this server is actually responding to that, this is a point-to-point communication. Only these two endpoints are now engaged in the conversation, and the intermediate router's job is to just facilitate the transmission of the packet corresponding to that content over the multiple hops, which actually gets delivered to the right user. So, we can see that as the number of users actually increases then the number of the times the same content is transmitted will also increase as the number of users increase. So, that is the inefficiency.

So, now what would be the better approach to do this distribution? So, assuming the same content, same movie, or same video file is being accessed by or requested by all the users in the previous slide now, what is the better approach that we want to bring on to the table? So, let us say the content itself let us name it as movie number 1, that is, M1, and if let us say M1 gets transmitted to this router, the first top router, one time, the content is now available at the router router R1 the content itself is M1, and maybe this same content goes to this router, now the content is available at this router, and let us say all of these guys on the right-hand side, four of them requires the same content M1. So, all that I can do is go to this router and ask if can you give me M1, and he gives. So, this router can also ask this guy, can you give me M1 and he gives.

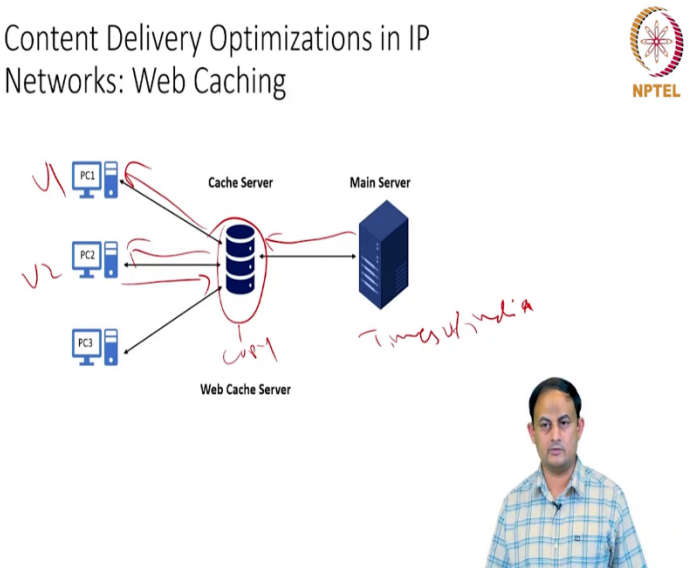
So, finally, M1 comes from this router to this end user, and again the same M1 comes from this router to this guy, and this guy can access one time and then one time to this guy and one time to

this guy, and this is where the content can be this how it can be delivered. So, the end server actually gives it movie 1, movie M1 only one time, and these routers actually intelligently coordinate among themselves and the content movie they made available for all the end users.

So, effectively every link in this one only one time the movie got transmitted from one hop to another hop in simple terms. So, this is actually a very good way of transmission. So, earlier, if four users were requesting, four times the server was giving, but now the server give only one time, and some of the routers requested only one time, and they were able to in turn serve to the many other routers in the network, and they may also serve many of the end users in the network. So, this is actually a very very efficient way of doing the distribution.

Now, the question is how do I achieve this, or in the IP network, is it possible to do it? So, in a true sense, it is not possible to do it, but we have found out several intermediate ways of doing it. So, we are not doing the exact same thing as shown in this figure but something which is along these lines, a compromise between how many times I may not necessarily do only one-time transmission but something which is in the ground. Now, let us take a look what are the mechanisms that are available to go somewhere here, but we are not there.

(Refer Slide Time: 18:11)



So, many of us have heard this term called as the Caching. So, Web Caching is one mechanism. So, I was referring to the timesofindia.com example. So, let us say this is the server of the Times of India News website, and there are some consumers on the left-hand side. What I do is maybe

let us take the example in organization. So, maybe IIT Indore is there, and there are a bunch of users inside the IIT Indore Network who are interested in visiting the timesofindia.com website. And this Caching server can once the user number 1 using the PC number 1 in this figure actually makes a request, the content is actually delivered to that user in the process; this Caching Server also keeps a copy of a local copy of the same web page and then when the user U2 from another PC makes a request the content instead of going all the way to the Times of India server the cached content can be served are given to the end user.

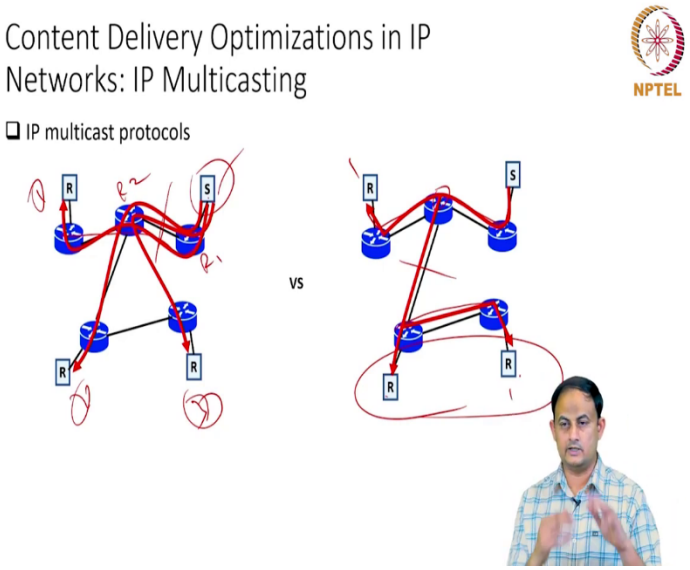
So, many times we might also notice when we actually visit websites which are popular like google.com, the page gets downloaded quite quickly. So, this is primarily as the page itself is cached when I say the page, it is actually the search page, the blank space with the search bar that page, the content itself is actually not Cached in the server if you are using the Google the page.

On the other hand, in this case, the timesofindia.com, the content might also be stored because the same content is actually accessible, particularly when the user U1 and U2 are parallelly accessing, and along the timeline they are actually close by. So, within a fraction when user 1 accesses, the content is delivered, and within a fraction of sometime the, user U2 also makes the same request, the content can be delivered, because by that time, in a few seconds or few minutes of time, the contents on the timesofindia.com web server may not have changed. So, it serves as a good; it makes sense to Cache the server and then send.

But that is one way of doing it, and this obviously comes with some negatives what is the U2 if the content on the main server changes, then if you are keeping local cached content, then the updated content may not be served; it goes with that assumption. So, here, in this case, even if you cache the content, if I serve the cached content to the end users, the utility of that news.

So, all the page the entire page of the timesofindia.com website may not have changed only parts of that might have changed, or one link or one image might have changed, but it is still is useful to user 2, this is one way of doing it. So, obviously, there is a limitation on how much you can cache, and because as the amount of the data that you want increases, the cache increases then the storage requirement will also increase. So, that goes with that negative impact.

(Refer Slide Time: 21:42)



So, that one way, the second is you also built some protocols, we have also made the routers available aware of such a distribution is happening. So, for example, on the left-hand side. So, there is one source which is actually generating the content, maybe a video streaming content, maybe a video conferencing application is running, and the presentation is made from the source, S, and these are the users number 1, 2 and 3 are the users who are interested in that communication. If you use plain point-to-point communication, every transmission originating from S, needs to be transmitted three times, one for this user using this link and the second time for this user, third time for this user.

So, from this point onward till wherever the routers are common on the path, there are three times in this case, the three parallel transmissions are going to the same router from, let us say, this router number as 1 and this router as R2 between R1 and R2, three parallel transmissions of the same content destined for three different users is passing through. Now, this actually, again does not make sense. So, three times the same packets carrying the same content are transmitted from R1 to R2.

So, in order to deal with this, in order to reduce this transmission burden, we designed something called as the multicasting route protocols where we actually take the end users and then tell the nearest router I am going to join so and so multicast group. So, add me to that group, and let us see this guy come and he joins. So, there is one transmission already going on from S, to they will say R1, and now this guy comes you go nearest point, and then you join the same

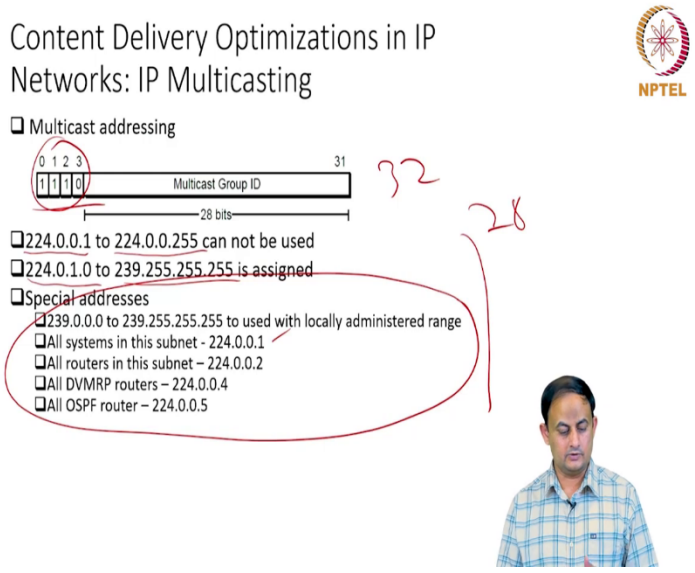
transmission, and when this guy asks, I also want to join then you find out who is the nearest guy who can give you that content.

So, it is this router turns out to be the nearest point where you can get that content, you join that group. So, the routers themselves are organized in some kind of tree, kind of architecture where they make themselves aware that so and so content is being transmitted through me and it belongs to so and so group, and now whoever wants to join that group, I am going to give that content to that particular router or the end user that is what it says.

So, we have built a protocol like group management protocols where the routers are aware, and the dynamic event users can dynamically request for certain content to be part of a certain group or not in the group, and then the routers will manage when the user is active, when the user is not active, and is asking for transmission, when he is not asking for the transmission, this is called as the multicasting.

There are a bunch of routers, and only a subset of them are actually interested in receiving the content; only to those subset of the routers, you transmit the content that is being originated at the S. So, this is called another way of reducing the burden on this one. So, currently, this is how the picture looks like if these two guys walk out from the group, then this transmission will be broken. So, only from S, the user number 1 or R1, the content is actually being transmitted.

(Refer Slide Time: 25:38)



So, there are a bunch of IP addresses that are reserved, which are basically used for this particular purpose. So, this is called IP multicasting; class D address or group D address is actually reserved for this particular purpose. And the class D address, specifically whoever is part of that multicasting group, they need to use this particular series of IP addresses, it starts with a triple 1 followed by 0, that is, an indication that the multicasting is actually being used.

And it is actually triple 1 followed by 0, and then out of the 32 bits, 4 bits are taken the remaining 28 bits actually can take any value that is technically what it means these 4 bits are having the fixed values of triple 1 followed by 0 and the rest of the 28 bits can actually vary. But not all of them can be used for multicasting. You can assign it to the end-users; there are certain IP address series which are actually reserved for some specific purpose.

So, from 224.0.0.1 to 224.0.0.2, it cannot be used, and another series from, 224.0.1.0 to 239.255.255.255 this actually can be used for the multicasting purpose, and within that series as well, some of the IP addresses or particularly these IP addresses are actually reserved for a specific purpose. So, if you use the IP address 224.0.0.1, this is actually used to mean that you want to communicate only to all those users within this particular subnet.

And the second 224.0.0.2 will also mean that we actually are transmitting; the communication should go to all the users which are within the subnet, but yet interested in the receiving the multi-cast content. And there are a couple of other IP address series also reserved for some other purpose, particularly for the routing application, they are actually reserved. So, that is a way of doing the optimization.