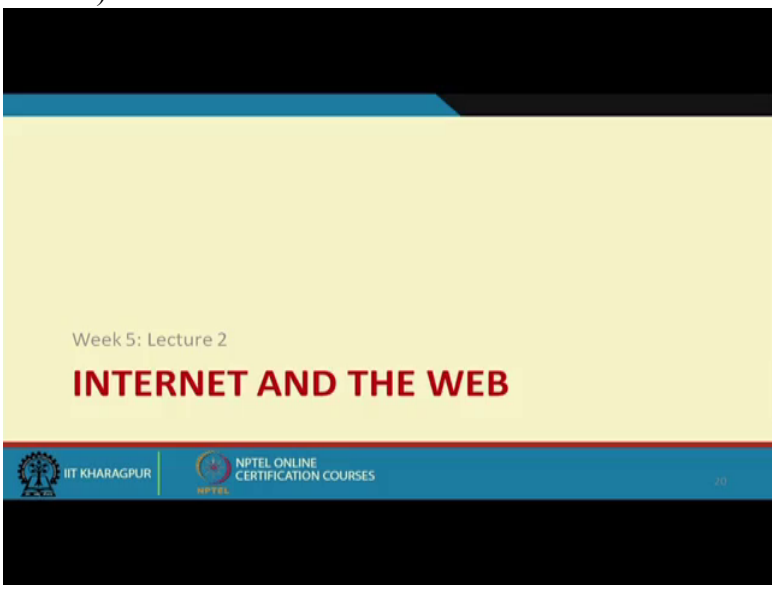
Course on E-Business
Professor Mamata Jenamani
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur
Module 05
Lecture Number 24
Internet and The Web

(Refer Slide Time 00:19)

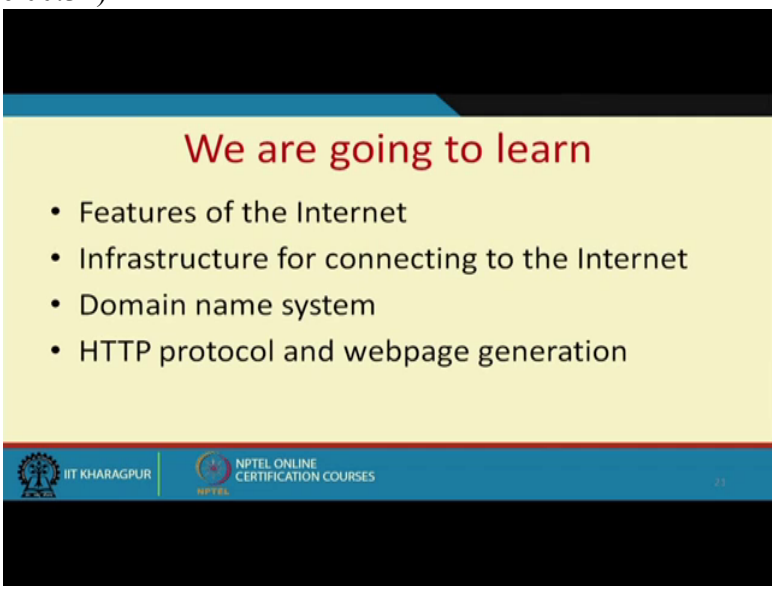
Welcome back. Since last lecture we have started about the infrastructure for E- Business. And we will continue with the same area,

(Refer Slide Time 00:29)



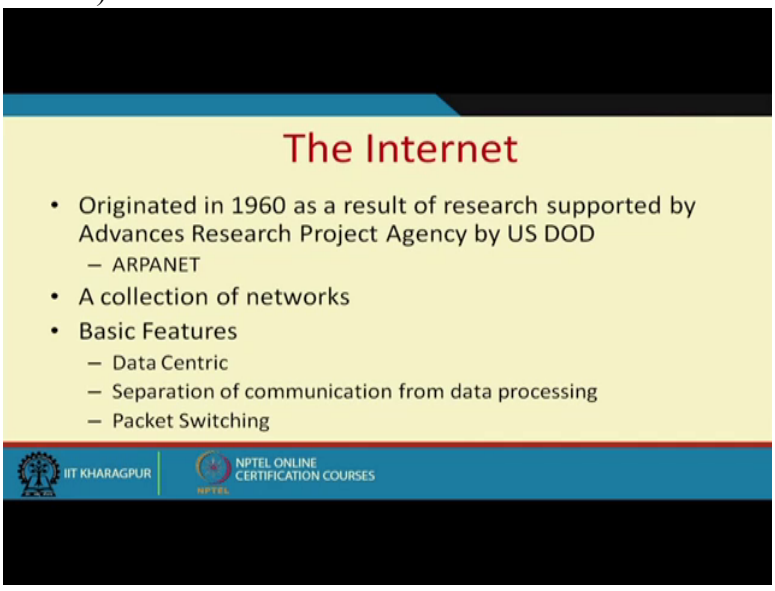
and we will start today. We will start talking about internet and the web.

(Refer Slide Time 00:34)



So in this lecture we are going to learn what are the features of the internet then we have to see what is the infrastructure behind this E- Business infrastructure and how to connect to this internet. We are going to talk about Domain Name System, H T T P protocol and how web pages are generated.

(Refer Slide Time 00:56)



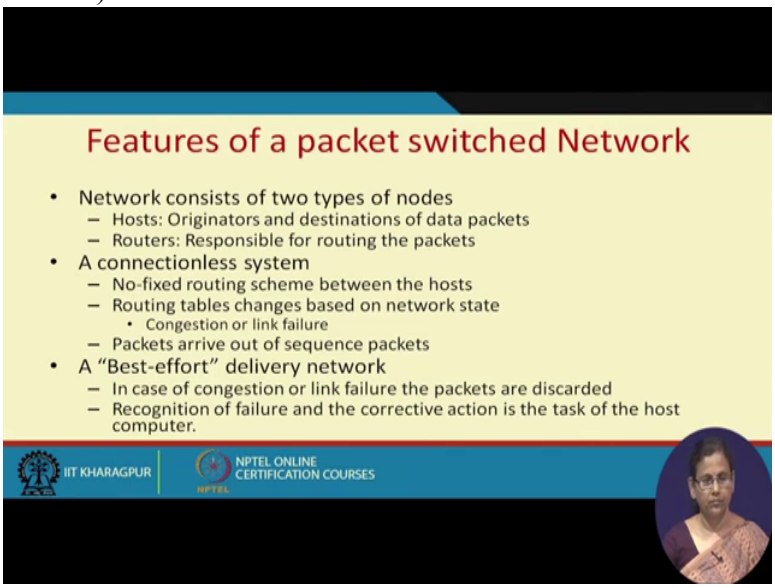
Now this internet is, though it became commercialized in 1990, it was originated in 1960 in one of the projects by U S Department of Defense. The project was ARPANet.

This is basically, nobody owns the internet. It is just a collection of networks. And it has 3 basic features. First of all, it is datacentric. Only data can be sent over it. Of course the idea has now changed. But even if we send voices and all, all of them are actually treated as data

packets and they are sent over the internet. Then second is this internet is designed to separate communication from data processing. So data processing is not a feature of the internet.

Next thing is it is packet switching. Let us try to understand what is the idea of a packet switched network.

(Refer Slide Time 02:04)



Here the network consists of two types of nodes. Some are called hosts and some are called routers. The hosts are the two end systems. For example you are sending data from your web server, your E- Business web server to the customer. Then customer's system is one of the hosts. Your server is another host. So this is the originator and destination of data packets. Then we have another kind of nodes which are called routers. These routers are simply responsible for moving the data packets in a specific direction depending on the address attached to it.

This is a connectionless system. By connectionless we mean, whenever we send data packets they need not follow the same route. They are broken into small, small packets and they are, they need not be sent on the same route. There is no fixed, which means; there is no fixed routing scheme between the hosts. Then the routing, then the connection is maintained using the routing table which is changed based on the network state. And if there is any congestion over any link then it can change. Now these packets arrive, because it connectionless and the packets are sent through whatever path is available and less congested; the packets can arrive in out of sequence.

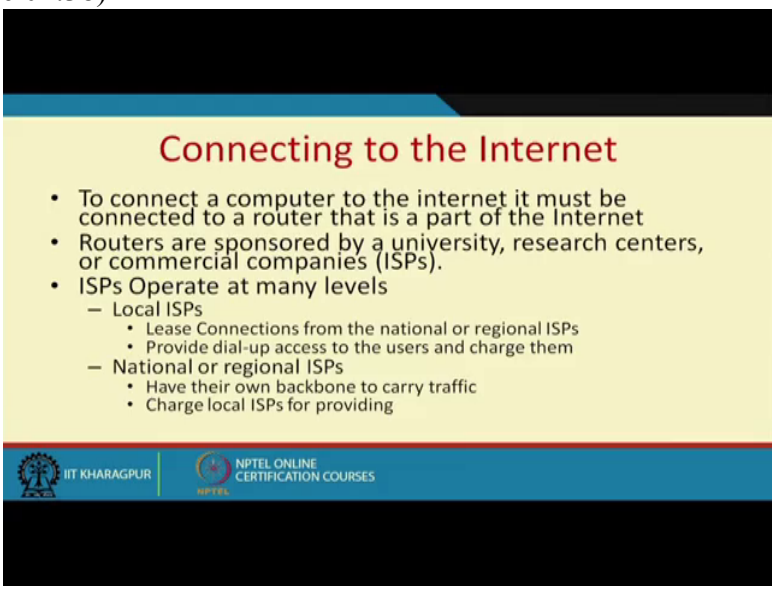
Then the next property of this packet switch network is, which is this case internet is, it is a Best effort delivery network, which means in case of congestion or link failure there is, the internet does not take any responsibility if the packets are lost. The packets are simply discarded if there is any kind of problem in between.

Now to recognize, it is the work of the host computers to recognize that there is, there are some packets missing and it has

(Refer Slide Time 04:34)

to resend it.

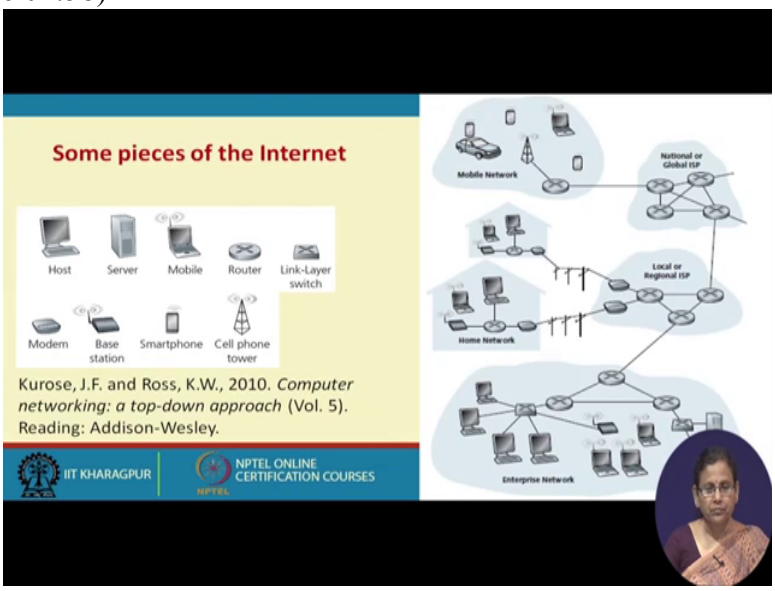
(Refer Slide Time 04:38)



Now how to connect to the internet is not to be discussed, because all of us, at least whosoever is watching this video, all of us know how to connect to the internet. But we need

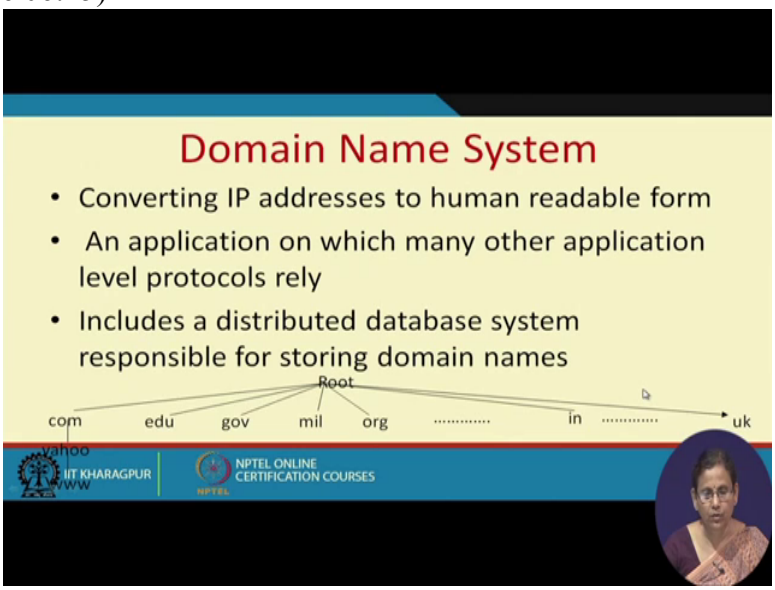
to, in order to connect to the internet; we need some local I S P who in turn connected to the regional I S P.

(Refer Slide Time 04:58)



In fact, as an organization if you would like to connect to the internet and your users can be anybody and they can and they can use any, any kind of computing device, it can be a mobile, it can be a p c; it can be anything. So this particular diagram shows how exactly you can, your organization can connect to its customer over the internet. So there will be, I mean, your enterprise network there will be some server and that server in turn will be connected through some proxy, last class we were discussing, it will be connected, probably it might be going through some proxy then there will be some router. Router will be diverting the traffic in a specific region depending on the I P. Now this, in the, within this local I S Ps or in the national level I S P or in the global I S P, these packets through the router are routed in a specific direction and through a number of cell phone towers they will be reaching your home computer. Your home computer can be your mobile, your server or anything. So your home computer is the client and you are connecting to the server.

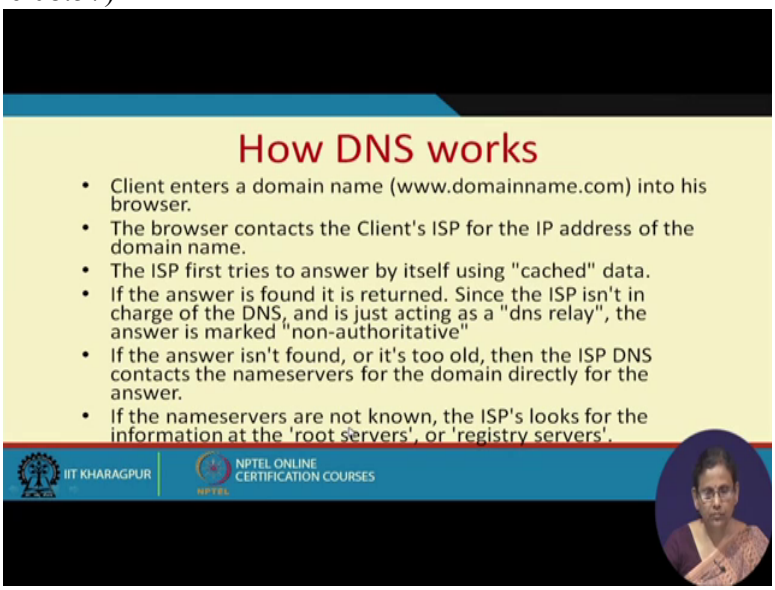
(Refer Slide Time 06:25)



Now when somebody connects to your business system he will be writing some kind of web address. So this web address, but as I have told you already, every computer in the network has certain I P address and that is how it is uniquely identified over the network. So therefore this name has to be, the name of the website now has to be converted to that I P address. Now the question is who does this conversion. So this converting this I P address to human readable form is done by some application called domain name system. So it is an application on which many other application level protocols rely. It includes a distributed database system responsible for storing the domain names. By distributed database system we mean a replica of at least a part of database system exists in many places in the internet. For example it exists in your own machine, it exists in your proxy's memory and so on so that repeatedly you do not have to go to the root server to find out this conversion between I P address to the domain name. Domain name is your web site address.

Now this domain name system is arranged in a hierarchical manner. There is a root server and with that root server is attached many specific domain servers. For example dot com is a server, dot edu is another server, dot gov is another server, dot o r g is another server, dot in is another server and so on. Again in turn in these specific domain servers, you, there will be this I P address to, I P address to your web site address conversion and replica of this is going to exist in many places over the internet. So that every time the packets need not have to consult this specific domain or root server to divert them.

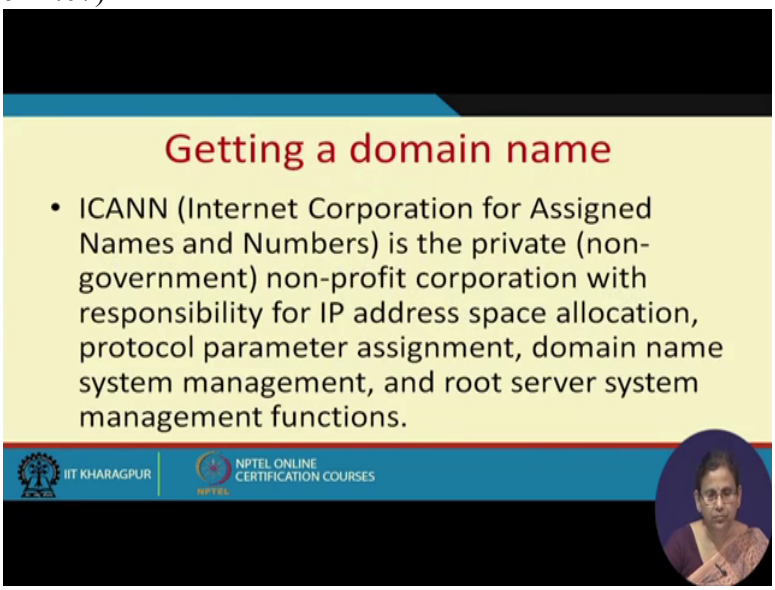
(Refer Slide Time 08:57)



Let us see how this D N S works. Whenever a client enters a domain name into his browser, the browser contacts the client's I S P for the I P address of the domain name. See why I S P? Though the main domain name to I P conversion table stays in your corresponding domain server, as I told you this is a distributed database system and a replica, a part of it is available if already some previous instances of that conversion has happened in that proxy, in I S P's proxy or your own organization proxy or in your local machine, then that name to address conversion is available there itself in that corresponding D N S.

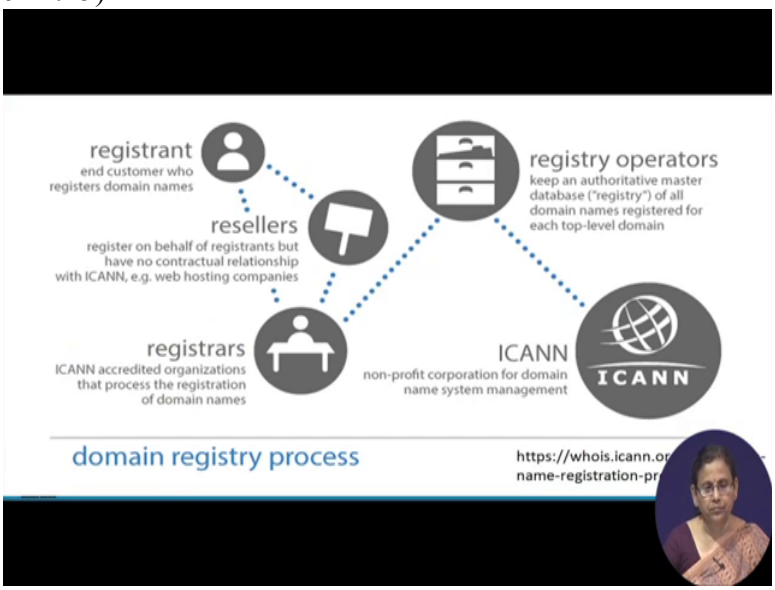
Then I S P first tries to answer it by using its cached data that by this cached data we mean this is the domain to, I P to domain conversion then if the answer is found it is returned. If it is not found then, you know, it is, it has to be forwarded to corresponding name server of that particular domain directly. By name server we mean dot com, dot in etc. Now in case the address is found, it is actually non-authoritative. By non-authoritative we mean it is simply whatever is exist in the local I S P server, that value is returned. Ok. Now if the name servers are not known, then the I S P looks for the information in the root server or the Registry server. And if nothing is there then you know that what kind of message you get; that this address is not found or something.

(Refer Slide Time 11:07)



Now the question is if you would like to have a website how would you like to get a domain name. The organization responsible for giving this domain name is actually a privately, private, I mean it's a non-government and non-profit corporation and it has taken the responsibility for I P address space allocation. This organization is called Internet Corporation for Assigned Names and Numbers I C A N N. Now

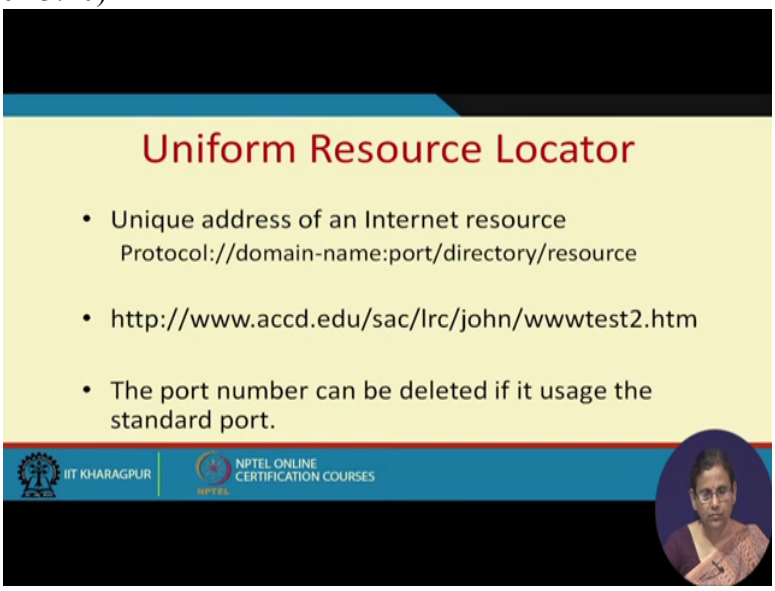
(Refer Slide Time 11:45)



this I C A N N in turn, this particular diagram we have referred from this I C A N N's website, this I C A N N is, has appointed a number of organizations who act as a registrars and they help in the process of registering the domain name. They are empowered to process the registration of domain name. Now these registrars in turn might be, might be giving certain, a part of their responsibility to certain resellers. So these resellers can register on behalf of the

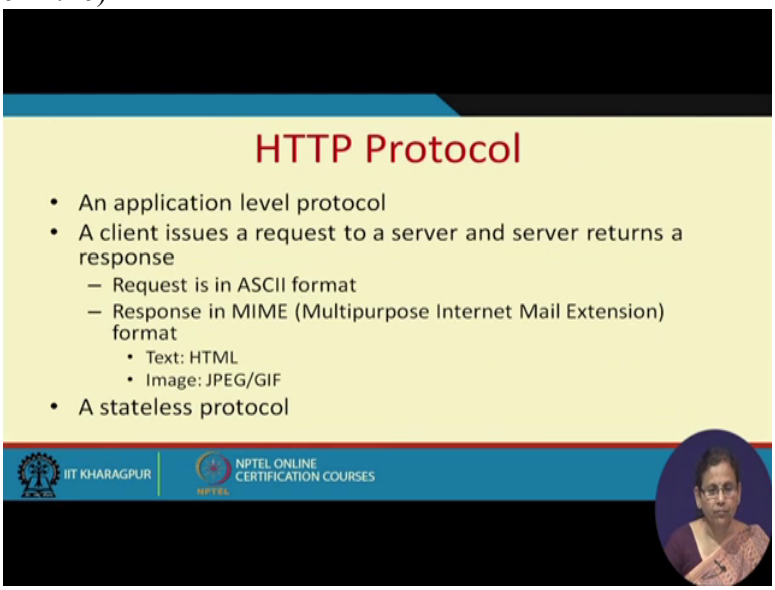
registrant but they not have any contractual relationship with I C A N N. These resellers and the registrant is the person, end customer who actually seeks for the domain name. Then your Registry operators like that of your, for various name servers they are actually responsible for this address conversion and registrar who is, who is appointed by I C A N N, is recognized by I C A N N connects these Registry operators for this conversion process.

(Refer Slide Time 13:10)



Now once you get the web site address, each page within this website can be uniquely identified. This unique identification is called the Uniform Resource Locator; so each of the pages is called the Uniform Resource Locator. So each page within your website or any resource that way is called a Uniform Resource Locator. Now if you look at the parts of a Uniform Resource Locator then it consists of the protocol name that you write as H T T P, F T P or whatever, then you have your domain name colon port; if the port is the default port then you don't write anything followed by the directory followed by the resource. This is one example. Then the port number can be deleted if it is a usage standard port that I already told you.

(Refer Slide Time 14:16)

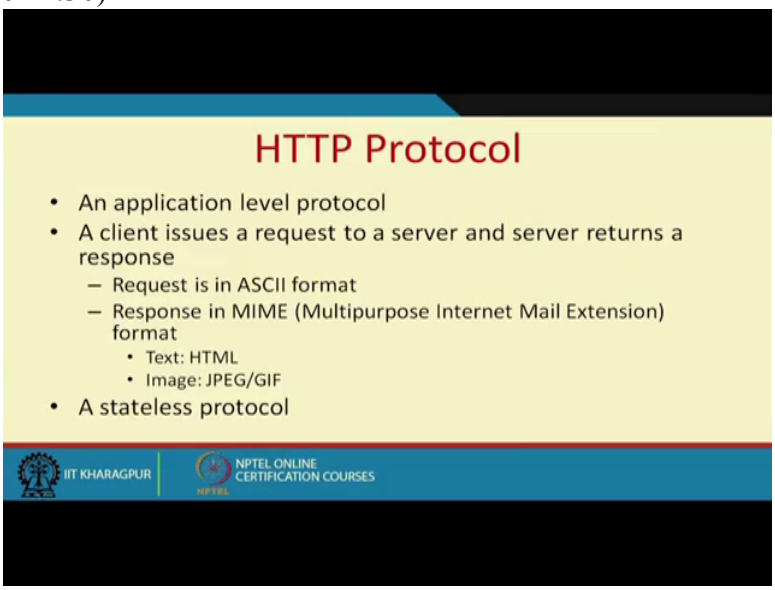


Now the protocol who is responsible for this, moving the data from your server to

(Refer Slide Time 14:28)

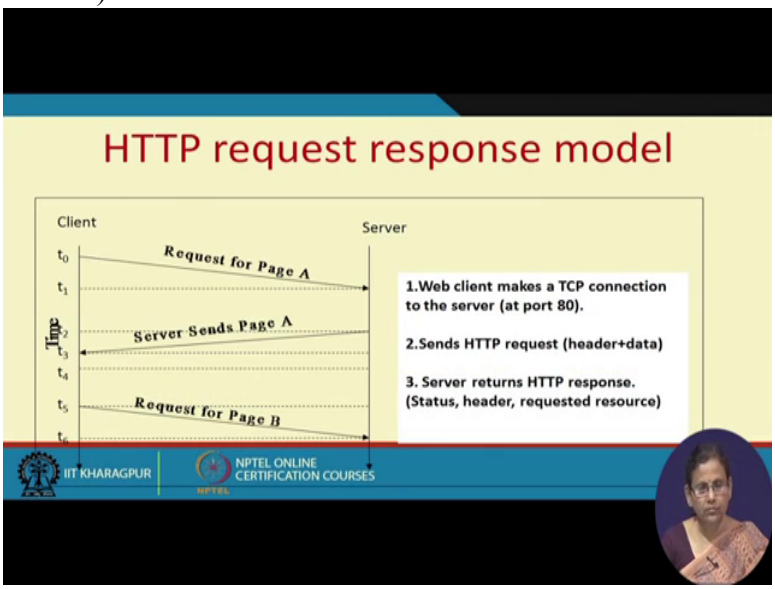
the host, your client is called the H T T P protocol. So this H T T P, the Hypertext

(Refer Slide Time 14:36)



Transfer Protocol is application level protocol and client issues a request to the server using this protocol and request, this request is responded by the server. While request is sent, it is sent in the ASCII format and while it is returned, it is returned in MIME format which can, along with the text you can also handle the other embedded resources like that of images etc. One important fact about H T T P protocol is it is a stateless protocol. By stateless protocol we mean the details that is provided by user in one request when the response is given back, that detail is forgotten by the server. Which means each request is treated as a new request.

(Refer Slide Time 15:37)

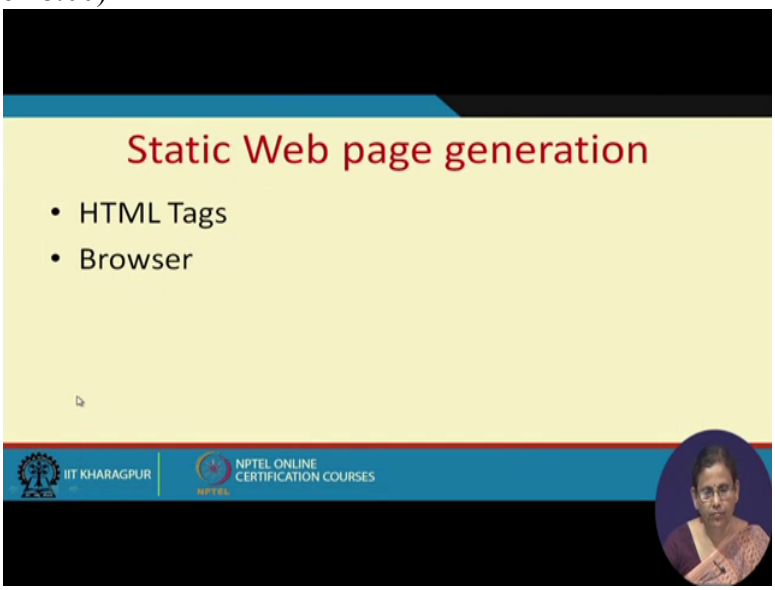


So this is a typical H T T P request response model. So here I would like to tell one more thing, so you must be thinking if each request is sent as a new request then how it is possible that while connecting to the internet using our let's say, mail service, something, let’s say

Gmail or something we always log in and every time if we are treated as new then how it is possible that every time we are remembered. There are various other methodologies which we are not going to discuss. But in its core, H T T P is a stateless protocol.

Now this is a typical H T T P Request Response model. A client can, and see there are many delays happening. The client sends a request to the server. And the server takes, and there is a delay, network delay from the starting of the request till the request reaches at the server. Then there is some delay at the server side where the server prepares the page with appropriate header and data and this is the web server. It in turn may have to connect to the application server or database server. After the page is prepared it is sent to the client. And when there is sent to the client there is some kind of delay as well. Then at the client side the client takes some time to, the client who, the client's machine has the browser. The client here means the browser. The browser takes some time to organize the page and show that detail, show that detail. Then the, in the mean time the server actually forgets who the client was and when the next request is sent this is actually treated as a new page. So the, when the server returns this H T T P response it not only sends the requested resource that is the requested page, it also sends some kind of header and status information which is also kept in the log file of the server and also sent to the browser of the client.

(Refer Slide Time 18:00)



When it comes to the web pages, these web pages can be either static

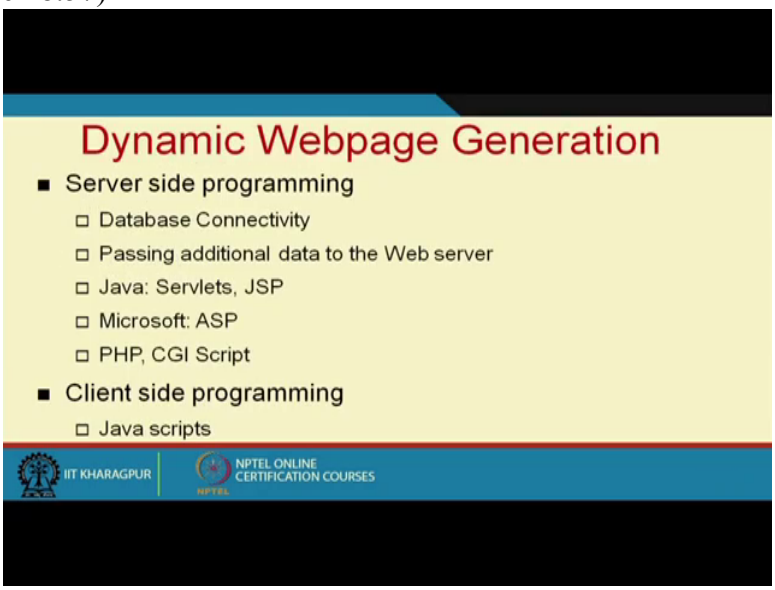
(Refer Slide Time 18:06)

or dynamic. These static web pages consist of H T M L tags within which the various, the text that is supposed to be displayed as well as the other embedded link to the other embedded resources are put. And it is the responsibility of the browser to convert, to understand the formatting that the user has made and it is presented in that manner. In order to maintain the uniformity among the web pages, corresponding cascading style sheets are often used which takes care of the uniform formatting across all the web pages. It has to be, it doesn't

(Refer Slide Time 18:56)

have to be repeated every time.

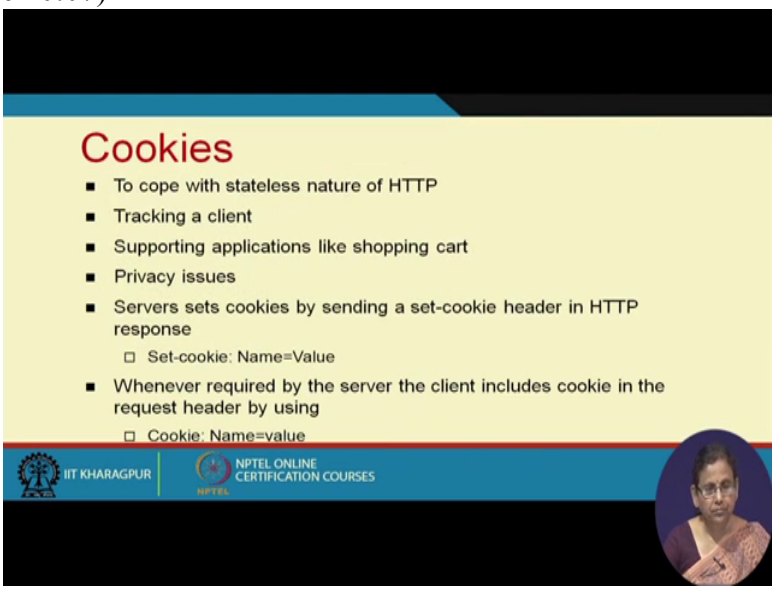
(Refer Slide Time 18:57)



Then some pages can be dynamic. This dynamism can happen in 2 different ways. It can be, it can happen in the client side where some Javascripts can be written but if you like to have database connectivity and because of the database connectivity suppose you are requesting the details of one employee.

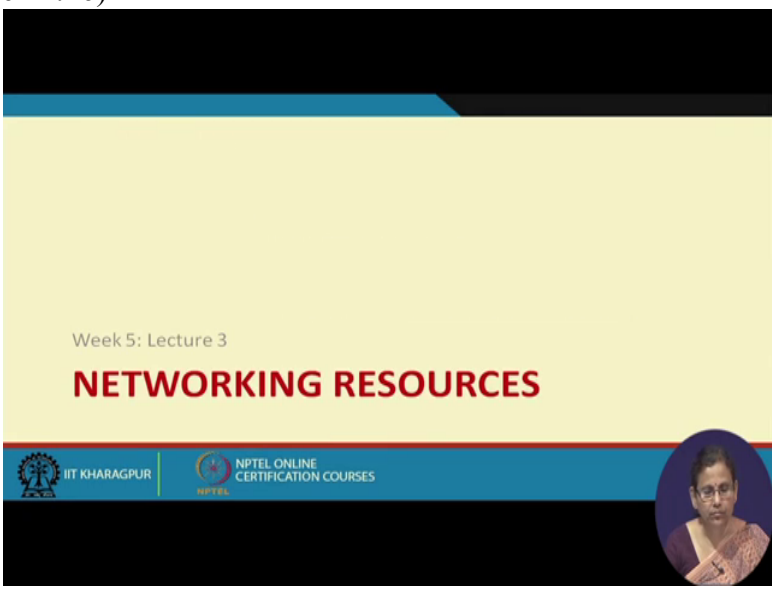
Suppose you are trying to access your employee database over the internet. You are asking the details of one employee. So next time you can ask the details of another employee. So every time you should give this employee id or employee's name the database will be searched and corresponding query results will be embedded within H T M L page and it will be returned back to you. So this technology is a server side technology where the database connectivity is required and your A S P, J S P, P H P, C G I scripts all these are the technologies behind these server-side programming which creates dynamic web pages so web pages from the server side.

(Refer Slide Time 20:07)



Again as I was telling you already that if you would like to maintain the state of H T T P protocol there are many ways it can be done. One of the ways is maintaining the cookies in your website. These cookies are some kind of textual, textual values which are stored in the client machine if the client permits. So these cookies also store some information by the website and they are, whenever a H T T P request is sent, they are again sent, this information embedded in the cookies is again sent back to the server side so that the server is aware of those details. So while collecting the end user data, these cookies play a great role. So they, not only this emails etc. they also support this applications like that which requires persistent connection which otherwise is not given by H T T P is maintained by cookies. These cookies support applications like that of shopping cart. So usually server sets the cookies in the server's header by a pair by embedding some kind of, embedding some kind of tag for keeping the cookies. This is set cookie tag can actually, the set cookie header can actually keep this name value pair about specific stuffs within the client's machine. Now whenever required by the server the client includes the cookie in the request header by using some procedure where some kind of header is again sent from the client side attaching this name value pair.

(Refer Slide Time 22:16)



So with this

(Refer Slide Time 22:19)

we finish our lecture in this series and in the next lecture we are going to talk about various networking resources. Thank you very much.