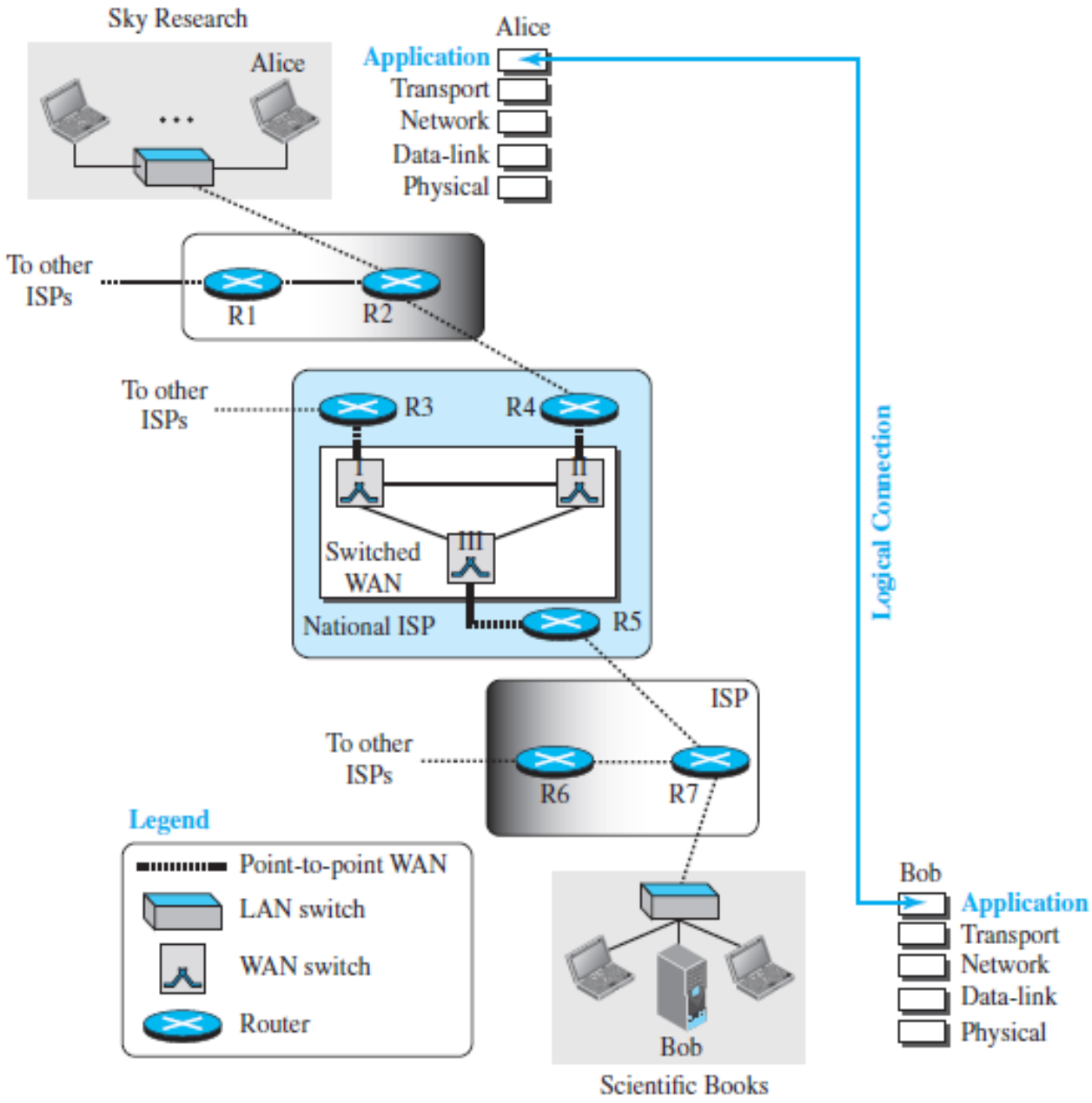
Application Layer

- The application layer enables the user, whether human or software, to access the network. It provides user interfaces and support for services such as electronic mail, file access and transfer, access to system resources, surfing the world wide web, and network management.
- The application layer is responsible for providing services to the user

Application-Layer Paradigms

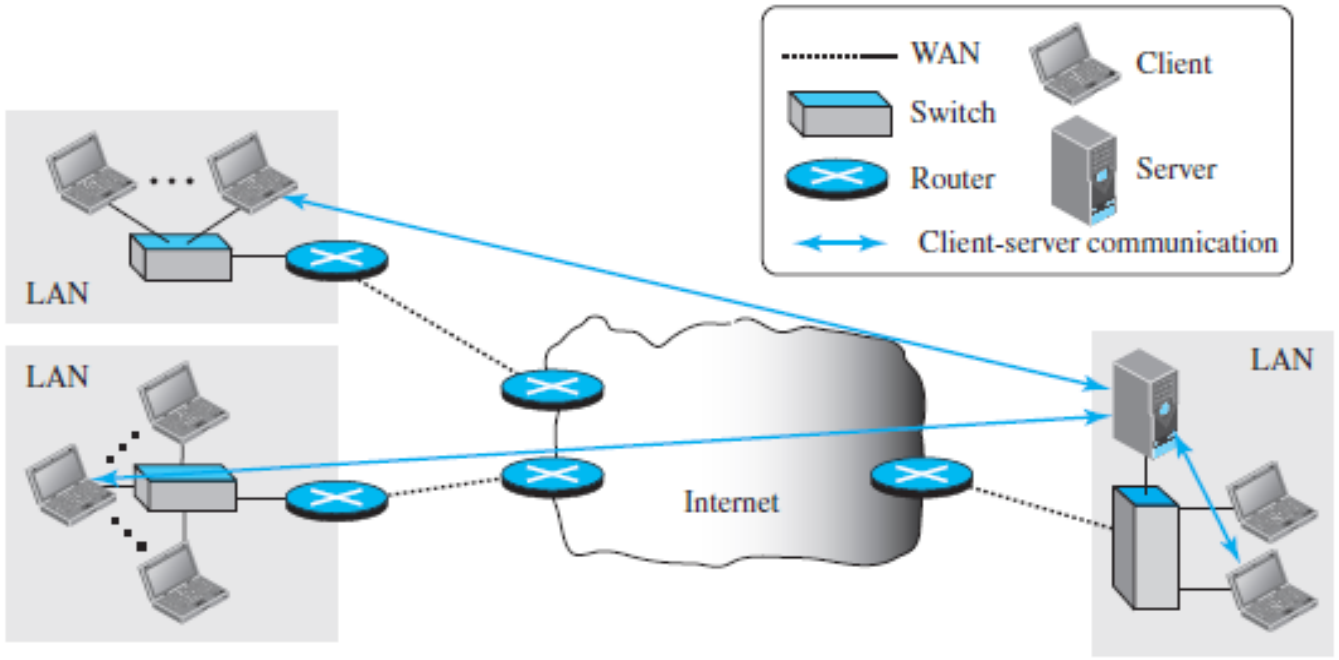
- The application layer provides services to the user.
- Communication is provided using a logical connection, which means that the two application layers assume that there is an imaginary direct connection through which they can send and receive messages.
- To use the Internet we need two application programs to interact with each other: one running on a computer somewhere in the world, the other running on another computer somewhere else in the world.



- Two paradigms have been developed during the lifetime of the Internet to establish a relation between these two programs: client-server architecture and peer to peer architecture.
  - **client-server paradigm**
  - **Peer-to-peer paradigm**

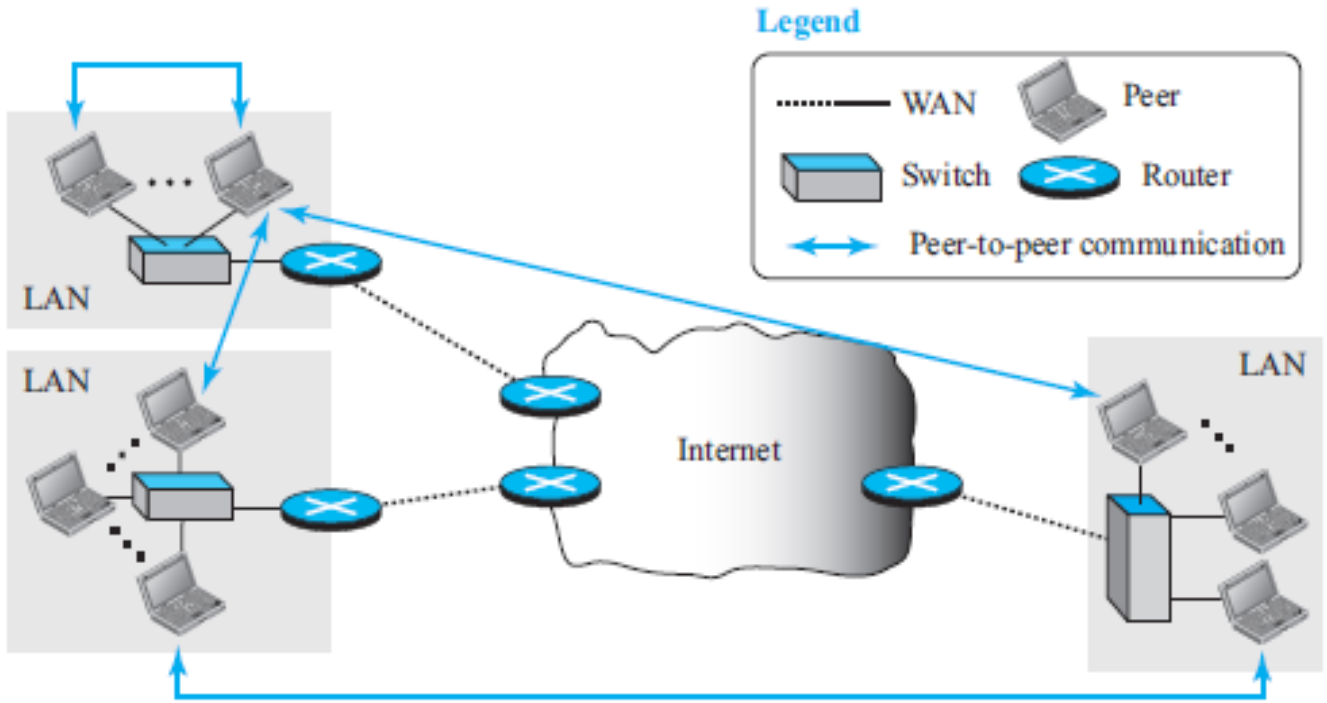
Traditional Paradigm: Client-Server

- The traditional paradigm is called the **client-server paradigm**.
- It was the most popular paradigm until a few years ago.
- In this paradigm, the service provider is an application program, called the server process; it runs continuously, waiting for another application program, called the client process, to make a connection through the Internet and ask for service.



Peer-to-peer Paradigm

- **Peer-to-peer paradigm** (often abbreviated *P2P paradigm*) has emerged to respond to the needs of some new applications.
- In this paradigm, there is no need for a server process to be running all the time and waiting for the client processes to connect.
- The responsibility is shared between peers. A computer connected to the Internet can provide service at one time and receive service at another time. A computer can even provide and receive services at the same time.



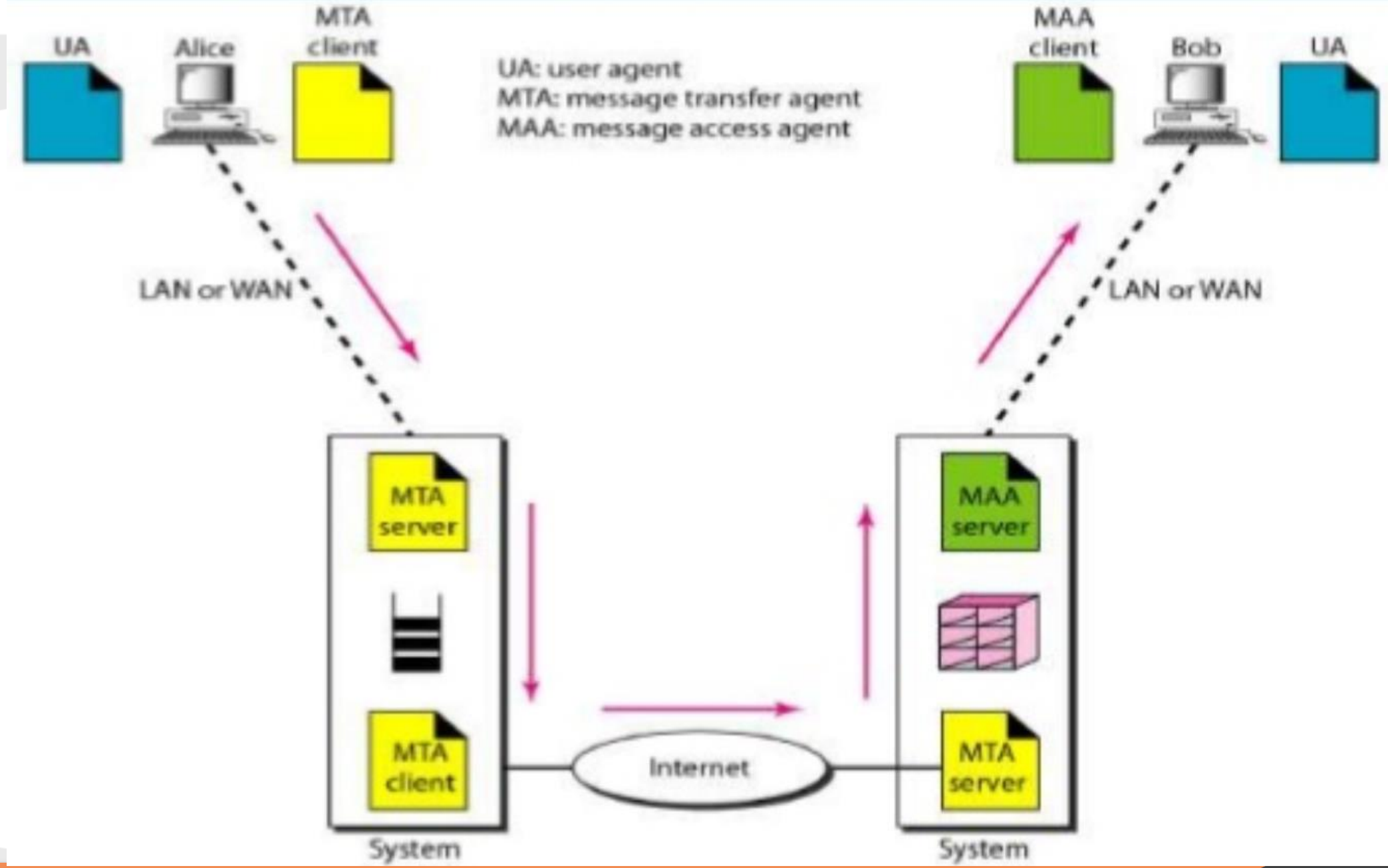
Break

ELECTRONIC MAIL

- One of the most popular Internet services is electronic mail (e-mail). The designers of the Internet probably never imagined the popularity of this application program.

- At the beginning of the Internet era, the messages sent by electronic mail were short and consisted of text only; they let people exchange quick memos.

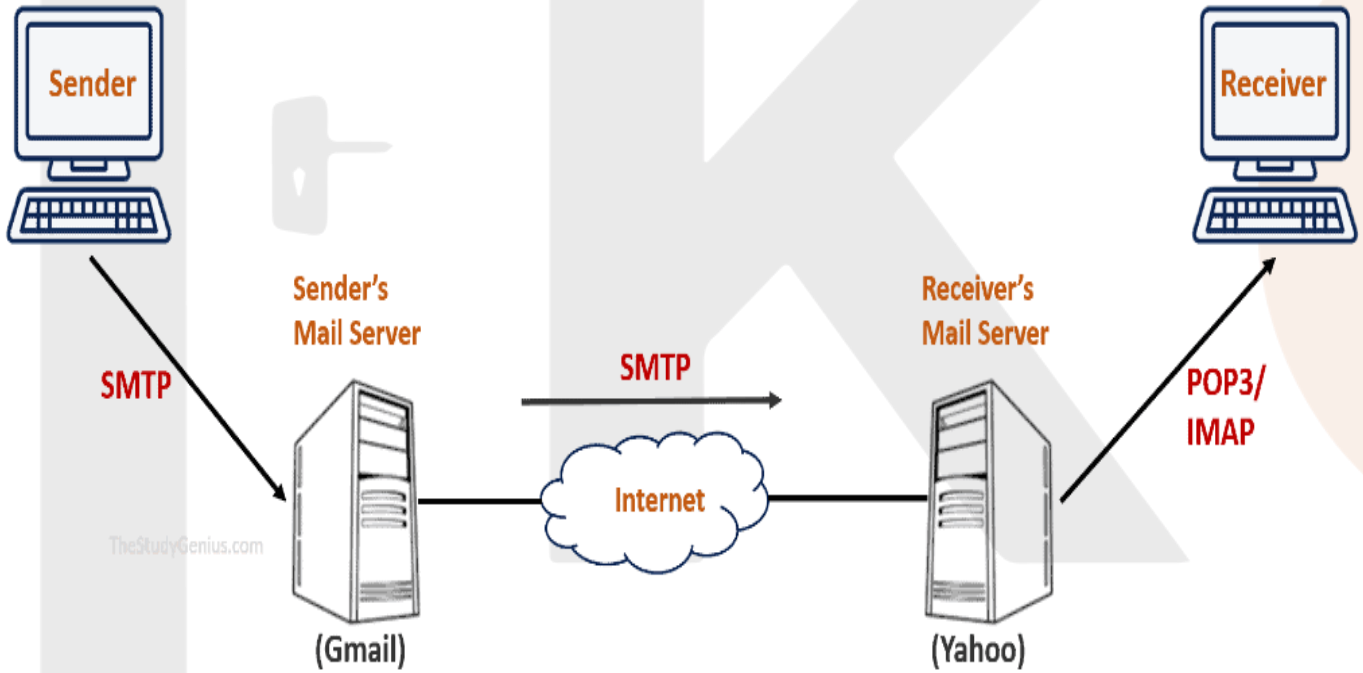
- Today, electronic mail is much more complex. It allows a message to include text, audio, and video. It also allows one message to be sent to one or more recipients.

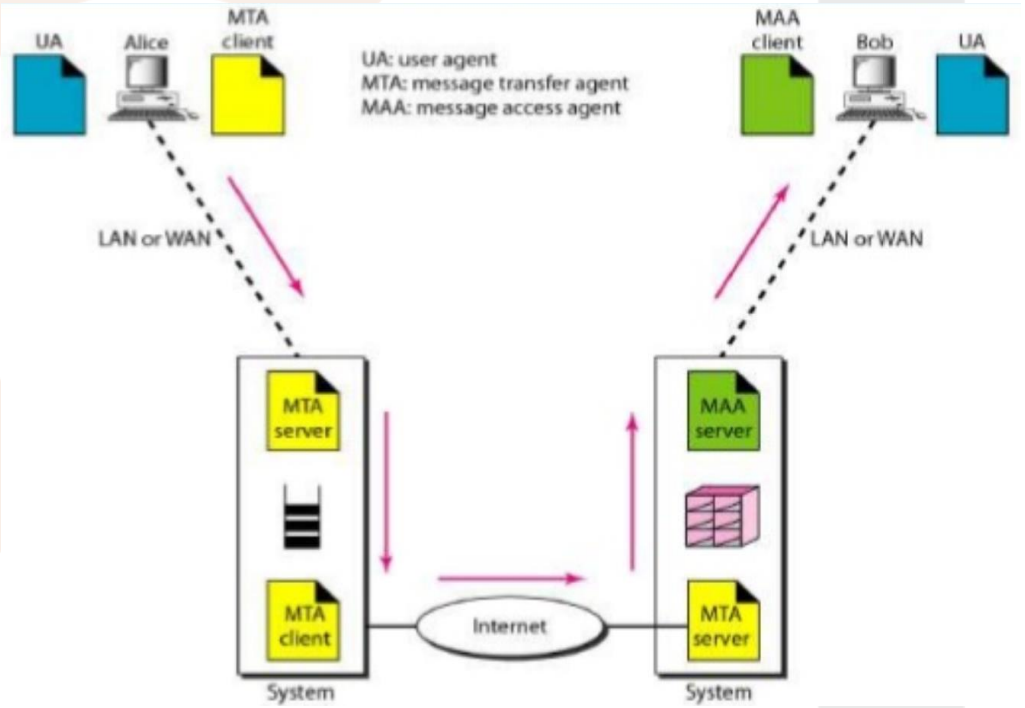


Message Transfer Agent: SMTP

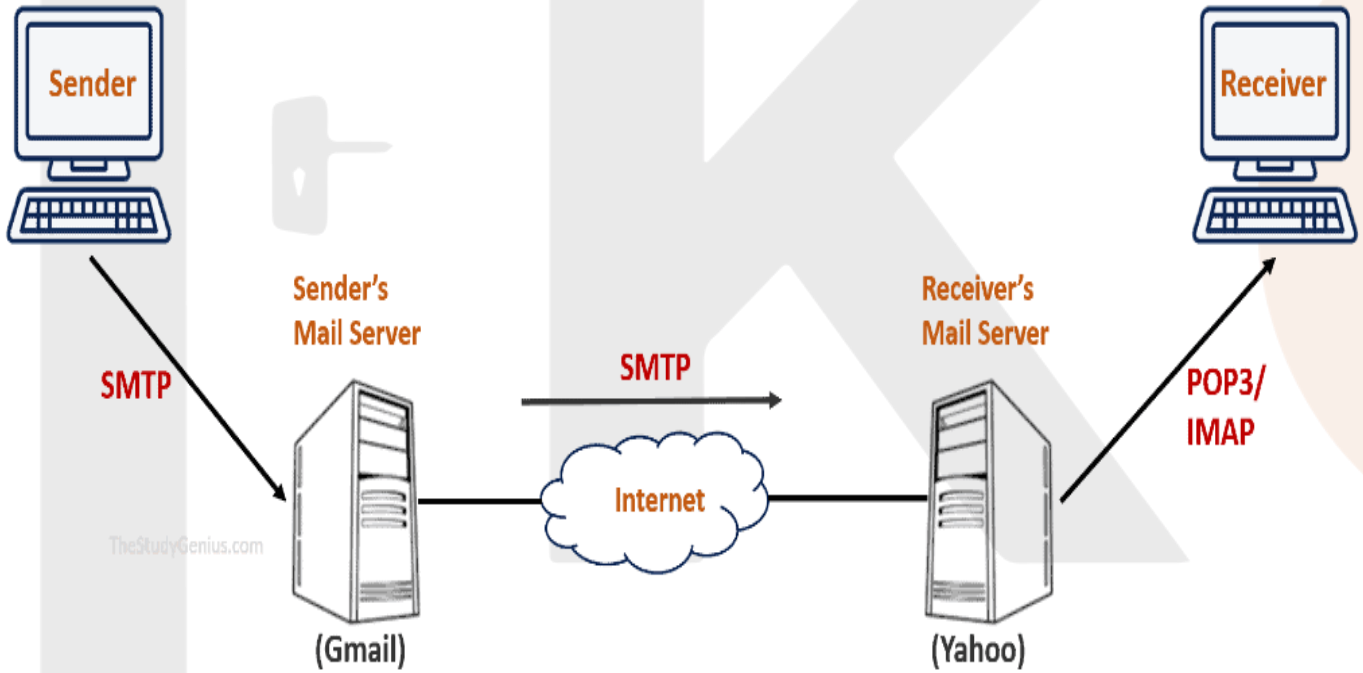
- The actual mail transfer is done through message transfer agents. To send mail, a system must have the client MTA, and to receive mail, a system must have a server MTA.

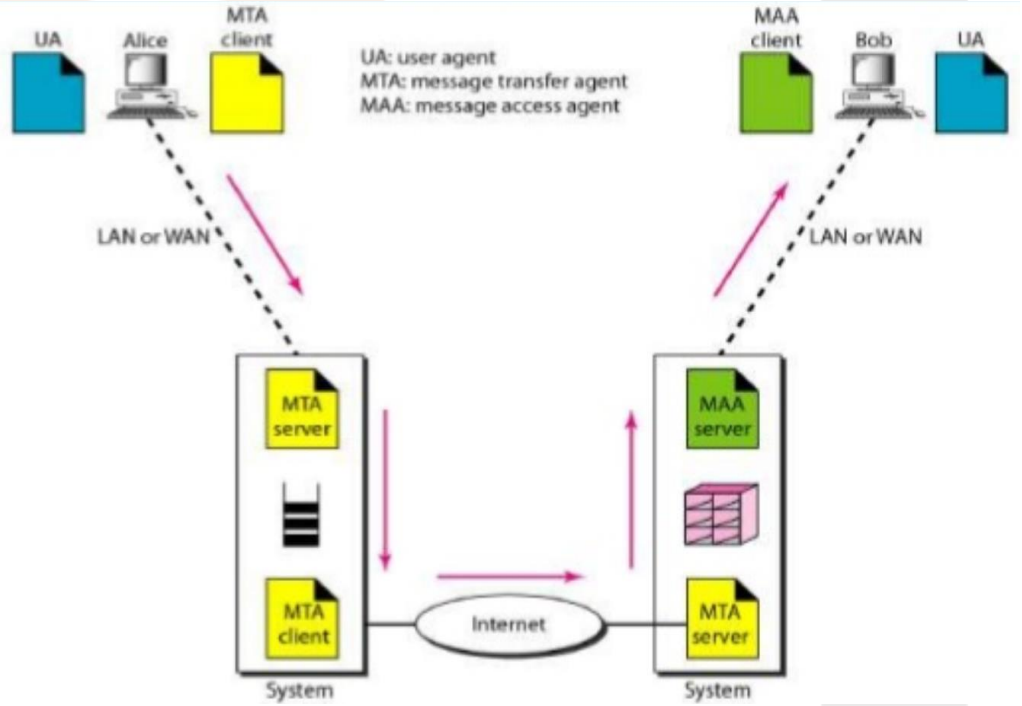
- The formal protocol that defines the MTA client and server in the Internet is called the Simple Mail Transfer Protocol (SMTP).





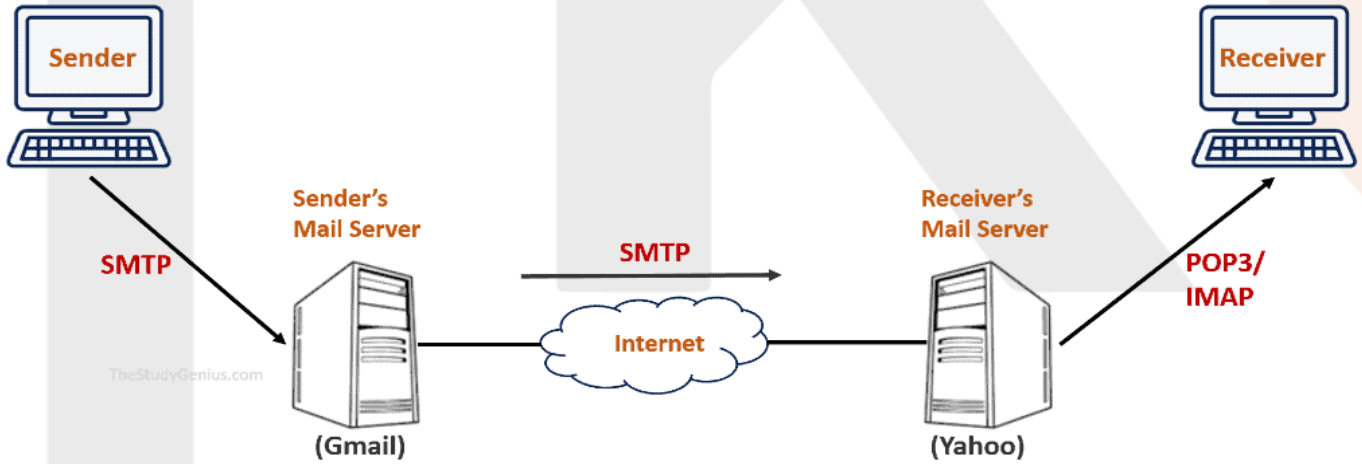
- SMTP is used two times, between the sender and the sender's mail server and between the two mail servers
- SMTP simply defines how commands and responses must be sent back and forth.
- Mail Transfer Phases: - The process of transferring a mail message occurs in three phases: connection establishment, mail transfer, and connection termination.





Message Access Agent: POP and IMAP

- The first and the second stages of mail delivery use SMTP. However, SMTP is not involved in the third stage because SMTP is a push protocol; it pushes the message from the client to the server.
- On the other hand, the third stage needs a pull protocol; the client must pull messages from the server.
- The third stage uses a message access agent. Currently two message access protocols are available: Post Office Protocol, version 3 (POP_3) and Internet Mail Access Protocol, version 4 ($IMAP_4$).



POP3

- Post Office Protocol, version 3 (POP3) is simple and limited in functionality. The client POP3 software is installed on the recipient computer; the server POP3 software is installed on the mail server.

- Mail access starts with the client when the user needs to download e-mail from the mailbox on the mail server. The client opens a connection to the server on TCP port 110. It then sends its user name and password to access the mailbox.

- The user can then list and retrieve the mail messages, one by one. POP3 has two modes: the delete mode and the keep mode. In the delete mode, the mail is deleted from the mailbox after each retrieval. In the keep mode, the mail remains in the mailbox after retrieval.

- The delete mode is normally used when the user is working at her permanent computer and can save and organize the received mail after reading or replying.

- The keep mode is normally used when the user accesses her mail away from her primary computer (e.g., a laptop). The mail is read but kept in the system for later retrieval and organizing.

IMAP4

- Another mail access protocol is Internet Mail Access Protocol, version 4 (IMAP4). IMAP4 is similar to POP3, but it has more features; IMAP4 is more powerful and more complex.

- POP3 is deficient in several ways. It does not allow the user to organize her mail on the server; the user cannot have different folders on the server. (Of course, the user can create folders on her own computer.)

- In addition, POP3 does not allow the user to partially check the contents of the mail before downloading. IMAP4 provides the following extra functions:

    - A user can check the e-mail header prior to downloading.
    - A user can search the contents of the e-mail for a specific string of characters prior to downloading.
    - A user can partially download e-mail. This is especially useful if bandwidth is limited and the e-mail contains multimedia with high bandwidth requirements.
    - A user can create, delete, or rename mailboxes on the mail server.
    - A user can create a hierarchy of mail boxes in a folder for e-mail storage.

MIME

- Electronic mail has a simple structure. Its simplicity, however, comes at a price. It can send messages only in NVT 7-bit ASCII format. In other words, it has some limitations. For example, it cannot be used for languages that are not supported by 7-bit ASCII characters (such as Hindi, French, German, Hebrew, Russian, Chinese, and Japanese).

- Also, it cannot be used to send binary files or video or audio data.

- Multipurpose Internet Mail Extensions (MIME) is a supplementary protocol that allows non-ASCII data to be sent through e-mail.

- MIME transforms non-ASCII data at the sender site to NVT ASCII data and delivers them to the client MTA to be sent through the Internet. The message at the receiving side is transformed back to the original data.

Web-Based Mail

- E-mail is such a common application that some websites today provide this service to anyone who accesses the site. Two common sites are Hotmail and Yahoo.
- The idea is very simple. Mail transfer from Alice's browser to her mail server is done through HTTP.
- The transfer of the message from the sending mail server to the receiving mail server is still through SMTP.
- Finally, the message from the receiving server (the Web server) to Bob's browser is done through HTTP.
- The last phase is very interesting. Instead of POP3 or IMAP4, HTTP is normally used. When Bob needs to retrieve his e-mails, he sends a message to the website (Hotmail, for example).
- The website sends a form to be filled in by Bob, which includes the log-in name and the password. If the log-in name and password match, the e-mail is transferred from the Web server to Bob's browser in HTML format.

Q Which of the following protocol pairs can be used to send and retrieve e-mails (in that order)? **(Gate-2019) (1 Marks)**

(a) SMTP, MIME

(b) SMTP, POP3

(c) IMAP, POP3

(d) IMAP, SMTP

Q Which of the following transport layer protocols is used to support electronic mail? **(Gate-2012) (1 Marks)**

(A) SMTP **(B)** IP **(C)** TCP **(D)** UDP

Q Consider different activities related to email:

m1: Send an email from a mail client to a mail server

m2: Download an email from mailbox server to a mail client

m3: Checking email in a web browser

Which is the application level protocol used in each activity? **(GATE-2011) (1 Marks)**

(A) m1: HTTP m2: SMTP m3: POP

(B) m1: SMTP m2: FTP m3: HTTP

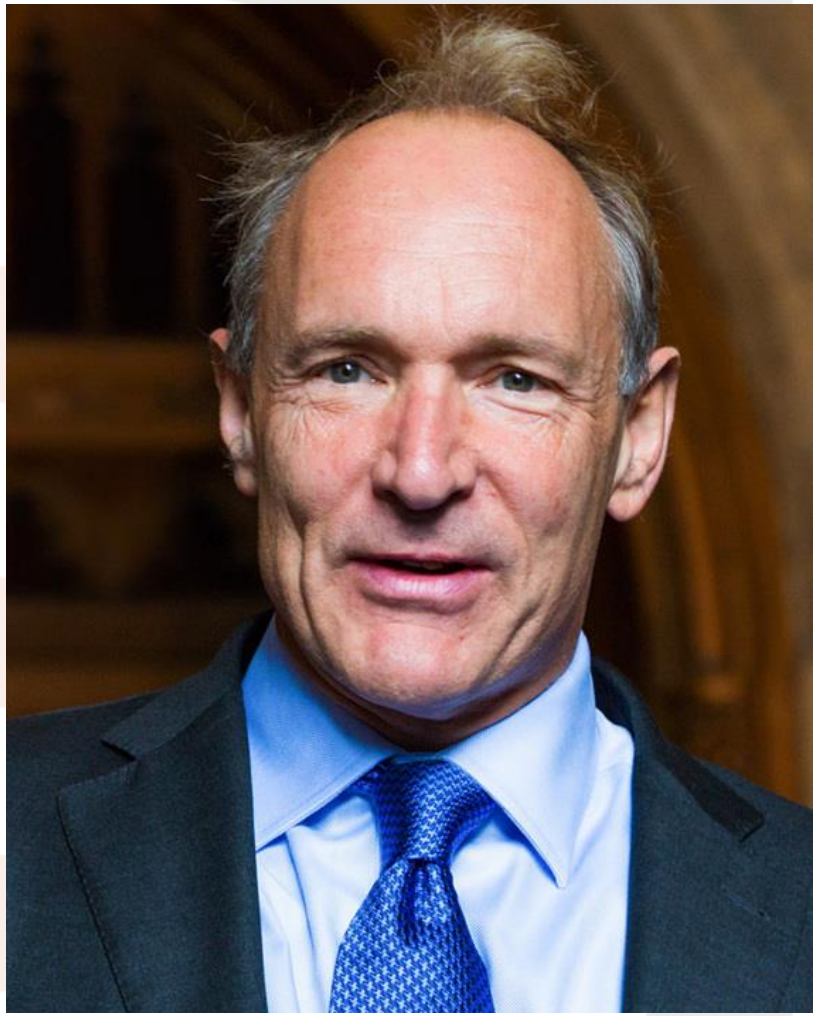
(C) m1: SMTP m2: POP m3: HTTP

(D) m1: POP m2: SMTP m3: IMAP

Break

WWW

- The idea of the Web was first proposed by Tim Berners-Lee in 1989 at *CERN*
- The World Wide Web (WWW) is a repository of information linked together from points all over the world.
- The WWW has a unique combination of flexibility, portability, and user-friendly features that distinguish it from other services provided by the Internet.
- The WWW project was initiated by CERN (European Laboratory for Particle Physics) to create a system to handle distributed resources necessary for scientific research.



- **ARCHITECTURE**

  - The WWW today is a distributed client server service, in which a client using a browser can access a service using a server. However, the service provided is distributed over many locations called sites.

  - Each site holds one or more documents, referred to as Web pages. Each Web page can contain a link to other pages in the same site or at other sites. The pages can be retrieved and viewed by using browsers.

- **Client (Browser)**
  - A variety of vendors offer commercial browsers that interpret and display a Web document, and all use nearly the same architecture.
  - Each browser usually consists of three parts: a controller, client protocol, and interpreters.
  - The controller receives input from the keyboard or the mouse and uses the client programs to access the document. After the document has been accessed, the controller uses one of the interpreters to display the document on the screen.




- **Server**
  - The Web page is stored at the server. Each time a client request arrives, the corresponding document is sent to the client.
  - To improve efficiency, servers normally store requested files in a cache in memory; memory is faster to access than disk.
  - A server can also become more efficient through multithreading or multiprocessing. In this case, a server can answer more than one request at a time.
  - A client that wants to access a Web page needs the address. To facilitate the access of documents distributed throughout the world, HTTP uses locators.

- ***Uniform Resource Locator (URL)***
    - A web page, as a file, needs to have a unique identifier to distinguish it from other web pages.
    - To define a web page, we need four identifiers in general: ***Protocol***, ***host, port, and path.***
    - ***Protocol*.** Which client-server application we are using is called protocol. Although most of the time the protocol is HTTP (Hyper Text Transfer Protocol), we can also use other protocols such as FTP (File Transfer Protocol).
    - ***Host.*** The host identifier can be the IP address of the server or the unique name to the server.
    - ***Port*.** The port, a 16-bit integer, is normally predefined for the client-server application.

Break

Cookies

- The World Wide Web was originally designed as a stateless entity. A client sends a request; a server responds. Their relationship is over. The original design of WWW, retrieving publicly available documents, exactly fits this purpose. Today the Web has other functions; some are listed here.
    - Some websites need to allow access to registered clients only.
    - Websites are being used as electronic stores that allow users to browse through the store, select wanted items, put them in an electronic cart, and pay at the end with a credit card.
    - Some websites are used as portals: the user selects the Web pages he wants to see.
    - Some websites are just advertising.

- For these purposes, the cookie mechanism was devised.
  - Creation and Storage of Cookies: - The creation and storage of cookies depend on the implementation; however, the principle is the same.
    - When a server receives a request from a client, it stores information about the client in a file or a string. The information may include the domain name of the client, the contents of the cookie (information the server has gathered about the client such as name, registration number, and so on), a timestamp, and other information depending on the implementation.
    - The server includes the cookie in the response that it sends to the client.
    - When the client receives the response, the browser stores the cookie in the cookie directory, which is sorted by the domain server name.
  - Using Cookies: - When a client sends a request to a server, the browser looks in the cookie directory to see if it can find a cookie sent by that server. If found, the cookie is included in the request.

- When the server receives the request, it knows that this is an old client, not a new one. Note that the contents of the cookie are never read by the browser or disclosed to the user. It is a cookie made by the server and eaten by the server.

Break

HTTP

- The Hypertext Transfer Protocol (HTTP) is a protocol used mainly to access data on the World Wide Web. The **Hyper Text Transfer Protocol (HTTP)** is used to define how the client-server programs can be written to retrieve web pages from the Web.
- HTTP uses the services of TCP on well-known port 80, the client uses a temporary port number.
- It is a connection-oriented and reliable protocol.
- HTTP functions as a combination of FTP and SMTP.

- It is similar to FTP because it transfers files and uses the services of TCP. However, it is much simpler than FTP because it uses only one TCP connection. There is no separate control connection; only data are transferred between the client and the server.

- HTTP is like SMTP because the data transferred between the client and the server look like SMTP messages.

- Unlike SMTP, the HTTP messages are not destined to be read by humans; they are read and interpreted by the HTTP server and HTTP client (browser).

- SMTP messages are stored and forwarded, but HTTP messages are delivered immediately.

- The commands from the client to the server are embedded in a request message. The contents of the requested file or other information are embedded in a response message.

Proxy Server

- HTTP supports proxy servers. A proxy server is a computer that keeps copies of responses to recent requests.
- The HTTP client sends a request to the proxy server. The proxy server checks its cache. If the response is not stored in the cache, the proxy server sends the request to the corresponding server.
- Incoming responses are sent to the proxy server and stored for future requests from other clients.
- The proxy server reduces the load on the original server, decreases traffic, and improves latency. However, to use the proxy server, the client must be configured to access the proxy instead of the target server.

Nonpersistent versus Persistent Connections

- ***Nonpersistent Connections***: - In a **nonpersistent connection**, one TCP connection is made for each request/response. The following lists the steps in this strategy:
  - The client opens a TCP connection and sends a request.
  - The server sends the response and closes the connection.
  - The client reads the data until it encounters an end-of-file marker; it then closes the connection.
  - **For example:** If a file contains links to N different pictures in different files (all located on the same server), the connection must be opened and closed N + 1 times.

- **Disadvantage** The nonpersistent strategy imposes high overhead on the server because the server needs N + 1 different buffer each time a connection is opened.

Persistent Connections

- HTTP version 1.1 specifies a **persistent connection** by default.
- In a persistent connection, the server leaves the connection open for more requests after sending a response.
- The server can close the connection at the request of a client or if a time-out has been reached.

- **Advantages**
  - Time and resources are saved using persistent connections.
  - Only one set of buffers and variables needs to be set for the connection at each site.
  - The round-trip time for connection establishment and connection termination is saved.

- ***It is important to know that HTTP is a stateless protocol as:***
  - HTTP server does not maintain any state. It forgets about the client after sending the response.
  - It treats every new request independently.

- ***HTTP Security***
  - HTTP per se does not provide security.
  - HTTP can be run over the Secure Socket Layer (SSL). In this case, HTTP is referred to as HTTPS.
  - HTTPS provides confidentiality, client and server authentication, and data integrity.

Q Identify the correct sequence in which the following packets are transmitted on the network by a host when a browser requests a webpage from a remote server, assuming that the host has just been restarted. **(GATE-2016) (2 Marks)**

(A) HTTP GET request, DNS query, TCP SYN

(B) DNS query, HTTP GET request, TCP SYN

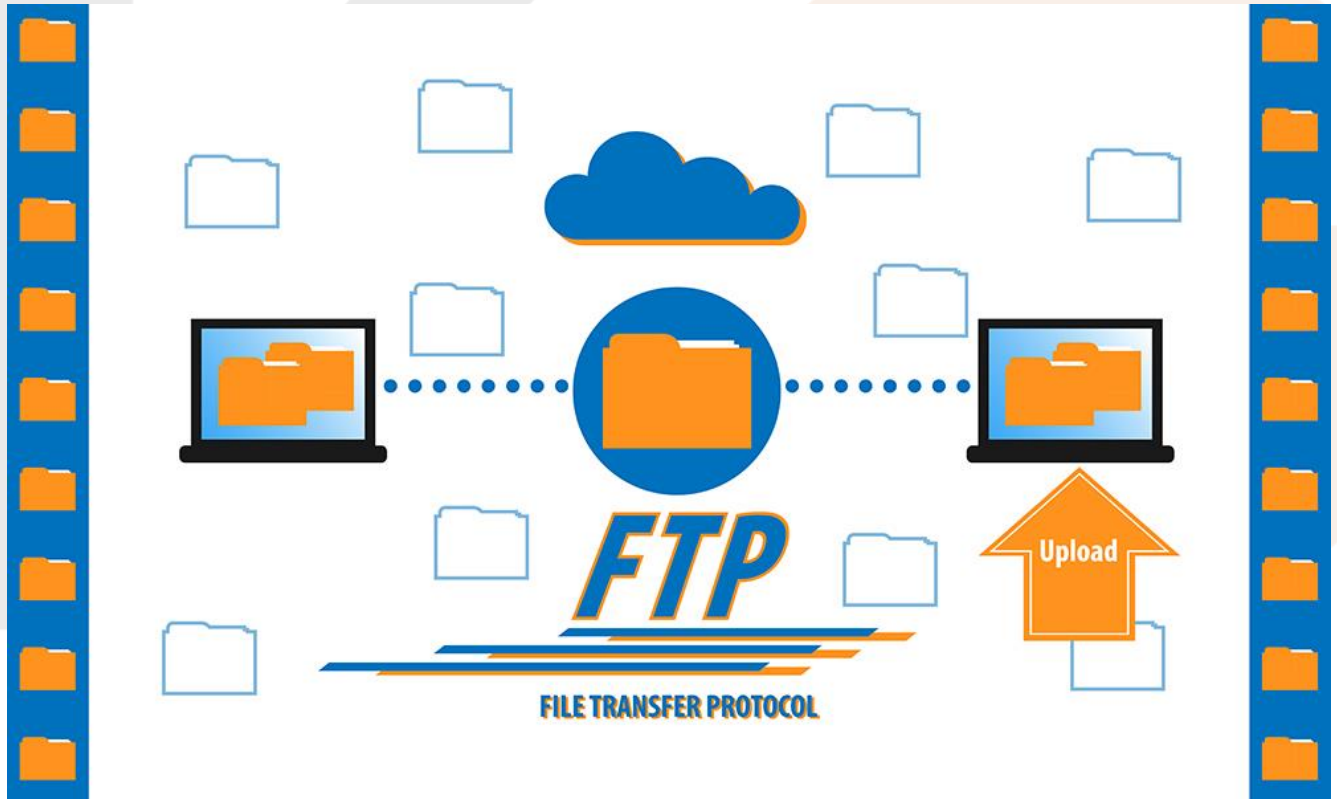
(C) DNS query, TCP SYN, HTTP GET request

(D) TCP SYN, DNS query, HTTP GET request

Break

FILE TRANSFER

- Transferring files from one computer to another is one of the most common tasks expected from a networking or internetworking environment. As a matter of fact, the greatest volume of data exchange in the Internet today is due to file transfer.



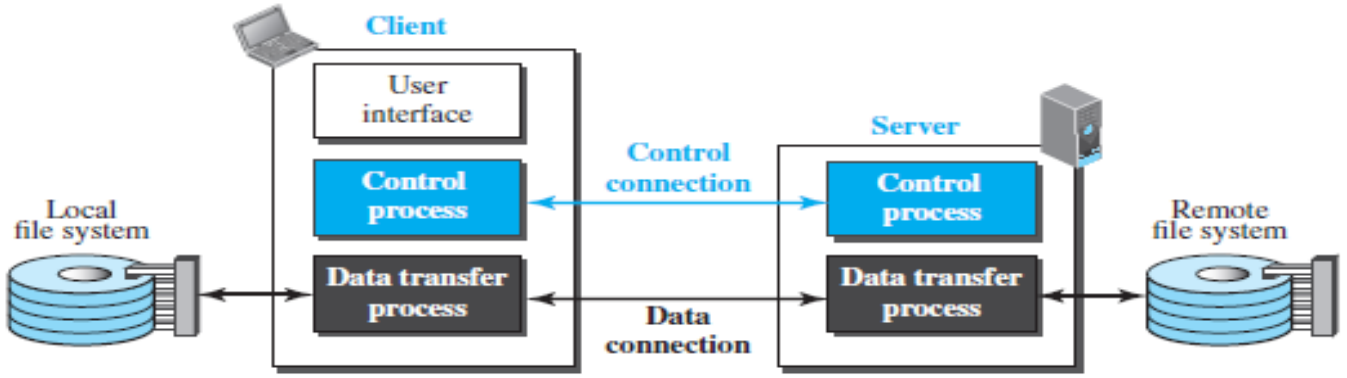
File Transfer Protocol (FTP)

- It is used for exchanging files over the internet and enables the users to upload and download the files from the internet.

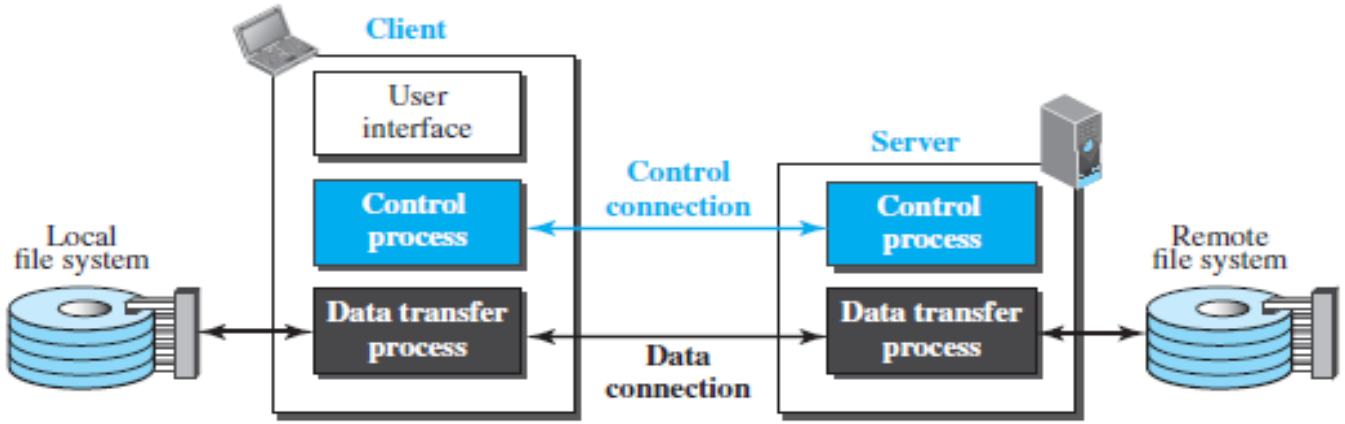
- File Transfer Protocol (FTP) is the standard mechanism provided by TCP/IP for copying a file from one host to another. Although transferring files from one system to another seems simple and straightforward, some problems must be dealt with first. For example, two systems may use different file name conventions.

- Two systems may have different ways to represent text and data. Two systems may have different directory structures.

- All these problems have been solved by FTP in a very simple and elegant approach. FTP differs from other client/server applications in that it establishes two connections between the hosts.

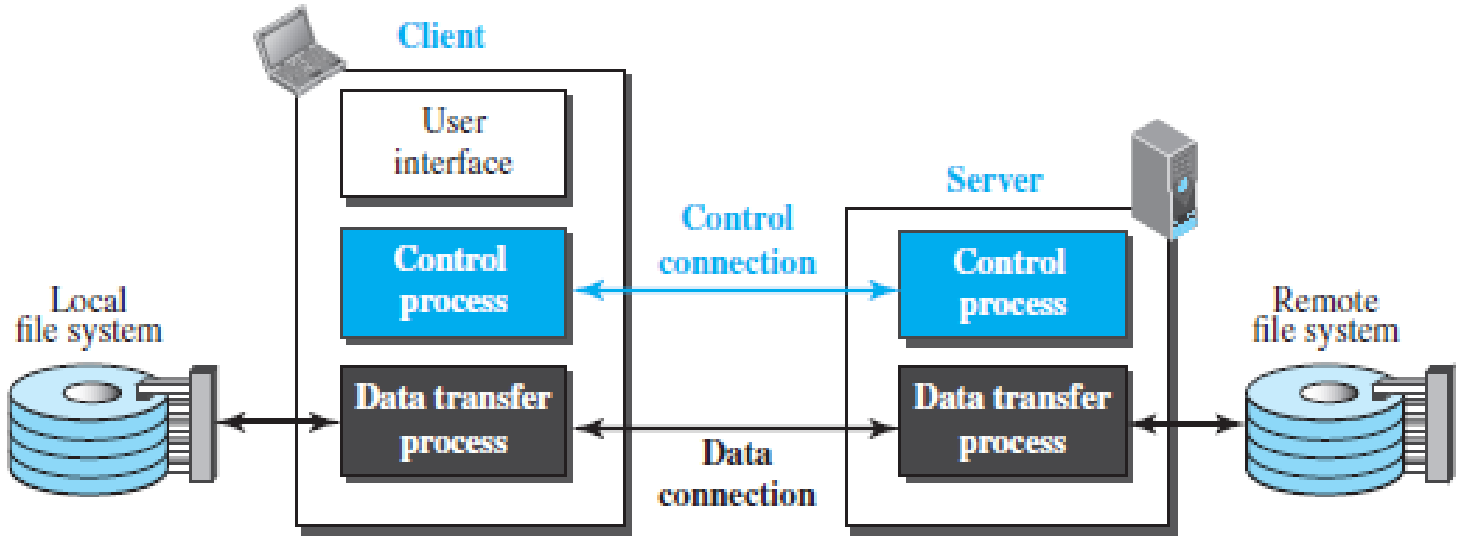


- One connection is used for data transfer, the other for control information (commands and responses). Separation of commands and data transfer makes FTP more efficient.
- The control connection uses very simple rules of communication. We need to transfer only a line of command or a line of response at a time.
- The data connection, on the other hand, needs more complex rules due to the variety of data types transferred. However, the difference in complexity is at the FTP level, not TCP. For TCP, both connections are treated the same.
- FTP uses two well-known TCP ports: Port 21 is used for the control connection, and port 20 is used for the data connection.
- The control connection remains connected during the entire interactive FTP session.
- The data connection is opened and then closed for each file transfer activity.



Security for FTP

- The FTP protocol was designed when security was not a big issue.
- Although FTP requires a password, the password is sent in plaintext (unencrypted), which means it can be intercepted and used by an attacker.
- The data transfer connection also transfers data in plaintext, which is insecure.
- To be secure, one can add a Secure Socket Layer between the FTP application layer and the TCP layer. In this case FTP is called SSL-FTP.



Q In one of the pairs of protocols given below, both the protocols can use multiple TCP connections between the same client and the server. Which one is that?

(GATE-2015) (1 Marks)

(A) HTTP, FTP

(B) HTTP, TELNET

(C) FTP, SMTP

(D) HTTP, SMTP

Q The transport layer protocols used for real time multimedia, file transfer, DNS and email, respectively are: **(GATE-2013) (1 Marks)**

(A) TCP, UDP, UDP and TCP

(B) UDP, TCP, TCP and UDP

(C) UDP, TCP, UDP and TCP

(D) TCP, UDP, TCP and UDP

Q Consider the following clauses:

i. Not inherently suitable for client authentication.

ii. Not a state sensitive protocol.

iii. Must be operated with more than one server.

iv. Suitable for structured message organization.

v. May need two ports on the serve side for proper operation.

The option that has the maximum number of correct matches is **(GATE-2007) (2 Marks)**

(A) IMAP-(i), FTP-(ii), HTTP-(iii), DNS-(iv), POP3-(v)

(B) FTP-(i), POP3-(ii), SMTP-(iii), HTTP-(iv), IMAP-(v)

(C) POP3-(i), SMTP-(ii), DNS-(iii), IMAP-(iv), HTTP-(v)

(D) SMTP-(i), HTTP-(ii), IMAP-(iii), DNS-(iv), FTP-(v)

Q Which one of the following uses UDP as the transport protocol? **(GATE-2007) (1 Marks)**

(A) HTTP **(B)** Telnet **(C)** DNS **(D)** SMTP

Q Match the following: **(GATE-2007) (2 Marks)**

(P) SMTP	(1) Application layer
(Q) BGP	(2) Transport layer
(R) TCP	(3) Data link layer
(S) PPP	(4) Network layer
	(5) Physical layer

(A) P – 2 Q – 1 R – 3 S – 5

(B) P – 1 Q – 4 R – 2 S – 3

(C) P – 1 Q – 4 R – 2 S – 5

(D) P – 2 Q – 4 R – 1 S – 3

Q which of the following is not a client -server application? **(GATE-2010) (1 Marks)**

a) Internet chat

b) Web browsing

c) E-mail

d) Ping

Q The protocol data unit (PDU) for the application layer in the Internet stack is **(GATE-2012) (1 Marks)**

(A) Segment

(B) Datagram

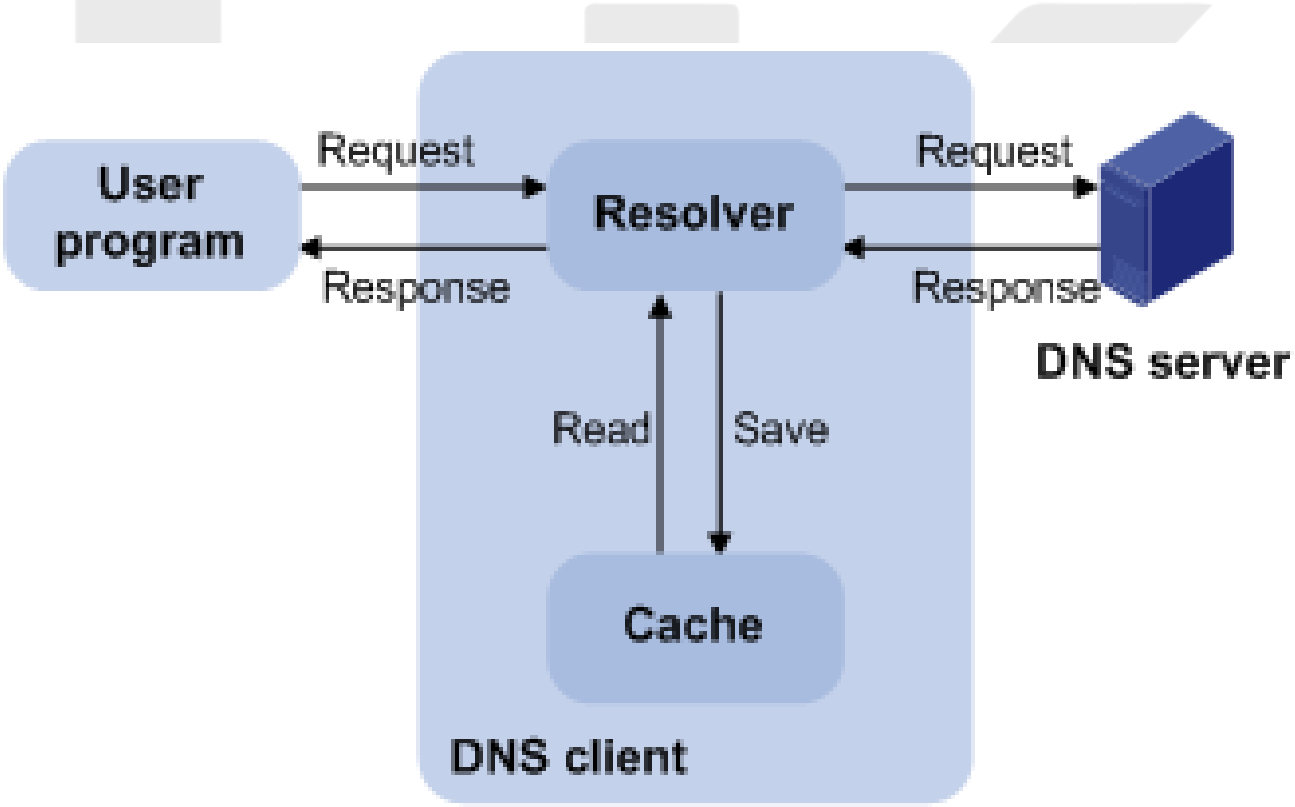
(C) Message

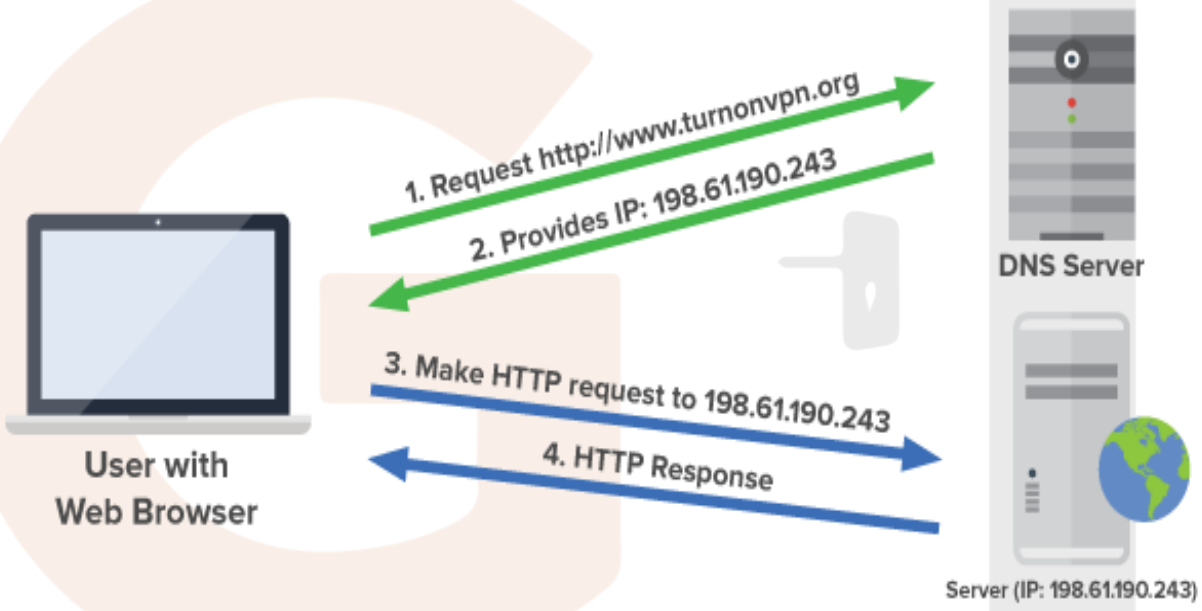
(D) Frame

DNS

- As we know human beings are not comfortable in remembering numbers so to remember IP address of a website or mail account in internet is difficult.
- Secondly IP addresses of mail or websites keeps on changing, so we have to come up with one more level of addressing which is easy to remember and do not change with time.
- Solution is Name addressing, i.e. we give some names to websites and mail account like we do to humans in real world.
- But then if someone write a name of the website in the browser we need some mechanism to convert it back into IP address.
- Domain Name System solve this problem.

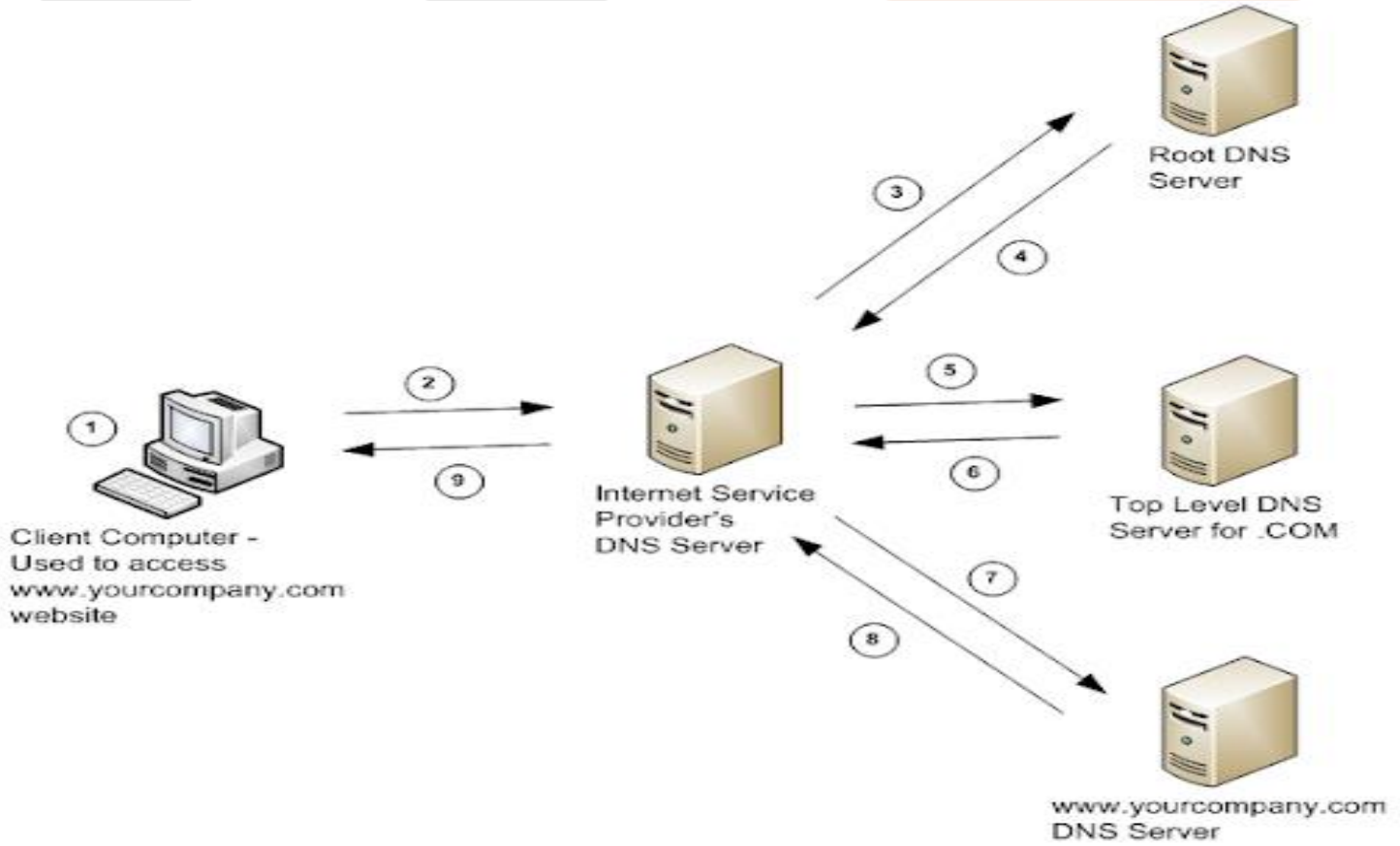
- This diagram perfectly represent how DNS works, A user of a website may know the name of the website; however, the IP protocol needs the IP address.
- The DNS client program sends a request to a DNS server to map the Web-site address to the corresponding IP address.



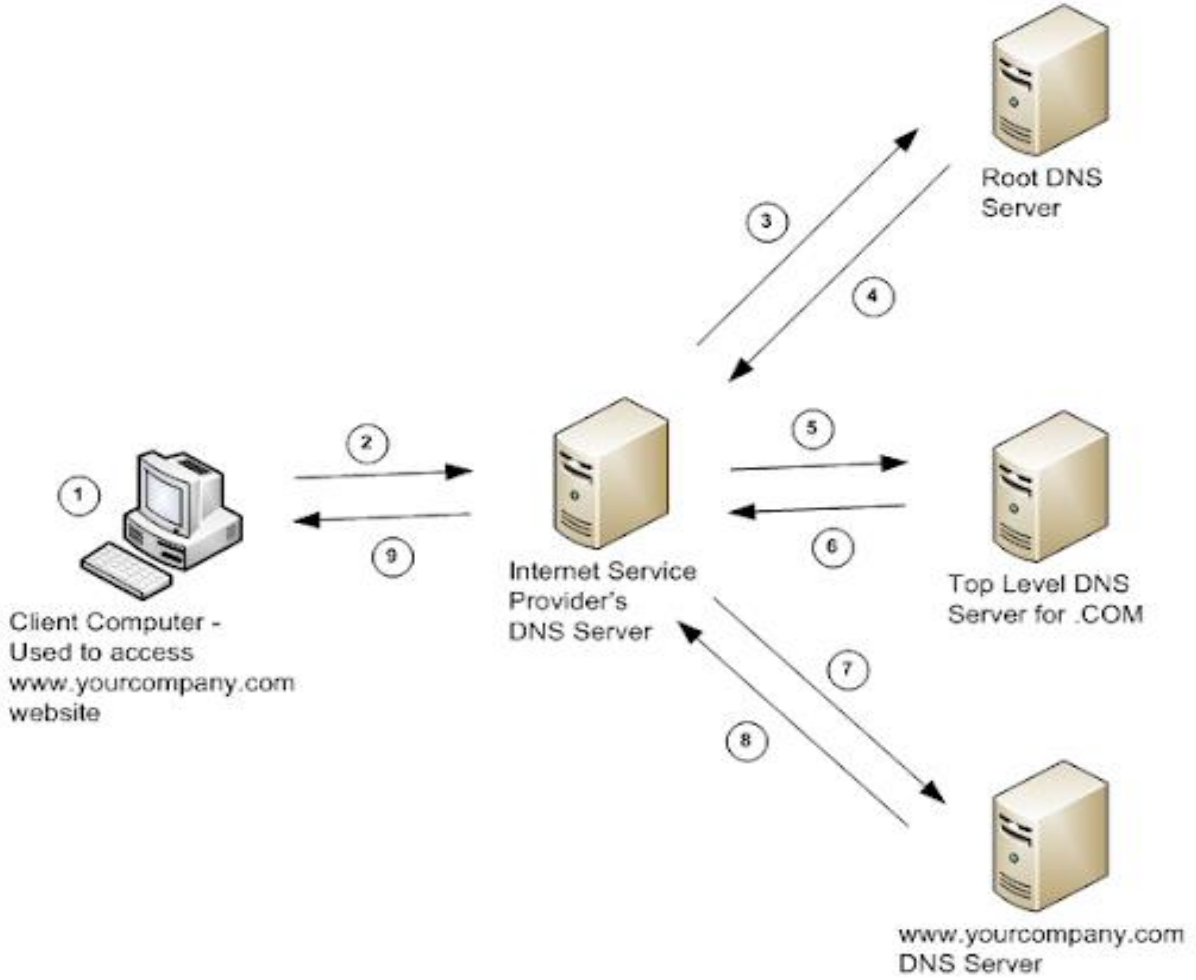


- today we divide this huge amount of information into smaller parts and store each part on a different computer. In this method, the host that needs mapping can contact the closest computer holding the needed information.

- It is Very difficult to find out the ip address associated to 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 delay for that to happen organization of database is very important.



- **Hierarchy of Name Servers**
  - **Root name servers** – It is contacted by name servers that cannot resolve the name. It contacts authoritative name server if name mapping is not known. It then gets the mapping and return the IP address to the host.
  - **Top level 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 names and IP addresses of each authoritative name server for the second level domains.
  - **Authoritative name servers** This is organization's DNS server, providing authoritative hostName to IP mapping for organization servers. It can be maintained by 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 authoritative domain name server which actually contains the IP address. So, the authoritative domain server will return the associative ip address.



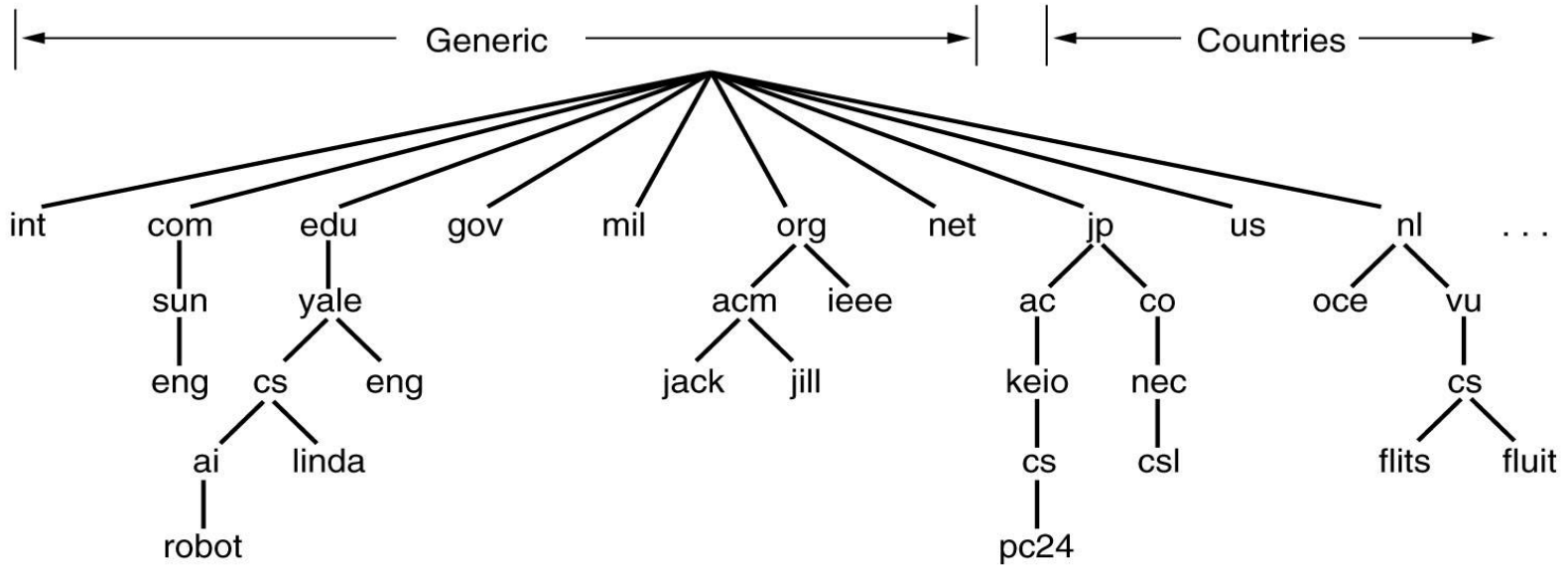
NAME SPACE

- To be unambiguous, the names must be unique because the addresses are unique. A name space that maps each address to a unique name can be organized in two ways: flat or hierarchical.

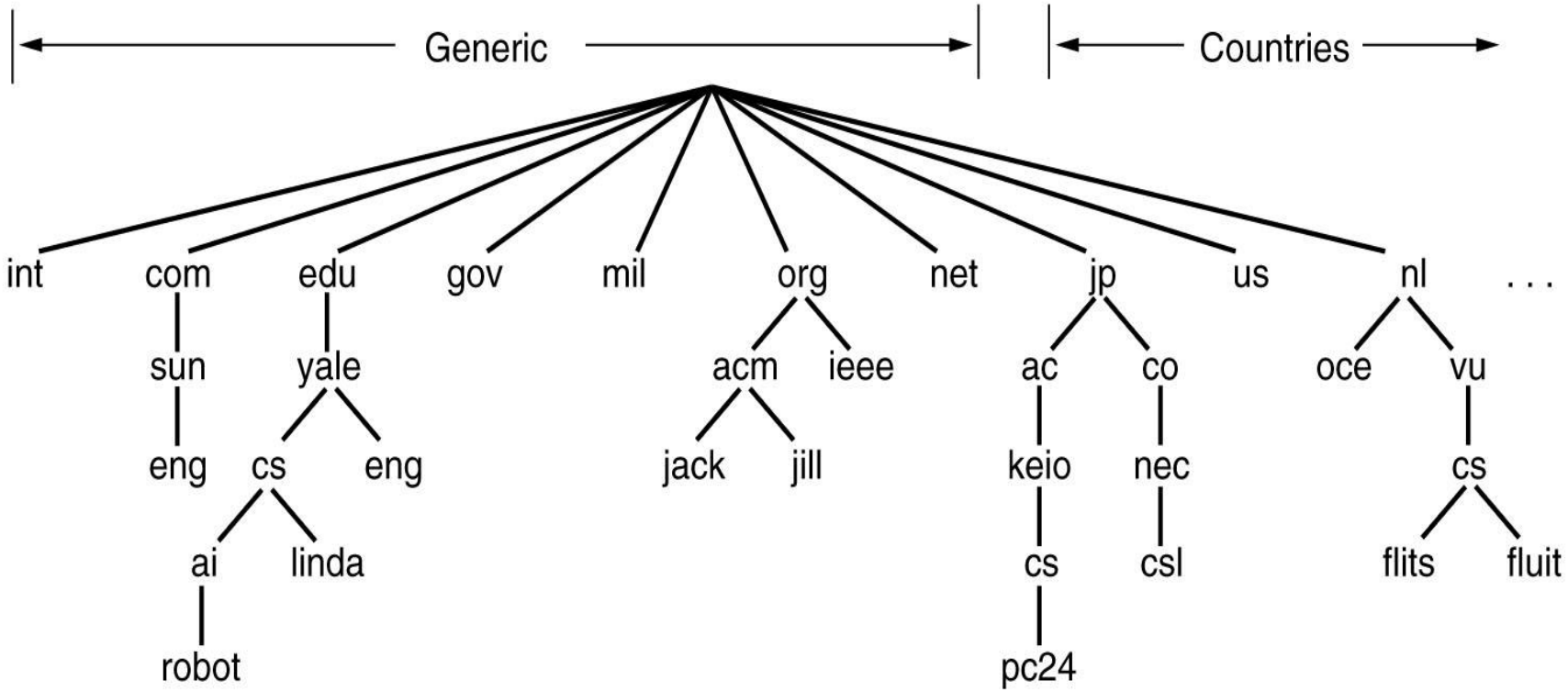
- **Flat Name Space**
  - In a flat name space, a name is assigned to an address. A name in this space is a sequence of characters without structure.
  - The main disadvantage of a flat name space is that it cannot be used in a large system such as the Internet because it must be centrally controlled to avoid ambiguity and duplication.
  - So, Solution is Hierarchical Name Space

Hierarchical Name Space

- In a hierarchical name space, each name is made of several parts. The first part can define the nature of the organization
- the second part can define the name of an organization, the third part can define departments in the organization, and so on.
- In this case, the authority to assign and control the name spaces can be decentralized. A central authority can assign the part of the name that defines the nature of the organization and the name of the organization.
- The responsibility of the rest of the name can be given to the organization itself.
- The management of the organization need not worry that the prefix chosen for a host is taken by another organization because, even if part of an address is the same, the whole address is different

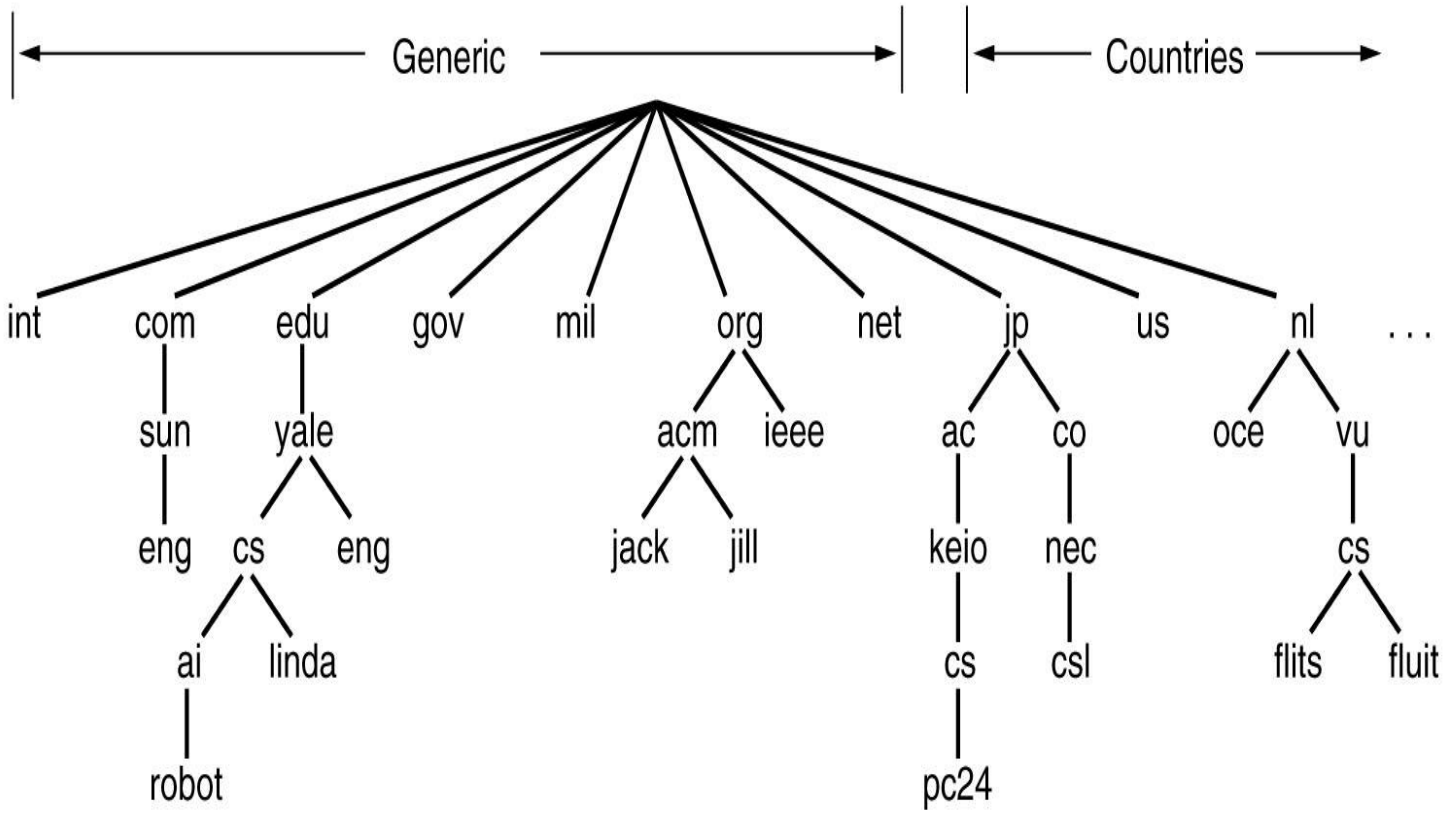


- Generic domain: .com(commercial) .edu(educational) .mil(military) .org (non-profit organization) .net (similar to commercial) all these are generic domain.
- Country domain .in (india) .us .uk
- Inverse domain if we want to know what is the domain name of the website. Ip to domain name mapping. So, DNS can provide both the mapping for example to find the ip addresses of www.unacademy.com then we have to type nslookup www.unacademy.com



DOMAIN NAME SPACE

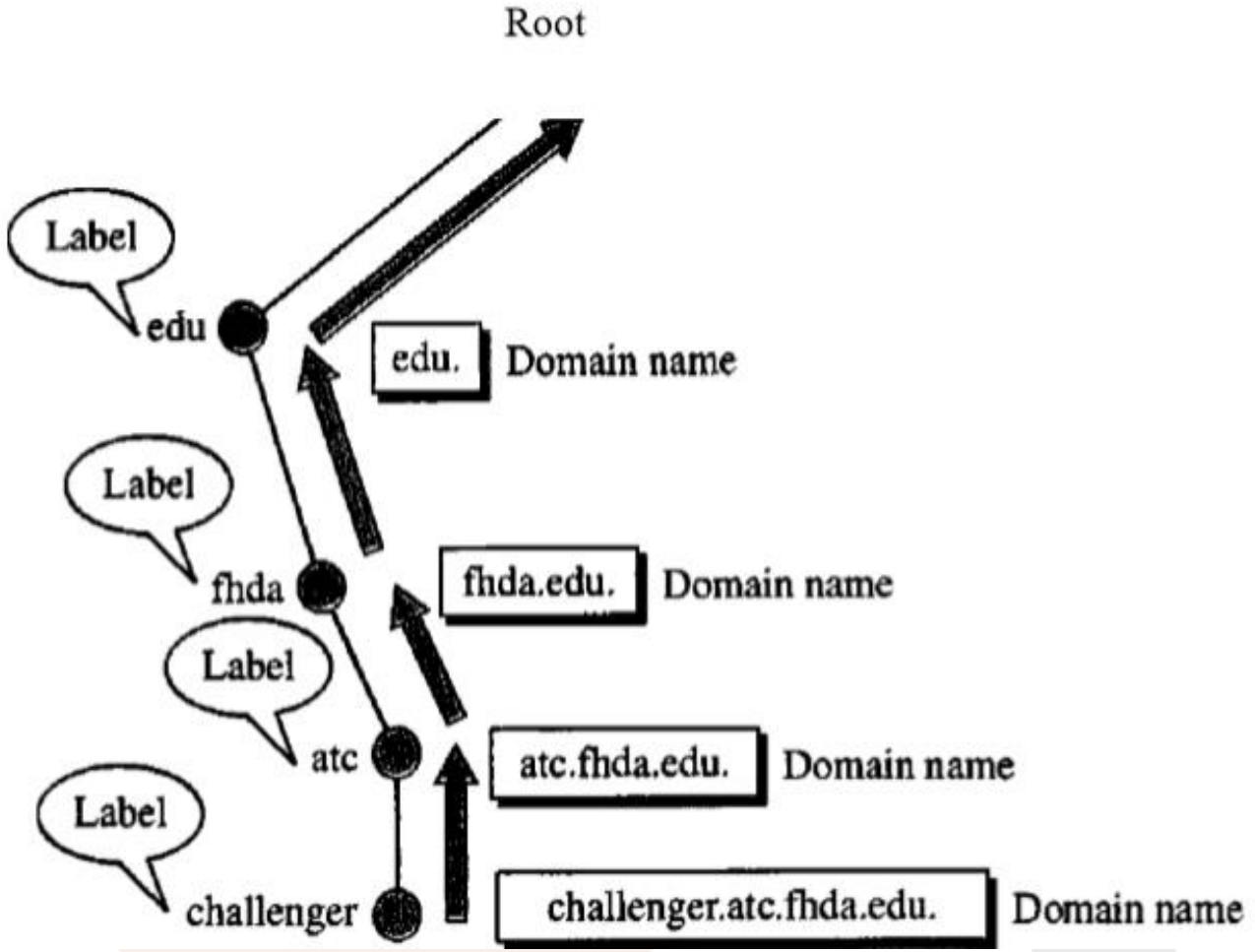
- To have a hierarchical name space, a domain name space was designed. In this design the names are defined in an inverted-tree structure with the root at the top. The tree can have only 128 levels: level 0 (root) to level 127



Label	*Description*
aero	Airlines and aerospace companies
biz	Businesses or firms (similar to "com")
com	Commercial organizations
coop	Cooperative business organizations
edu	Educational institutions
gov	Government institutions
info	Information service providers
int	International organizations
mil	Military groups
museum	Museums and other nonprofit organizations
name	Personal names (individuals)
net	Network support centers
org	Nonprofit organizations
pro	Professional individual organizations

- **Label**
  - Each node in the tree has a label, which is a string with a maximum of 63 characters.
  - The root label is a null string (empty string).
  - DNS requires that children of a node (nodes that branch from the same node) have different labels, which guarantees the uniqueness of the domain names.

- **Domain Name**
  - Each node in the tree has a domain name. A full domain name is a sequence of labels separated by dots (.).
  - The domain names are always read from the node up to the root. The last label is the label of the root (null). This means that a full domain name always ends in a null label, which means the last character is a dot because the null string is nothing.



Break

TELNET

- **Telnet** is an application protocol used on the Internet or local area network to provide a bidirectional interactive text-oriented communication facility using a virtual terminal connection. Telnet was developed in 1969.

- TELNET is an abbreviation for Terminal Network. It is the standard TCP/IP protocol for virtual terminal service as proposed by the International Organization for Standards (ISO). TELNET enables the establishment of a connection to a remote system in such a way that the local terminal appears to be a terminal at the remote system.

- TELNET was designed at a time when most operating systems, such as UNIX, were operating in a timesharing environment. In such an environment, a large computer supports multiple users. The interaction between a user and the computer occurs through a terminal, which is usually a combination of keyboard, monitor, and mouse.