UNIT 5 Application Layer

CO 1 Explain Basic concept, OSI reference Model. Services .Role of each layer in OSI Model. TCP/IP. Network devices. Transmission Media, Analog and Digital Transmission

CO 2 Apply Channel allocation . Framing. Frame control and Error Control Techniques

CO 3 Describe the function of Network Layer, Logical addressing and Subletting, Routing Mechanism

CO 4 Explain the function of Session and Presentation Layer.

CO 5 Explain different Protocol used at different Application layer HTTP.SNMP..FTP.TELNET. VPN

Domain Name System (DNS)

Domain Name System (DNS) is a hostname for IP address translation service. DNS is a distributed database implemented in a hierarchy of name servers. It is an application layer protocol for message exchange between clients and servers. It is required for the functioning of the Internet.

Requirement

Every host is identified by the IP address but remembering numbers is very difficult for people also the IP addresses are not static therefore a mapping is required to change the domain name to the IP address. So DNS is used to convert the domain name of the websites to their numerical IP address.

Domain

There are various kinds of DOMAIN:

1. Generic domains: .com(commercial), .edu(educational), .mil(military),

.org(nonprofit organization), .net(similar to commercial) all these are generic domains.

2. Country domain: .in (India) .us .uk

3. Inverse domain: if we want to know what is the domain name of the website. Ip to domain name mapping. So DNS can provide both the mapping.

Organization of Domain

It is very difficult to find out the IP address associated with a website because there are millions of websites and with all those websites we should be able to generate the IP address immediately, there should not be a lot of delays for that to happen organization of the database is very important.

•DNS record: Domain name, IP address what is the validity?? what is the time to live ?? and all the information related to that domain name. These records are stored in a tree-like structure.

•Namespace: Set of possible names, flat or hierarchical. The naming system maintains a collection of bindings of names to values – given a name, a resolution mechanism returns the corresponding value.

•Name server: It is an implementation of the resolution mechanism. DNS (Domain Name System) = Name service in Internet – A zone is an administrative unit, and a domain is a subtree.

Name-to-Address Resolution

The host requests the DNS name server to resolve the domain name. And the name server returns the IP address corresponding to that domain name to the host so that the host can future connect to that IP address.

- Hierarchy of Name Servers Root name servers: It is contacted by name servers that cannot resolve the name. It contacts the authoritative name server if name mapping is not known. It then gets the mapping and returns the IP address to the host.

- Top-level domain (TLD) server: It is responsible for com, org, edu, etc, and all top-level country domains like uk, fr, ca, in, etc. They have info about authoritative domain servers and know the names and IP addresses of each authoritative name server for the second-level domains.

- Authoritative name servers are the organization's DNS servers, providing authoritative hostnames to IP mapping for organization servers. It can be maintained by an organization or service provider. In order to reach cse.dtu.in we have to ask the root DNS server, then it will point out to the top-level domain server and then to the authoritative domain name server which actually contains the IP address. So the authoritative domain server will return the associative IP address.

Domain Name Server

The client machine sends a request to the local name server, which, if the root does not find the address in its database, sends a request to the root name server, which in turn, will route the query to a top-level domain (TLD) or authoritative name server. The root name server can also contain some hostName to IP address mappings. The Top-level domain (TLD) server always knows who the authoritative name server is. So finally the IP address is returned to the local name server which in turn returns the IP address to the host.

Domain Name Server

Working of DNS

The working of DNS starts with converting a hostname into an IP Address. A domain name serves as a distinctive identification for a website. It is used in place of an IP address to make it simpler for consumers to visit websites. Domain Name System works by executing the database whose work is to store the name of hosts which are available on the Internet. The top-level domain server stores address information for top-level domains such as .com and .net,

.org, and so on. If the Client sends the request, then the DNS resolver sends a request to DNS Server to fetch the IP Address. In case, when it does not contain that particular IP Address with a hostname, it forwards the request to another DNS Server. When IP Address has arrived at the resolver, it completes the request over Internet Protocol.

World Wide Web

The World Wide Web or Web is basically a collection of information that is linked together from points all over the world. It is also abbreviated as WWW.

•World wide web provides flexibility, portability, and user-friendly features.

•It mainly consists of a worldwide collection of electronic documents (i.e, Web Pages).

•It is basically a way of exchanging information between computers on the Internet.

•The WWW is mainly the network of pages consists of images, text, and sounds on the Internet which can be simply viewed on the browser by using the browser software.

•It was invented by Tim Berners-Lee.

Components of WWW

The Components of WWW mainly falls into two categories:

1. Structural Components
2. Semantic Components

Architecture of WWW

The WWW is mainly a distributed client/server service where a client using the browser can access the service using a server. The Service that is provided is distributed over many different locations commonly known as sites/websites.

•Each website holds one or more documents that are generally referred to as web pages.

•Where each web page contains a link to other pages on the same site or at other sites.

•These pages can be retrieved and viewed by using browsers.

In the above case, the client sends some information that belongs to site A. It generally sends a request through its browser (It is a program that is used to fetch the documents on the web).

and also the request generally contains other information like the address of the site, web page(URL).

The server at site A finds the document then sends it to the client. after that when the user or say the client finds the reference to another document that includes the web page at site B.

The reference generally contains the URL of site B. And the client is interested to take a look at this document too. Then after the client sends the request to the new site and then the new page is retrieved.

Now we will cover the components of WWW in detail.

1.Client/Browser

The Client/Web browser is basically a program that is used to communicate with the webserver on the Internet.

•Each browser mainly comprises of three components and these are:

o Controller

o Interpreter

o Client Protocols

•The Controller mainly receives the input from the input device, after that it uses the client programs in order to access the documents.

•After accessing the document, the controller makes use of an interpreter in order to display the document on the screen.

•An interpreter can be Java, HTML, javascript mainly depending upon the type of the document.

•The Client protocol can be FTP, HTTP, TELNET.

2. Server

The Computer that is mainly available for the network resources and in order to provide services to the other computer upon request is generally known as the server.

•The Web pages are mainly stored on the server.

•Whenever the request of the client arrives then the corresponding document is sent to the client.

•The connection between the client and the server is TCP.

•It can become more efficient through multithreading or multiprocessing. Because in this case, the server can answer more than one request at a time.

3. URL

URL is an abbreviation of the Uniform resource locator.

•It is basically a standard used for specifying any kind of information on the Internet.

•In order to access any page the client generally needs an address.

•To facilitate the access of the documents throughout the world HTTP generally makes use of Locators.

URL mainly defines the four things:

•Protocol

It is a client/server program that is mainly used to retrieve the document. A commonly used protocol is HTTP.

•Host Computer

It is the computer on which the information is located. It is not mandatory because it is the name given to any computer that hosts the web page.

•Port

The URL can optionally contain the port number of the server. If the port number is included then it is generally inserted in between the host and path and is generally separated from the host by the colon.

•Path

It indicates the pathname of the file where the information is located.

4.HTML

HTML is an abbreviation of Hypertext Markup Language.

•It is generally used for creating web pages.

•It is mainly used to define the contents, structure, and organization of the web page.

5.XML

XML is an abbreviation of Extensible Markup Language. It mainly helps in order to define the common syntax in the semantic web.

Features of WWW

Given below are some of the features provided by the World Wide Web:

- Provides a system for Hypertext information
- Open standards and Open source
- Distributed.
- Mainly makes the use of Web Browser in order to provide a single interface for many services.
- Dynamic
- Interactive
- Cross-Platform

Advantages of WWW

Given below are the benefits offered by WWW:

- It mainly provides all the information for Free.
- Provides rapid Interactive way of Communication.
- It is accessible from anywhere.
- It has become the Global source of media.
- It mainly facilitates the exchange of a huge volume of data.

Disadvantages of WWW

There are some drawbacks of the WWW and these are as follows;

- It is difficult to prioritize and filter some information.
- There is no guarantee of finding what one person is looking for.
- There occurs some danger in case of overload of Information.
- There is no quality control over the available data.
- There is no regulation.

HTTP Protocol

HTTP stands for Hypertext Transfer Protocol and is mainly used to access the data on the world wide web i.e (WWW). The HTTP mainly functions as the combination of FTP(File Transfer Protocol) and SMTP(Simple Mail Transfer Protocol).

- HTTP is one of the protocols used at the Application Layer.
- The HTTP is similar to FTP because HTTP is used to transfer the files and it mainly uses the services of TCP.
- Also, HTTP is much simpler than FTP because there is only one TCP connection.
- In HTTP, there is no separate control connection, as only data is transferred between the client and the server.
- The HTTP is like SMTP because the transfer of data between the client and server simply looks like SMTP messages. But there is a difference unlike SMTP, the HTTP messages are not destined to be read by humans as they are read and interpreted by HTPP Client(that is browser) and HTTP server.
- Also, SMTP messages are stored and then forwarded while the HTTP messages are delivered immediately.
- The HTTP mainly uses the services of the TCP on the well-known port that is port 80.

•HTTP is a stateless protocol.

•In HTTP, the client initializes the transaction by sending a request message, and the server replies by sending a response.

•This protocol is used to transfer the data in the form of plain text, hypertext, audio as well as video, and so on.

Working of HTTP

The HTTP makes use of Client-server architecture. As we have already told you that the browser acts as the HTTP client and this client mainly communicates with the webserver that is hosting the website.

The figure shows the HTTP transaction

The format of the request and the response message is similar. The Request Message mainly consists of a request line, a header, and a body sometimes. A Response message consists of the status line, a header, and sometimes a body.

At the time when a client makes a request for some information (say client clicks on the hyperlink) to the webserver. The browser then sends a request message to the HTTP server for the requested objects.

After that the following things happen:

•There is a connection that becomes open between the client and the webserver through the TCP.

•After that, the HTTP sends a request to the server that mainly collects the requested data.

•The response with the objects is sent back to the client by HTTP

•At last, HTTP closes the connection.

Let us take a look at the format of the request message and response message:

Request Line and Status line

The first line in the Request message is known as the request line, while the first line in the Response message is known as the Status line.

where,

Request Type

This field is used in the request line. The are several request types that are defined and these are mentioned in the table given below;

Name of Method

Actions

GET

This method is used to request a document from the server.

HEAD

This method mainly requests information about a document and not the document itself

POST

This method sends some information from the client to the server.

PUT

This method sends a document from the server to the client.

TRACE

This method echoes the incoming request.

CONNECT

This method means reserved

OPTION

In order to inquire about the available options.

URL

URL is a Uniform Resource locator and it is mainly a standard way of specifying any kind of information on the Internet.

HTTP Version

The current version of the HTTP is 1.1.

Status Code

The status code is the field of the response message.The status code consists of three digits.

Status Phrase

This field is also used in the response message and it is used to explain the status code in the form of text.

Header

The header is used to exchange the additional information between the client and the server. The header mainly consists of one or more header lines. Each header line has a header name, a colon, space, and a header value.

The header line is further categorized into four:

- **General Header**

It provides general information about the message and it can be present in both request and response.

- **Request Header**

It is only present in the request message and is used to specify the

configuration of the client and the format of the document preferred by the client

- **Response Header**

This header is only present in the response header and mainly specifies the configuration of the server and also the special information about the request.

- **Entity Header**

It is used to provide information about the body of the document.

Body

It can be present in the request message or in the response message. The body part mainly contains the document to be sent or received.

Features of HTTP

The HTTP offers various features and these are as follows:

1. **HTTP is simple**

The HTTP protocol is designed to be plain and human-readable.

2. **HTTP is stateless**

Hypertext transfer protocol(HTTP) is a stateless protocol, which simply means that there is no connection among two requests that are being consecutively carried out on the same connection. Also, both the client and the server know each other only during the current requests and thus the core of the HTTP is itself a stateless one, On the other hand, the HTTP cookies provide in making use of stateful sessions.

3. **HTTP is extensible**

The HTTP can be integrated easily with the new functionality by providing a simple agreement between the client and the server.

4. **HTTP is connectionless**

As the HTTP request is initiated by the browser (HTTP client) and as per the request information by the user, after that the server processes the request of the client and then responds back to the client

Advantages of HTTP

Given below are the benefits of using HTTP:

1. There is no runtime support required to run properly.
2. As it is connectionless so there is no overhead in order to create and maintain the state and information of the session.
3. HTTP is usable over the firewalls and global application is possible.
4. HTTP is platform-independent.
5. HTTP reports the errors without closing the TCP connection.
6. Offers Reduced Network congestions.

Disadvanatges of HTTP

There are some drawbacks of using the HTTP protocol:

- HTTP is not optimized for mobile.
- HTTP is too verbose.
- It can be only used for point-to-point connections.
- This protocol does not have push capabilities.
- This protocol does not offer reliable exchange without the retry logic.

The HTTP supports proxy servers. A proxy server is basically a computer that keeps the copies of the responses to recent requests. The proxy server mainly reduces the load on the originals server.In order to use the proxy server, the client must be configured in order to access the proxy instead of the target server.

HTTP Connections

HTTP connections can be further classified into two:

- **Persistent Connection**

Non persistent Connection Let us discuss them one by one:

Persistent Connection

In the persistent HTTP connection, all the requests and their corresponding responses are sent over the same TCP connections. The 1.1 version of the HTTP specifies a persistent connection by default.

In this type of connection, the server leaves the connection open for more requests after sending a response. Also, the server can close the connection at the request of the client or upon reaching the time-out.

In a Persistent connection, a single TCP connection is mainly used for sending multiple objects one after the other.

Usually, the length of the data is sent along with each response. There are some cases when the server does not know the length of the data this happens when the document is created dynamically and in such cases, the server informs the client that length is not known and closes the connection after sending the data so in order let the client Inform about the end of the data.

2. Nonpersistent Connection

In the Non-persistent HTTP connection, one TCP connection is made for each request/response; it means there is a separate for each object.

Following are the steps used;

•The client opens a TCP connection and then sends a request.

•After that, the server sends the response and then closes the connection.

•Then the client reads the data and until it encounters an end-of-file marker then it closes the connection.

This connection imposes a high overhead on the server because N different buffers are required by the server, and the start procedure is slow each time when a connection is opened.

The non-persistent connection is supported by the HTTP 1.0 version.

Electronic mail

Electronic mail (e-mail) is a computer-based program that allows users to send and receive messages. E-mail is the electronic version of a letter, but with time and flexibility advantages. While a letter can take anywhere from a week to a couple of months to reach its intended destination, an e-mail is sent virtually almost instantly.

Messages in the mail contain not just text but also photos, audio, and video data. A person sending an e-mail is a sender, and the person receiving it is the recipient.

Electronic mail is one of the most well-known network services. Electronic mail is a computer-based service that allows users to communicate with one another by exchanging messages. Email information is transmitted via email servers and uses a variety of TCP/IP protocols. For example, the simple mail transfer protocol (SMTP) is a protocol that is used to send messages. Similarly, IMAP or POP receives messages from a mail server.

Features Of Electronic Mail

•Spontaneity: In a couple of seconds, you may send a message to anybody on the globe.

•Asynchronous: You may send the e-mail and let the recipient view it at their leisure.

•Attachments of data, pictures, or music, frequently in compressed forms, can be delivered as an e-mail to a person anywhere in the world.

•Addresses can be stored in an address book and retrieved instantly.

•Through an e-mail, a user can transfer multiple copies of a message to various individuals.

Services offered by Electronic Mail

Composition: Creating messages and responses is referred to as composition. Transfer: Sending mail from the sender to the receiver is known as a transfer. Reporting: Mail delivery confirmation is known as reporting. It allows users to see

Displaying: It refers to presenting messages so that the user can understand them.

Disposition: This stage concerns the recipient's actions after receiving mail, such as

as saving it, deleting it before reading it, or after reading it.

Components Of Electronic Mail

The following are the essential components of an e-mail system:

1.**User Agent (UA)**

2.Message Transfer Agent (MTA)

Message Access Agent

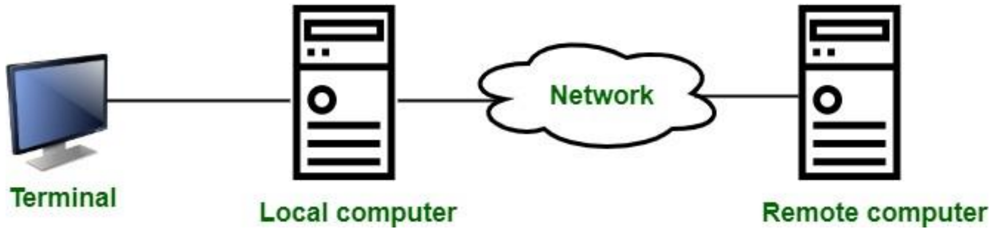
User Agent (UA) The User-Agent is a simple software that sends and receives mail. It is also known as a mail reader. It supports a wide range of instructions for sending, receiving

Remote Login

Remote Login is a process in which user can login into remote site i.e. computer and use services that are available on the remote computer. With the help of

remote login a user is able to understand result of transferring and result of processing from the remote computer to the local computer.

Figure – Remote login



It is implemented using Telnet.

Procedure of Remote Login :

1.When the user types something on local computer, then local operating system accepts character.

2. Local computer does not interpret the characters, it will send them to TELNET client.

3.TELNET client transforms these characters to a universal character set called Network Virtual Terminal (NVT) characters and it will pass them to the local TCP/IP protocol Stack.

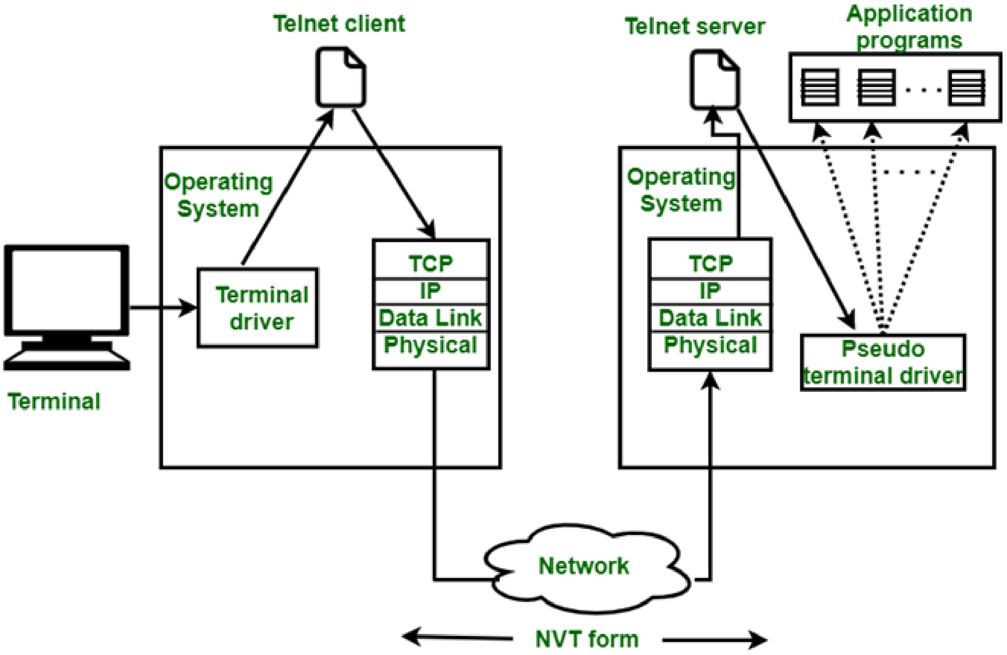
4.Commands or text which is in the form of NVT, travel through Internet and it will arrive at the TCP/IP stack at remote computer.

5.Characters are then delivered to operating system and which later on passed to TELNET server.

6.Then TELNET server changes that characters to characters which can be understandable by remote computer.

7.Remote operating system receives character from a pseudo-terminal driver, which is a piece of software that pretends that characters are coming from a terminal.

8.Operating system then passes character to the appropriate application program.



NVT Character Set :

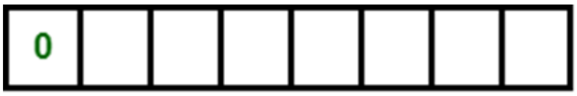
With NVT Character set, TELNET client translates characters into NVT form and deliver to network.

•**TELNET** server translates data and commands from NVT form to the other form that will be understandable by remote computer.

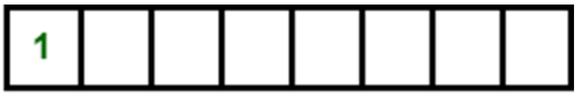
•**NVT uses 2 sets of characters**, one for data and other for control. Size of both characters is 8-bit bytes.

•For data, NVT is an 8-bit character set in which 7 lowest bits are same as ASCII and highest order bit is 0.

•For control characters, NVT uses an 8-bit character set in which the highest bit is set to 1.



Data character



Control character

File Transfer Protocol

FTP (File Transfer Protocol) is a network protocol for transmitting files between computers over Transmission Control Protocol/Internet Protocol (TCP/IP) connections. Within the TCP/IP suite, FTP is considered an application layer protocol.

In an FTP transaction, the end user's computer is typically called the local host. The second computer involved in FTP is a remote host, which is usually a server. Both computers need to be connected via a network and configured properly to transfer files via FTP. Servers must be set up to run FTP services, and the client must have FTP software installed to access these services.

Although many file transfers can be conducted using Hypertext Transfer Protocol (HTTP) -- another protocol in the TCP/IP suite -- FTP is still commonly used to transfer files behind the scenes for other applications, such as banking services. It is also sometimes used to download new applications via web browsers.

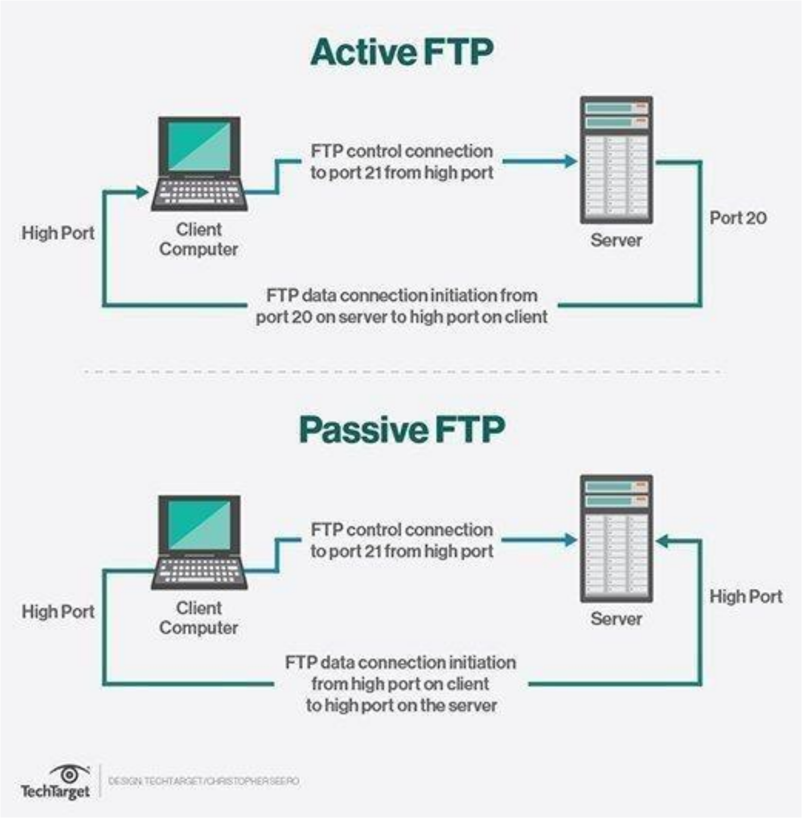
How does FTP work?

FTP is a client-server protocol that relies on two communications channels between the client and server: a command channel for controlling the conversation and a data channel for transmitting file content.

Here is how a typical FTP transfer works:

1. A user typically needs to log on to the FTP server, although some servers make some or all of their content available without a login, a model known as anonymous FTP.
2. The client initiates a conversation with the server when the user requests to download a file.
3. Using FTP, a client can upload, download, delete, rename, move and copy files on a server.

FTP sessions work in active or passive modes:


Active FTP
FTP control connection to port 21 from high port
High Port
Client Computer
Server
Port 20
FTP data connection initiation from port 20 on server to high port on client
Passive FTP
FTP control connection to port 21 from high port
High Port
Client Computer
Server
High Port
FTP data connection initiation from high port on client to high port on the server
TechTarget
DESIGN: TECHTARGET/CHRISTOPHER SEERO

- Active mode. After a client initiates a session via a command channel request, the server creates a data connection back to the client and begins transferring data.

- Passive mode. The server uses the command channel to send the client the information it needs to open a data channel. Because passive mode has the client initiating all connections, it works well across firewalls and network address translation gateways.

Users can work with FTP via a simple command-line interface -- from a console or terminal window in Microsoft Windows, Apple macOS or Linux -- or with a dedicated graphical user interface. Web browsers can also serve as FTP clients

Why is FTP important and what is it used for?

FTP is a standard network protocol that can enable expansive file transfer capabilities across IP networks. Without FTP, file and data transfer can be managed with other mechanisms -- such as email or an HTTP web service -- but those other options lack the clarity of focus, precision and control that FTP enables.

FTP is used for file transfers between one system and another, and it has several common use cases, including the following:

•Backup. FTP can be used by backup services or individual users to backup data from one location to a secured backup server running FTP services.

•Replication. Similar to backup, replication involves duplication of data from one system to another but takes a more comprehensive approach to provide higher availability and resilience. FTP can also be used to facilitate this.

•Access and data loading. FTP is also commonly used to access shared web hosting and cloud services as a mechanism to load data onto a remote system.

FTP types

There are several different ways an FTP server and client software can conduct a file transfer using FTP:

•**Anonymous FTP**. This is the most basic form of FTP. It provides support for data transfers without encrypting data or using a username and password. It's most commonly used for download of material that is allowed for unrestricted distribution. It works on port

•**Password-protected FTP**. This is also a basic FTP service, but it requires the use of a username and password, though the service might not be encrypted or secure. It also works on port 21.

•**FTP Secure (FTPS).** Sometimes referred to as FTP Secure Sockets Layer (FTP- SSL), this approach enables implicit Transport Layer Security (TLS) as soon as

an FTP connection is established. FTPS was initially used to help enable a more secure form of FTP data transfer. It typically defaults to using port 990.

• **FTP over explicit SSL/TLS (FTPES**). This approach enables explicit TLS support by upgrading an FTP connection over port 21 to an encrypted connection. This is a commonly used approach by web and file sharing services to enable secure file transfers.

• **Secure FTP (SFTP).** This is technically not an FTP protocol, but it functions similarly. Rather, SFTP is a subset of the Secure Shell (SSH) protocol that runs over port 22. SSH is commonly used by systems administrators to remotely and securely access systems and applications, and SFTP provides a mechanism within SSH for secure file transfer.

Network management

•Network management is the procedure of administering, managing and working a data network using a network management system. Current network management systems use software and hardware to constantly collect and analyse data and push out configuration changes for increasing performance, reliability, and security.

•It involves configuring monitoring and possibly reconfiguring components in a network with the goal of providing optimal performance, minimum downtime, proper security, accountability and flexibility.

Features

There are various features of network management which are as follows −

Network automation

- One defining feature of a modern network management system is network automation. This is the procedure of automating the configuring, handling, testing, deploying, and operating of physical and virtual devices inside a network. Network service availability increases when everyday network tasks and functions are automated and repetitive processes are controlled and managed automatically.

Network administration

Network administration encompasses tracking network resources, including switches, routers, and servers. It also includes performance monitoring and software updates.

Network Operation

This contains smooth network functioning as created and intended, including close monitoring of activities to quickly and effectively address and fix problems as they occur and preferably even before users are aware of the problem.

Network assurance

Network assurance features are often included in modern network management systems. These features help improve network performance, customer experience, and security. Assurance systems help network analytics, application analytics, and policy analytics, as well as AI and ML, to achieve full network assurance.

Network provisioning

Network provisioning involves network resource configuration for the purposes of supporting any given service, like voice functions or accommodating additional users.

Network maintenance

Network maintenance covers upgrades and fixes to network resources. It also consists of proactive and remediation activities executed by working with network administrators, such as replacing network gear like routers and switches.

Network analytics

Network analytics is a software tool that compares incoming information against preprogrammed operational models and makes functional decisions for improving network performance.

Data compression

Data compression is the function of presentation layer in OSI reference model. Compression is often used to maximize the use of bandwidth across a network or to optimize disk space when saving data.

There are two general types of compression algorithms:

1. Lossless compression
2. Lossy compression

Lossless Compression

Lossless compression compresses the data in such a way that when data is decompressed it is exactly the same as it was before compression i.e. there is no loss of data.

A lossless compression is used to compress file data such as executable code, text files, and numeric data, because programs that process such file data cannot tolerate mistakes in the data.

Lossless compression will typically not compress file as much as lossy compression techniques and may take more processing power to accomplish the compression.

Lossless Compression Algorithms

The various algorithms used to implement lossless data compression are :

1. Run length encoding
2. Differential pulse code modulation
3. Dictionary based encoding

1. Run length encoding

- This method replaces the consecutive occurrences of a given symbol with only one copy of the symbol along with a count of how many times that symbol occurs. Hence the names ‘run length’.
- For example, the string AAABBCDDDD would be encoded as 3A2BIC4D.
- A real life example where run-length encoding is quite effective is the fax machine. Most faxes are white sheets with the occasional black text. So, a run- length encoding scheme can take each line and transmit a code for while then the number of pixels, then the code for black and the number of pixels and so on.
- This method of compression must be used carefully. If there is not a lot of repetition in the data then it is possible the run length encoding scheme would actually increase the size of a file.

2.Differential pulse code modulation

- In this method first a reference symbol is placed. Then for each symbol in the data, we place the difference between that symbol and the reference symbol used.
- For example, using symbol A as reference symbol, the string AAABBC DDDD would be encoded as AOOOl123333, since A is the same as reference symbol, B has a difference of 1 from the reference symbol and so on.

3. Dictionary based encoding

- One of the best known dictionary based encoding algorithms is Lempel-Ziv (LZ) compression algorithm.
- This method is also known as substitution coder.
- In this method, a dictionary (table) of variable length strings (common phrases) is built.

- This dictionary contains almost every string that is expected to occur in data.
- When any of these strings occur in the data, then they are replaced with the corresponding index to the dictionary.

- In this method, instead of working with individual characters in text data, we treat each word as a string and output the index in the dictionary for that word.
- For example, let us say that the word "compression" has the index 4978 in one particular dictionary; it is the 4978th word is usr/share/dict/words. To compress a body of text, each time the string "compression" appears, it would be replaced by 4978.

Lossy Compression

Lossy compression is the one that does not promise that the data received is exactly the same as data send i.e. the data may be lost.

This is because a lossy algorithm removes information that it cannot later restore.

Lossy algorithms are used to compress still images, video and audio.

Lossy algorithms typically achieve much better compression ratios than the lossless algorithms.

Audio Compression

•Audio compression is used for speech or music.

•For speech, we need to compress a 64-KHz digitized signal; For music, we need to compress a 1.411.MHz signal

•Two types of techniques are used for audio compression:

1. Predictive encoding
2. Perceptual encoding

Predictive encoding

- In predictive encoding, the differences between the samples are encoded instead of encoding all the sampled values.
- This type of compression is normally used for speech.
- Several standards have been defined such as GSM (13 kbps), G. 729 (8 kbps), and G.723.3 (6.4 or 5.3 kbps).

Perceptual encoding

- Perceptual encoding scheme is used to create a CD-quality audio that requires a transmission bandwidth of 1.411 Mbps.
- MP3 (MPEG audio layer 3), a part of MPEG standard uses this perceptual encoding.
- Perceptual encoding is based on the science of psychoacoustics, a study of how people perceive sound.
- The perceptual encoding exploits certain flaws in the human auditory system to encode a signal in such a way that it sounds the same to a human listener, even if it looks quite different on an oscilloscope.
- The key property of perceptual coding is that some sounds can mask other sound. For example, imagine that you are broadcasting a live flute concert and all of a sudden someone starts striking a hammer on a metal sheet. You will not be able to hear the flute any more. Its sound has been masked by the hammer.
- Such a technique explained above is called frequency masking-the ability of a loud sound in one frequency band to hide a softer sound in another frequency band that would have been audible in the absence of the loud sound.
- Masking can also be done on the basis of time. For example: Even if the hammer is not striking on a metal sheet, the flute will be inaudible for a short period of time because the ears turn down its gain when they start and take a finite time to turn up again.
- Thus, a loud sound can numb our ears for a short time even after the sound has stopped. This effect is called temporal masking.

MP3

•MP3 uses these two phenomena, i.e. frequency masking and temporal masking to compress audio signals.

•In such a system, the technique analyzes and divides the spectrum into several groups. Zero bits are allocated to the frequency ranges that are totally masked.

•A small number of bits are allocated to the frequency ranges that are partially masked.

•A larger number. of bits are allocated to the frequency ranges that are not masked.

•Based on the range of frequencies in the original analog audio, MP3 produces three data rates: 96kbps, 128 kbps and