Course Title: **Web Technology**

Subject Code: **CO 427**

Syllabus

S.No.	Contents
1.	**Inter-Networking:** Internet, Growth of Internet, Owners of the Internet, Anatomy of Internet, APRANET and Internet history of the World Web, Basic Internet Terminology, Net etiquette. Working of Internet: Packet switching technology, Internet Protocols: TCP/IP, Router. Internet Addressing Scheme: Machine Addressing (IP address), E-mail Address, Resource Addresses.
2.	**Internet Applications:** E-mail, file transfer (FTP), telnet, usenet, Internet chat, Web.
3.	**Evolution of Web:**Web 1.0: Hypertext & linking documents, HTTP, Client-Server, peer-to-peer; Web Browser (Lynx, Mosiac, Netscape, Internet Explorer, Firefox, and Safari, the mobile web); Impact: Opportunities & Challenges. Web 2.0: From 1.0 to 2.0; Framework; Technologies: Client-side & server-side; Web 2.0 development technologies; Examples: social networking sites, blogs, wikis, video sharing sites, hosted services(web services, location-based services), web applications, mashups & folksonomies; Practical Usage. Web 3.0: From 2.0 to 3.0; Semantic Web: What, How, Why; From Web 3.0 to Web 4.0
4.	**Web Development:** Phases; Web Page, Website, and Web Application: Example, Technology Framework for development.cClient-side technology: HTML (HTML 5).Client-side scripting: JavaScript. Server-side technology: PHP.Server-side scripting: Server-side JavaScript.Web application development frameworks: Django& Ruby on Rails. Web Database: Database Connectivity: JDBC, ODBC; Database-to-web connectivity.
6.	**Web Search and Mining:**Web IR System: Search Engines, Web Crawling, Search Engine Optimization, Web Analytics, Web Mining Taxonomy; Web Mining Framework; Social Web Mining.Text Mining: Opinion Mining, Recommendation System, Topic Detection and Tracking.

Web Technology

- **Web technology** refers to the means by which computers communicate with each other using **markup languages** and **multimedia packages**. It gives us a way to interact with hosted information, like websites.

- Web technology involves the use of **hypertext markup language (HTML) and cascading style sheets (CSS)**.

If you want to be a web developer you should be able to explain various types of web technologies....

Types of Web technologies

- **Browsers**
  - **Google Chrome** - Currently, the most popular browser brought to you by Google
  - **Safari** - Apple's web browser
  - **Firefox** - Open-source browser supported by the Mozilla Foundation
  - **Internet Explorer** - Microsoft's browser.
- **HTML**
- **CSS**

Programming Languages

- **Javascript** - used by all web browsers, Meteor, and lots of other frameworks
- **Coffeescript** - is a kind of “dialect” of javascript. It is viewed as simpler and easier on your eyes as a developer but it complies (converts) back into javascript
- **Python** -used by the Django framework and used in a lot of mathematical calculations
- **Ruby** - used by the Ruby on Rails framework
- **PHP** - used by Wordpress
- **Go** - newer language, built for speed.
- **Objective-C** - the programming language behind iOS (your iPhone), lead by Apple
- **Swift** - Apple’s newest programming language
- **Java** - Used by Android (Google) and a lot of desktop applications.

Frameworks

- **Node.js** - a server-side javascript framework
- **Ruby on Rails** - a full-stack framework built using ruby
- **Django** - a full-stack framework built using python
- **Ionic** - a mobile framework
- **Phonegap / Cordova** - a mobile framework that exposes native api's of iOS and Android for use when writing javascript
- **Bootstrap** - a UI (user interface) framework for building with HTML/CSS/Javascript
- **Foundation** - a UI framework for building with HTML/CSS/Javascript
- **Wordpress** - a CMS (content management system) built on PHP. Currently, about 20% of all websites run on this framework
- **Drupal** - a CMS framework built using PHP.
- **.NET** - a full-stack framework built by Microsoft
- **Angular.js** - a front-end javascript framework.
- **Ember.js** - a front-end javascript framework.
- **Backbone.js** - a front-end javascript framework.

Libraries

- jQuery
- Underscore

Databases

- **MongoDB** - is an open-sourced NoSQL database and is currently the only database supported by Meteor.
- **Redis** - is the most popular key-value store. It is lighting fast for retrieving data but doesn't allow for much depth in the data storage.
- **PostgreSQL** - is a popular open-sourced SQL database.
- **MySQL** - is another popular open-sourced SQL database. MySQL is used in Wordpress websites.
- **Oracle** - is an enterprise SQL database.
- **SQL Server** - is an SQL server manager created by Microsoft.

Client (or Client-side) & Server (or Server-side)

- **Server-side** is the systems that run on the **server**, and **client-side** is the software that runs on a user's web browser. **Client-side** web development involves interactivity and displaying data, **server-side** is about working behind the scenes to manage data.

Front-end & Back-end

- The term "**front-end**" refers to the user interface, while "**back-end**" means the server, application and database that work behind the scenes to deliver information to the user

API

- An Application Programming Interface (**API**) is a tool set that programmers can use in helping them create software. ... An **example** is the Apple (iOS) **API** that's used to detect touchscreen interactions. APIs are tools. They allow you as a programmer to deliver solid solutions fairly rapidly.

Protocols

- **HTTP** - This protocol is how each website gets to your browser. Whenever you type a website like “http://google.com” this protocol requests the website from google’s server and then receives a response with the HTML, CSS, and javascript of the website.

- **DDP** - is a new protocol created in connection with Meteor. The DDP protocol uses websockets to create a consistent connection between the client and the server. This constant connection lets websites and data on those websites update in real-time without refreshing your browser.

- **REST** - is a protocol mainly used for API’s. It has standard methods like GET, POST, and PUT that let information be exchanged between applications.

Data Formats

- **JSON** - is quickly becoming the most popular data format
- **XML** - was the main data format early in the web days and predominantly used by Microsoft systems
- **CSV** - is data formatted by commas. Excel data is typically formatted this way.

Inter-Networking

Inter-Net

- The Internet is the global system of interconnected computer networks that uses the Internet protocol suite (TCP/IP) to communicate between networks and devices.

- An electronic communications network that connects computer networks and organizational computer facilities around the world.

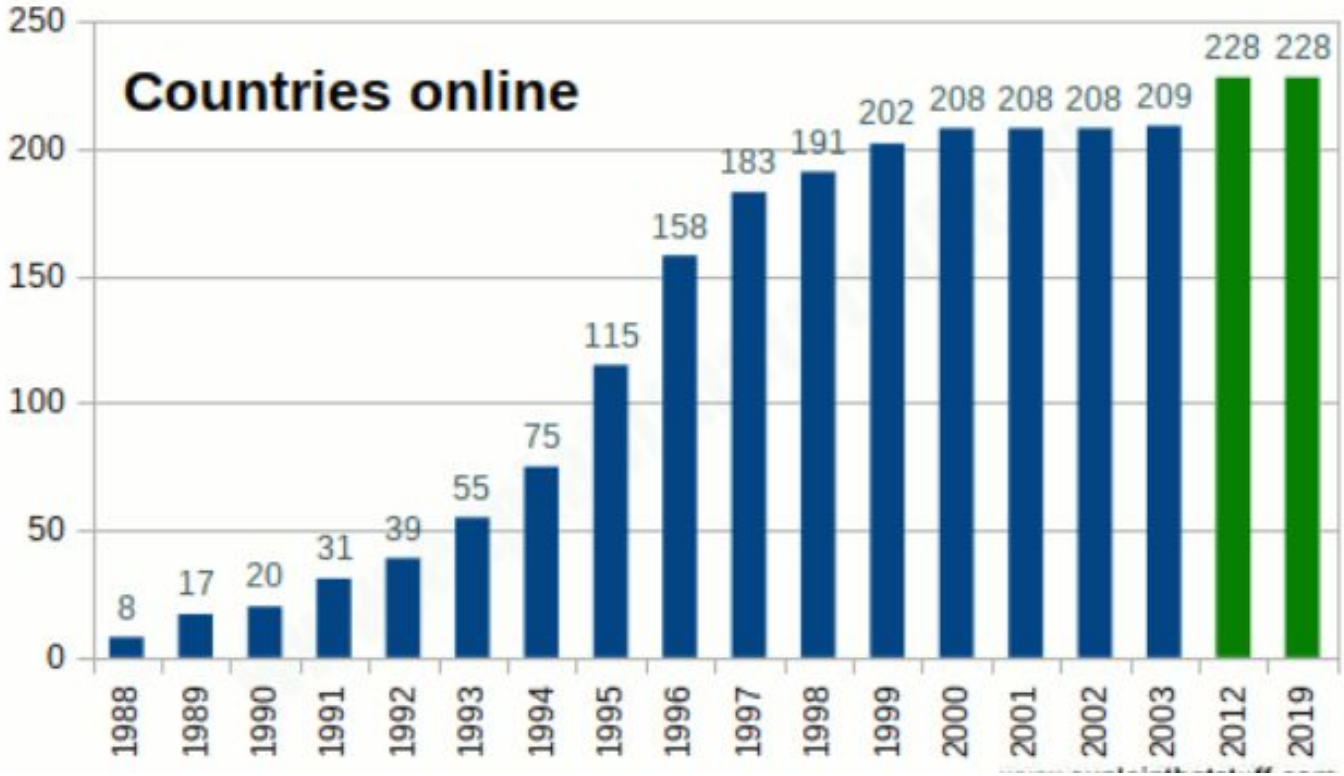
ARPANET

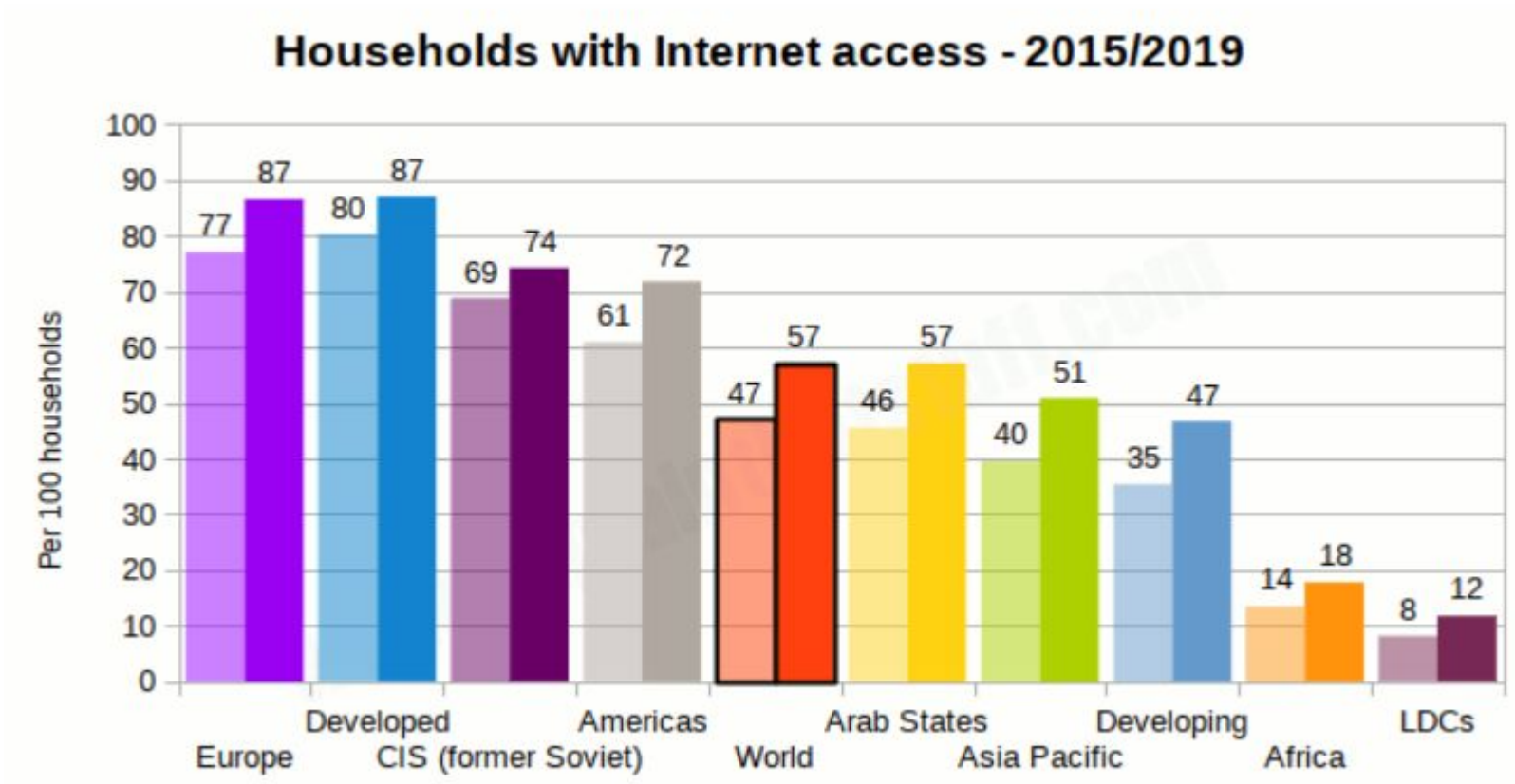
- The Internet, sometimes called simply "the Net," is a worldwide system of computer networks -- a network of networks in which users at any one computer can, if they have permission, get information from any other computer (and sometimes talk directly to users at other computers).

- It was conceived by the **Advanced Research Projects Agency (ARPA)** of the U.S. government in 1969 and was first known as the ARPANet. The original aim was to create a network that would allow users of a research computer at one university to "talk to" research computers at other universities.

- A side benefit of ARPANet's design was that, because messages could be routed or rerouted in more than one direction, the network could continue to function even if parts of it were destroyed.

Internet history of the world web




Households with Internet access - 2015/2019
Per 100 households
100
90
80
70
60
50
40
30
20
10
0
77
87
80
87
69
74
61
72
47
57
46
57
40
51
35
47
14
18
8
12
Europe
Developed
CIS (former Soviet)
Americas
World
Arab States
Asia Pacific
Developing
Africa
LDCs

- **What does the Internet do?**

- **How does Internet data move?**
  - Circuit Switching
  - Packet Switching

- **What are "clients" and "servers”**
  - Peer to Peer Communication (P2P)

- **Routers**

A brief history of the Internet

Precursors

- **1844:** Samuel Morse transmits the **first electric telegraph message**, eventually making it possible for people to send messages around the world in a matter of minutes.

- **1876:** Alexander Graham Bell (and various rivals) **develop the telephone**.

- **1940:** George Stibitz **accesses a computer in New York using a teletype** (remote terminal) in New Hampshire, connected over a telephone line.

- **1945:** Vannevar Bush, a US government scientist, publishes a paper called As We May Think, anticipating the **development of the World Wide Web** by half a century.

- **1958: Modern modems are developed** at Bell Labs. Within a few years, AT&T and Bell begin selling them commercially for use on the public telephone system.

1960s: Preparing for a global network

- **1964:** Paul Baran, a researcher at RAND, **invents the basic concept of computers communicating by sending "message blocks"** (small packets of data); Welsh physicist Donald Davies has a very similar idea and coins the name "packet switching," which sticks.

- **1963:** J.C.R. Licklider envisages a network that **can link people** and user-friendly **computers together**.

- **1964:** Larry Roberts, a US computer scientist, experiments with **connecting computers over long distances.**

- **1960s:** Ted Nelson **invents hypertext**, a way of linking together separate documents that eventually becomes **a key part of the World Wide Web**.

- **1966:** Inspired by the work of Licklider, Bob Taylor of the US government's Advanced Research Projects Agency (ARPA) hires Larry Roberts **to begin developing a national computer network.**

- **1969:** The **ARPANET computer network is launched**, initially linking together four scientific institutions in California and Utah.

1970s: The modern Internet appears

- **1971:** Ray Tomlinson **sends the first email**, introducing the @ sign as a way of separating a user's name from the name of the computer where their mail is stored.

- **1973:** Bob Metcalfe **invents Ethernet**, a convenient way of linking computers and peripherals (things like printers) on a local network.

- **1974:** Vinton Cerf and Bob Kahn write an influential paper describing how computers linked on a network they called an "internet" could **send messages via packet switching**, using a protocol (set of formal rules) called TCP (Transmission Control Protocol).

- **1978:** TCP is improved by adding the concept of computer addresses (Internet Protocol or IP addresses) to which Internet traffic can be routed. This lays the **foundation of TCP/IP**, the basis of the modern Internet.

- **1978:** Ward Christensen sets **up Computerized Bulletin Board System** (a forerunner of topic-based Internet forums, groups, and chat rooms) so computer hobbyists can swap information.

1980s: The Internet gives birth to the Web

- **1983: TCP/IP is officially adopted** as the standard way in which Internet computers will communicate.

- **1982–1984: DNS (Domain Name System)** is developed, allowing people to refer to unfriendly IP addresses (12.34.56.78) with friendly and memorable names (like google.com).

- **1986:** The US National Science Foundation **(NSF) creates its own network,** NSFnet, allowing universities to piggyback onto the ARPANET's growing infrastructure.

- **1988:** Finnish computer scientist Jarkko Oikarinen **invents IRC (Internet Relay Chat),** which allows people to create "rooms" where they can talk about topics in real-time with like-minded online friends.

- **1989:** The Peapod grocery store pioneers online grocery shopping and e-commerce.

1990s: The Web takes off

- **1993:** Marc Andreessen writes **Mosaic, the first user-friendly web browser**, which later evolves into **Netscape and Mozilla.**

- **1993:** Oliver McBryan **develops the World Wide Web Worm**, one of the first search engines.

- **1995: E-commerce properly begins** when Jeff Bezos founds **Amazon.com and Pierre Omidyar sets up eBay.**

- **1996:** ICQ becomes the first user-friendly **instant messaging (IM) system** on the Internet.

- **1997:** Jorn Barger **publishes the first blog** (web-log).

- 1998: Larry Page and Sergey Brin **develop a search engine called BackRub** that they quickly decide to rename **Google.**

- **1999:** Kevin Ashton conceives the idea that everyday objects, and not just computers, could be part of the Internet. This idea is now known as the **Internet of Things.**

- **2000s: Internet and Web for all**

- **2003:** Virtually every country in the world is now **connected to the Internet.**

- **2004:** Harvard student **Mark Zuckerberg revolutionizes social networking with Facebook**, an easy-to-use website that connects people with their friends.

- **2006:** Jack Dorsey and Evan Williams **found Twitter**, an even simpler "microblogging" site where people share their thoughts and observations in off-the-cuff, 140-character status messages.

- **2017:** Russian president Vladimir Putin **approves a plan to create a private alternative to the Internet** to counter the historic dominance of the (traditional) Internet by the United States.

Internet Terminology

- **World Wide Web (WWW):** The World Wide Web ("WWW" or simply the "web") is a collection of electronic documents (called web pages) that are linked together like a spider web. These documents are stored on computers called servers located around the world.

- **Web Server:** A Web Server is a computer that stores web pages. It is responsible for accepting request(s) from users and serves them with web pages. Two important web server programs are: **IIS (Internet Information server) and Apache**, etc. Web servers are connected to the Internet 24 hours a day, seven days a week.

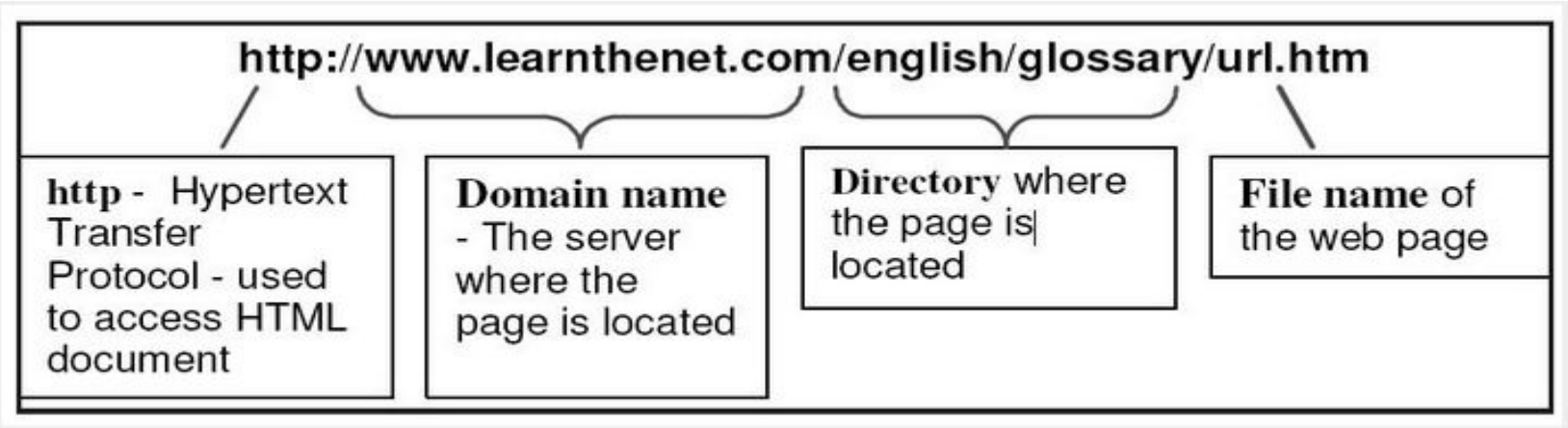
- **Hyperlink:** It is an element in an electronic document that links to another place in the same document or to an entirely different document or other resource.
- Hyperlinks usually appear as underlined text and in a different color, but they may also appear as graphics, such as buttons to click.
- Hyperlinks may be used to link another place in the same page, or another page, to play an audio or video file, to download a file, to set up a message to an e-mail address, and to link to other Internet resources.

- **HTML (Hypertext Markup Language):** It is a language that consists of certain key words called ‘Tags’, used for writing the documents on the web.

- **Web Page:** A web page is an **electronic document** written in a computer language called HTML (Hypertext Markup Language).
- Web pages can contain text, graphics, video, animation, and sound, as well as interactive features, such as data entry forms.
- Each page has a **unique address known as a URL** (Uniform Resource Locator) that identifies its location on the server. Web pages usually contain hyperlinks to other web pages.

- **Website:** A website (often shortened to just site) is one or more web pages, belonging to a particular company, institute, government or an individual.
- The first page is called the **home page**, which acts like an index, indicating the content on the site.
- By default the home page is named as **index.html**. From the home page, you can click hyperlinks to access other web pages.

- **URL (Uniform Resource Locator):** Every page on the web has a unique address, called Uniform Resource Locator, URL. A URL indicates where the **web page is stored** on the Internet. A sample URL might look like the following:



- **IP (Internet Protocol) Address:** Computers do not understand letters or symbols that humans use to communicate effectively. Computers understand numbers-specifically, 1s and 0s.
- Thus every host (a computer linked to the Internet) on the Internet has a unique host number. This number is called the Internet Protocol address, or IP address.
- The IP address is a unique address, generally written in the format nnn.nnn.nnn.nnn, where nnn represents a 3 digit number that varies between 0 and 255. For Example: 192.100.8.56

- **DNS (Domain Name System):** Every host (computer linked to Internet) has a unique host number called IP address. We can connect to any host through IP address only, but it is difficult to remember the 4-digit number of hosts.

- To resolve this, **domain name** is the only solution. Domain name, a unique name of the individual host computer on the Internet.
- Every computer on the Internet now have both a domain name and an IP address. To connect to any host through domain name requires some mechanism that will convert the domain name IP address.
- DNS, Domain Name System is the standard for resolving names to addresses. It is used mostly to translate between domain names and IP addresses.

Netiquette

- Make sure identification is clear in all communications.
- Review what you wrote and try to interpret it objectively.
- Don’t assume everyone understands where you’re coming from.
- Respect others’ privacy.
- Remember, if it’s on the internet, it’s everywhere.
- Follow the rules.
- Forgive and forget.

How Does the Internet Work

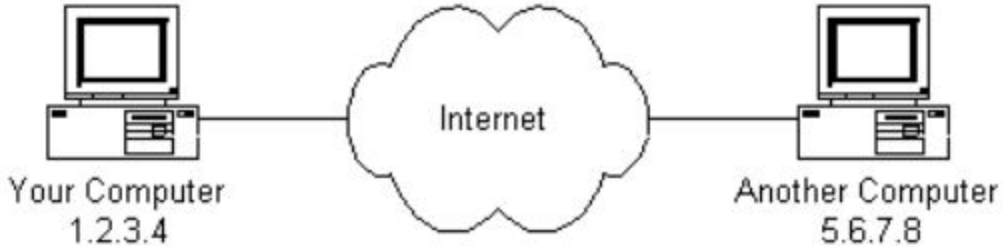
Internet addresses

- Internet is a global network of computers each computer connected to the Internet must have a **unique address.**

- Internet addresses are in the form **nnn.nnn.nnn.nnn** where **nnn** must be a number from **0 - 255**. This address is known as an **IP address**. (**IP stands for Internet Protocol**)

- The diagram in the next slide illustrates two computers connected to the Internet; one computer with IP address 1.2.3.4 and another computer with IP address 5.6.7.8.

- The Internet is represented as an abstract object in-between.



- If you connect to the Internet through an **Internet Service Provider (ISP),** you are usually assigned a **temporary IP address** for the duration of your dial-in session.

- If you connect to the Internet from a **local area network (LAN)** your computer might have a **permanent IP address** or it might obtain a temporary one from a **DHCP (Dynamic Host Configuration Protocol)** server.

- In both case, if you are connected to the Internet, your computer has a unique IP address.

Ping Program

Protocol Stacks and Packets

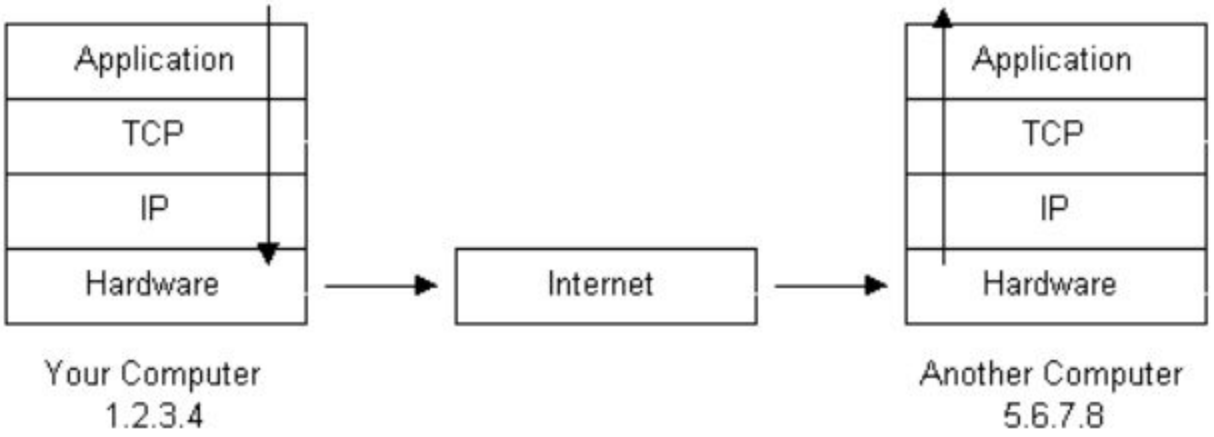
- **So till now computer is connected to the Internet and has a unique address. {Previous Step}**
- How does it 'talk' to other computers connected to the Internet?
  - An example should serve here:
  - Let's say your IP address is 1.2.3.4 and you want to send a message to the computer 5.6.7.8.
  - Message is **"Hello computer 5.6.7.8!".**
  - The message must be transmitted over any kind of wire connects the computer to the Internet.
  - Let's say you've dialed into your ISP from home and the message must be transmitted over the phone line.
  - Therefore the message must be translated from alphabetic text into electronic signals, transmitted over the Internet, then translated back into alphabetic text.
  - How is this accomplished? **Through the use of a protocol stack**.

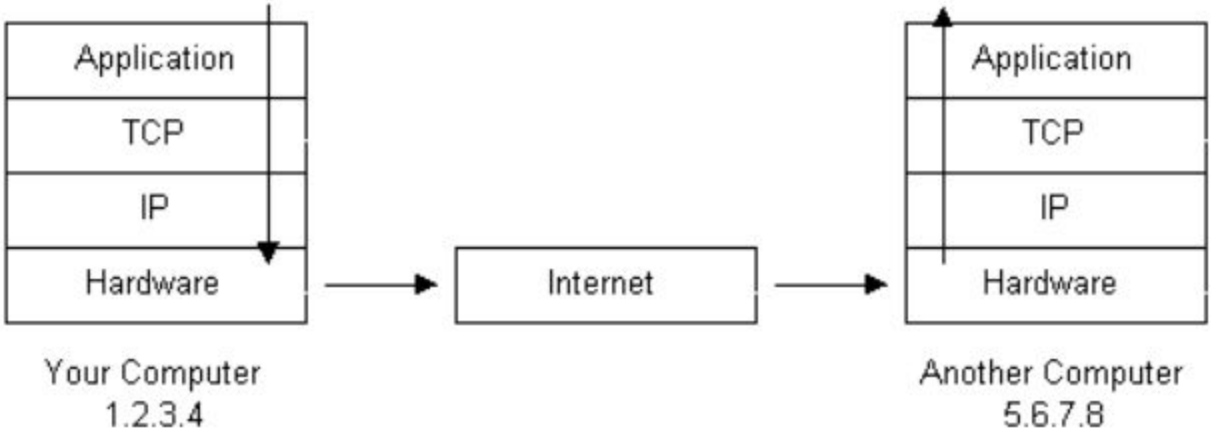
  - The protocol stack used on the Internet is referred to as the **TCP/IP protocol stack** because of the two major communication protocols used.

The TCP/IP stack

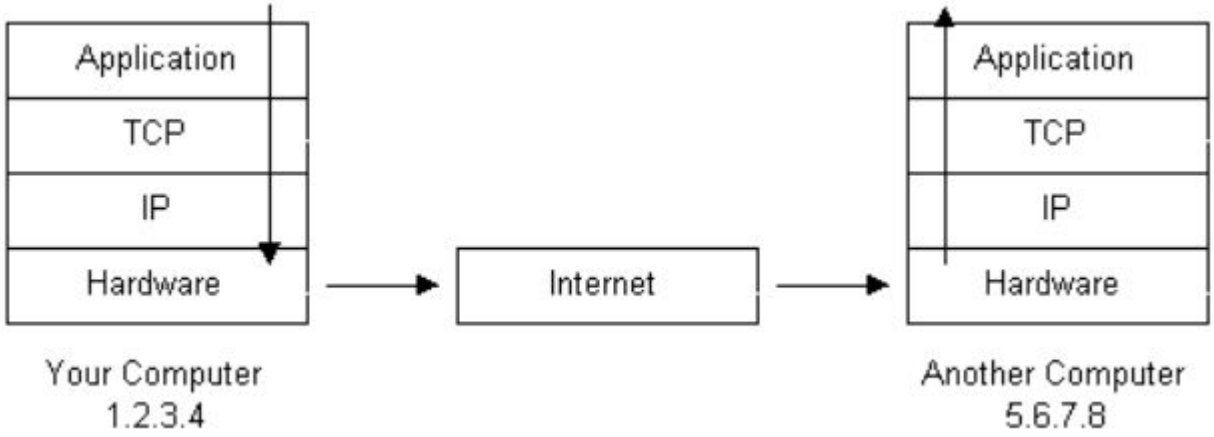
Protocol Layer	Comments
Application Layer	Protocols specific to applications such as WWW, e-mail, FTP, etc.
Transmission Control Protocol Layer	TCP directs packets to a specific application on a computer using a port number.
Internet Protocol Layer	IP directs packets to a specific computer using an IP address.
Hardware Layer	Converts binary packet data to network signals and back. (E.g. ethernet network card, modem for phone lines, etc.)

If we were to follow the path that the message "Hello computer 5.6.7.8!" took from our computer to the computer with IP address 5.6.7.8, it would happen something like this:





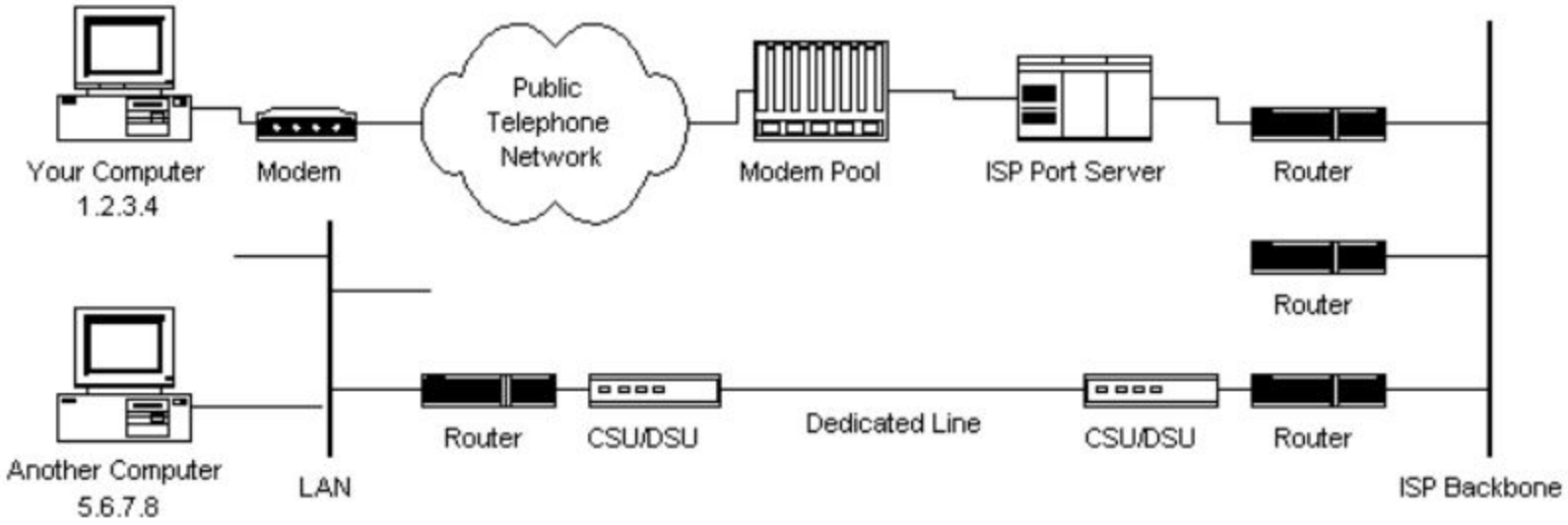
- The message would start at the top of the protocol stack on your computer and work it's way downward.
- If the message to be sent is long, each stack layer that the message passes through may **break the message up into smaller chunks of data**. This is because data sent over the Internet (and most computer networks) are sent in manageable chunks. On the Internet, these chunks of data are known as **packets.**
- The packets would go through the Application Layer and continue to the TCP layer. **Each packet is assigned a port number**.
- After going through the TCP layer, the packets proceed to the IP layer. This is where each **packet receives it's destination address, 5.6.7.8**.
- Now that our message packets have a port number and an IP address, they are ready to be sent over the Internet. The hardware layer takes care of turning our packets containing the **alphabetic text of message into electronic signals** and transmitting them over the phone line.



- On the other end of the phone line, **ISP has a direct connection to the Internet**. The ISPs router examines the destination address in each packet and determines where to send it. Often, **the packet's next stop is another router**.
- Eventually, the packets reach computer 5.6.7.8. Here, the packets start at the bottom of the destination computer's TCP/IP stack and work upwards.
- As the packets go upwards through the stack, all routing data that the sending computer's stack added (such as IP address and port number) is stripped from the packets.
- When the data reaches the top of the stack, **the packets have been re-assembled into their original form**, "Hello computer 5.6.7.8!"

Networking Infrastructure

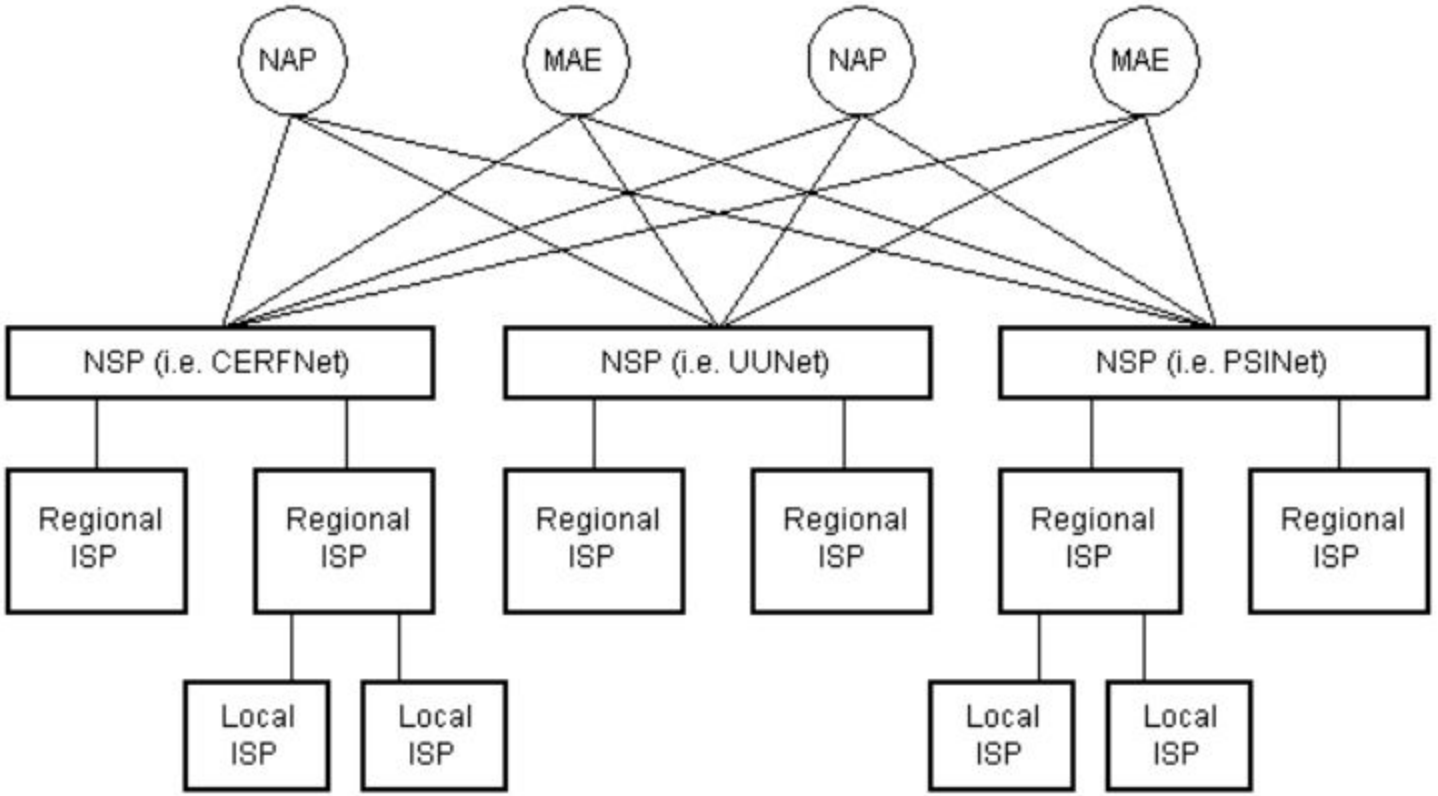
- **So till now the packets has travel from one computer to another over the Internet. {Previous Step}**
- But what's in-between? What actually makes up the Internet?



- The physical connection through the phone network to the Internet Service Provider might have been easy to guess, but beyond that might bear some explanation.

- The ISP maintains a pool of modems for their dial-in customers. This is managed by some form of computer (usually a dedicated one) which controls data flow from the modem pool to a backbone or dedicated line router.
- This setup may be referred to as a **port server**, as it 'serves' access to the network. **Billing and usage information is usually collected here as well**.
- Once the packets traverse through phone network and ISP's local equipment, they are routed onto the ISP's backbone or a backbone the ISP buys bandwidth from.
- From here the packets will usually journey through several routers and over several backbones, dedicated lines, and other networks until they find their destination, the computer with address 5.6.7.8.

Internet Infrastructure



- The Internet backbone is made up of many large networks which interconnect with each other. These large networks are known as **Network Service Providers or NSPs**. Some of the large NSPs are **UUNet, CerfNet, IBM, BBN Planet, SprintNet, PSINet**, as well as others.

- These networks **peer** with each other to exchange packet traffic. Each NSP is required to connect to three **Network Access Points** or **NAP**s.
- At the NAPs, packet traffic may jump from one NSP's backbone to another NSP's backbone. NSPs also interconnect at **Metropolitan Area Exchanges** or **MAE**s.
- MAEs serve the same purpose as the NAPs but are **privately owned**. NAPs were the original Internet interconnect points.
- Both NAPs and MAEs are referred to as **Internet Exchange Points or IXs**. NSPs also sell bandwidth to smaller networks, such as ISPs and smaller bandwidth providers.

The Internet Routing Hierarchy

- So how do packets find their way across the Internet?
- Does every computer connected to the Internet know where the other computers are?
- Do packets simply get 'broadcast' to every computer on the Internet?

The answer to both the preceding questions is 'NO'

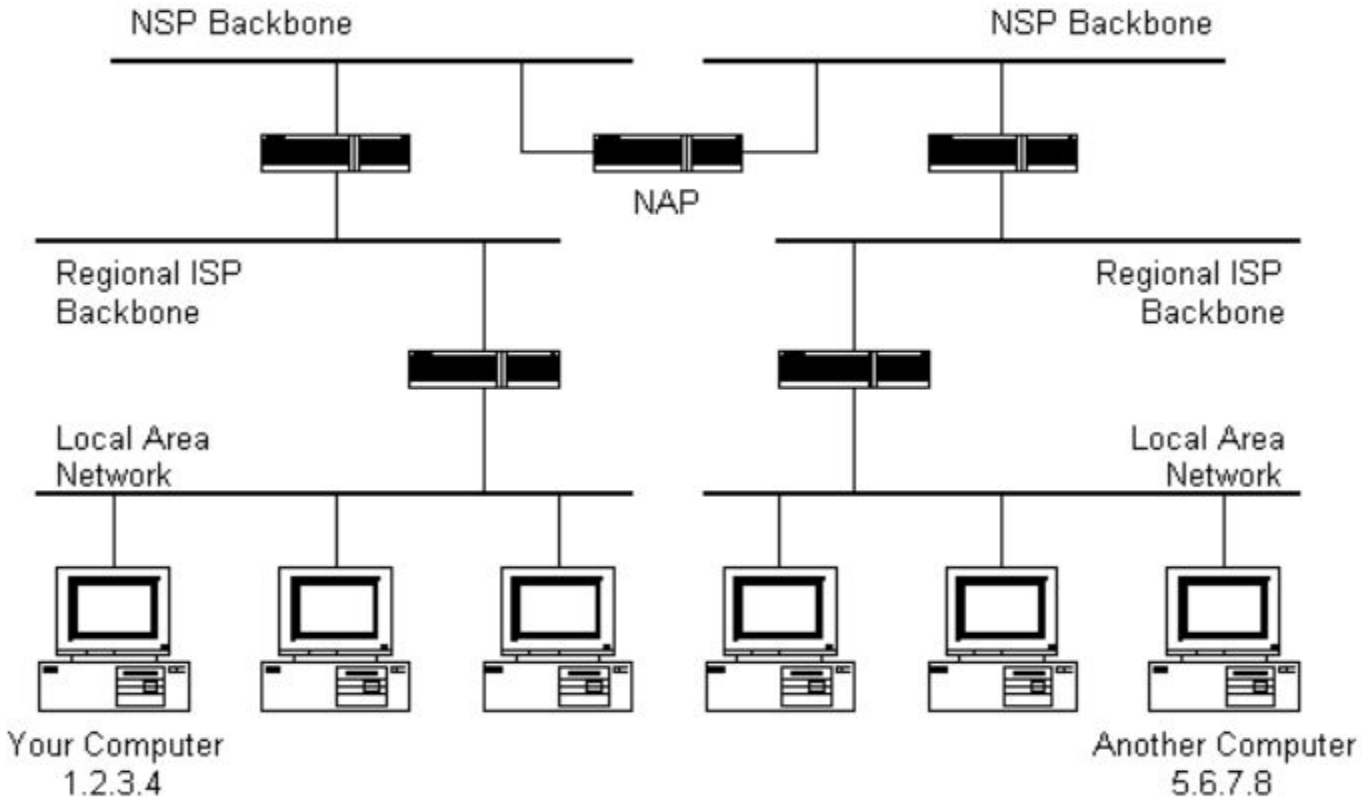
- No computer knows where any of the other computers are, and packets do not get sent to every computer. The information used to get packets to their destinations are contained in **routing tables** kept by each router connected to the Internet.

Routers are packet switches

- A router is usually connected between networks to route packets between them. Each router knows about it's sub-networks and which IP addresses they use. The router usually doesn't know what IP addresses are 'above' it.

- The black boxes connecting the backbones are routers. The larger NSP backbones at the top are connected at a NAP. Under them are several sub-networks, and under them, more sub-networks. At the bottom are two local area networks with computers attached.

- When a packet arrives at a router, the router examines the IP address put there by the IP protocol layer on the originating computer. The router checks it's routing table. If the network containing the IP address is found, the packet is sent to that network.



- If the network containing the IP address is not found, then the router sends the packet on a default route, usually up the backbone hierarchy to the next router. Hopefully the next router will know where to send the packet.
- If it does not, again the packet is routed upwards until it reaches a NSP backbone.
- The routers connected to the NSP backbones hold the largest routing tables and here the packet will be routed to the correct backbone, where it will begin its journey 'downward' through smaller and smaller networks until it finds it's destination.

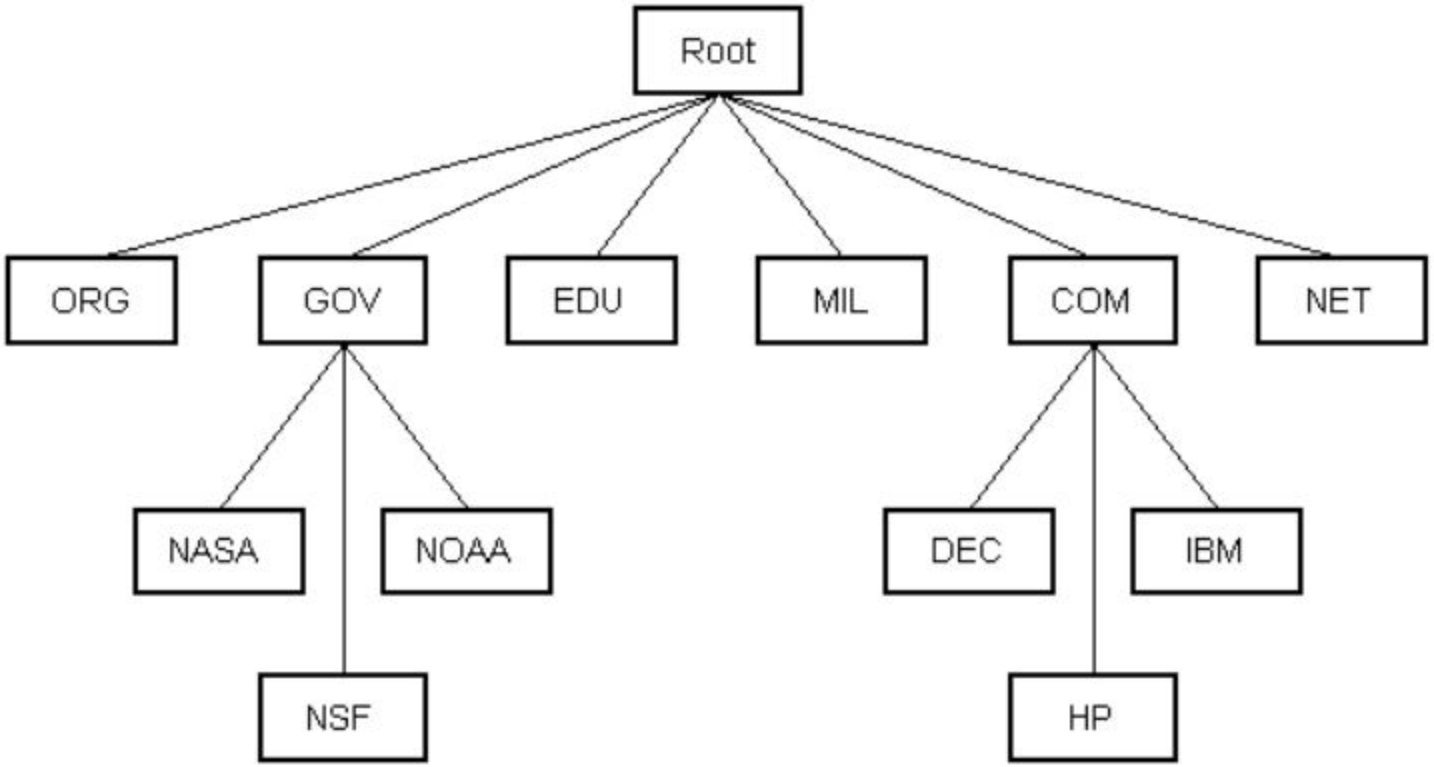
Domain Names and Address Resolution

- But what if you don't know the IP address of the computer you want to connect to?
- What if the you need to access a web server referred to as *www.anothercomputer.com*?
- How does your web browser know where on the Internet this computer lives?

The answer to all these questions is the Domain Name Service or DNS.

- The DNS is a distributed database which keeps track of computer's names and their corresponding IP addresses on the Internet.

- Many computers connected to the Internet host part of the DNS database and the software that allows others to access it.
- These computers are known as DNS servers. No DNS server contains the entire database; they only contain a subset of it.
- If a DNS server does not contain the domain name requested by another computer, the DNS server re-directs the requesting computer to another DNS server.



- The Domain Name Service is structured as a hierarchy similar to the IP routing hierarchy.
- The computer requesting a name resolution will be re-directed 'up' the hierarchy until a DNS server is found that can resolve the domain name in the request.

- At the top of the tree are the domain roots. Some of the older, more common domains are seen near the top.
- When an Internet connection is setup (e.g. for a LAN or Dial-Up Networking in Windows), one primary and one or more secondary DNS servers are usually specified as part of the installation.
- This way, any Internet applications that need domain name resolution will be able to function correctly.

- **For example,** when you enter a web address into your web browser, the browser first connects to your primary DNS server. After obtaining the IP address for the domain name you entered, the browser then connects to the target computer and requests the web page you wanted.

Application Protocols: HTTP and the World Wide Web

- One of the most commonly used services on the Internet is the World Wide Web (WWW). The application protocol that makes the web work is **Hypertext Transfer Protocol** or **HTTP**.

Do not confuse this with the Hypertext Markup Language (HTML). HTML is the language used to write web pages.

- HTTP is the protocol that **web browsers and web servers use to communicate** with each other over the Internet.
- It is an application level protocol because it sits on top of the TCP layer in the protocol stack and is used by specific applications to talk to one another.
- In this case the applications are **web browsers and web servers**.

- HTTP is a connectionless text based protocol.
- Clients (web browsers) send requests to web servers for web elements such as web pages and images.
- After the request is serviced by a server, the connection between client and server across the Internet is disconnected.
- A new connection must be made for each request. Most protocols are connection oriented. This means that the two computers communicating with each other keep the connection open over the Internet.
- HTTP does not however. Before an HTTP request can be made by a client, a new connection must be made to the server.

When you type a URL into a web browser, this is what happens:

- If the URL contains a domain name, the browser first **connects to a domain name server** and retrieves the corresponding **IP address** for the web server.
- The web browser connects to the web server and **sends an HTTP request** (via the protocol stack) for the desired web page.
- The web server receives the request and **checks for the desired page**. If the page exists, the web server sends it. If the server cannot find the requested page, it will send an **HTTP 404 error message**. (404 means **'Page Not Found'**.)
- The web browser receives the page back and **the connection is closed**.
- The browser then parses through the page and looks for other page elements it needs to complete the web page. These usually include images, applets, etc.
- For each element needed, the browser **makes additional connections** and HTTP requests to the server for each element.
- When the browser has finished loading all images, applets, etc. the page will be completely loaded in the browser window.

Application Protocols: SMTP and Electronic Mail

- Another commonly used Internet service is electronic mail. E-mail uses an application level protocol called **Simple Mail Transfer Protocol** or **SMTP**.
- SMTP is also a text based protocol, but unlike HTTP, SMTP is connection oriented. SMTP is also more complicated than HTTP. There are many more commands and considerations in SMTP than there are in HTTP.

- **When you open your mail client to read your e-mail, this is what typically happens:**

  - The mail client (Netscape Mail, Lotus Notes, Microsoft Outlook, etc.) opens a connection to it's default mail server. The mail server's IP address or domain name is typically setup when the mail client is installed.
  - The mail server will always transmit the first message to identify itself.

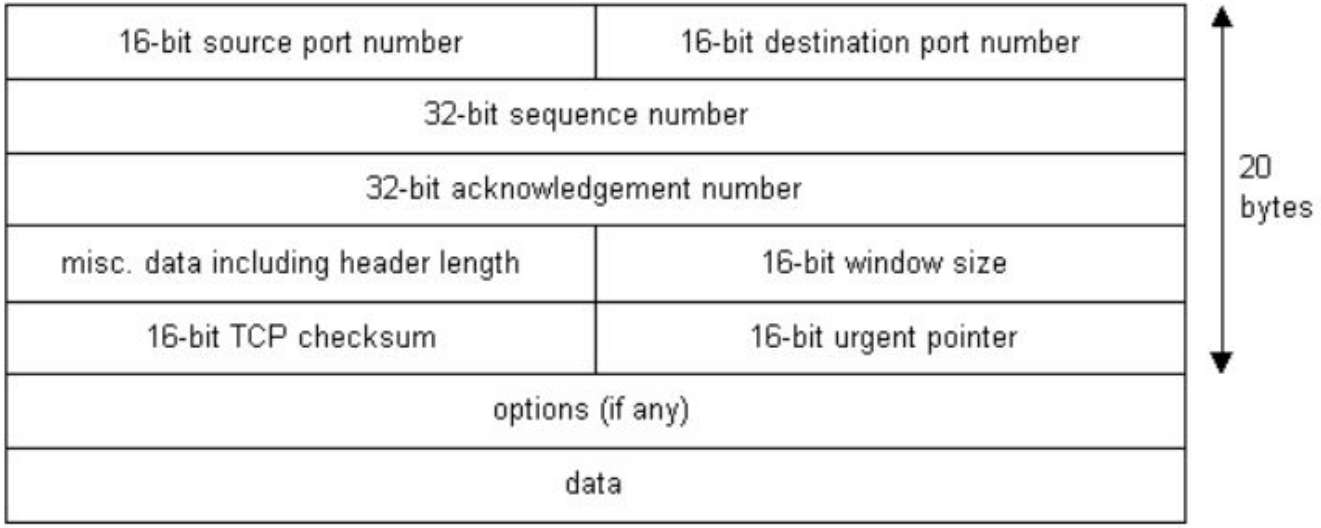
- The client will send an SMTP HELLO command to which the server will respond with a 250 OK message.
- Depending on whether the client is checking mail, sending mail, etc. the appropriate SMTP commands will be sent to the server, which will respond accordingly.
- This request/response transaction will continue until the client sends an SMTP QUIT command. The server will then say goodbye and the connection will be closed.

Transmission Control Protocol

- Under the application layer in the protocol stack is the TCP layer. When applications open a connection to another computer on the Internet, the messages they send (**using a specific application layer protocol**) get passed down the stack to the TCP layer.
- **TCP is responsible for routing application protocols to the correct application on the destination computer**.
- To accomplish this, **port numbers** are used. Ports can be thought of as separate channels on each computer.

- **For example,** you can surf the web while reading e-mail. This is because these two applications (**the web browser and the mail client**) used different port numbers. When a packet arrives at a computer and makes its way up the protocol stack, the TCP layer decides which application receives the packet based on a port number.

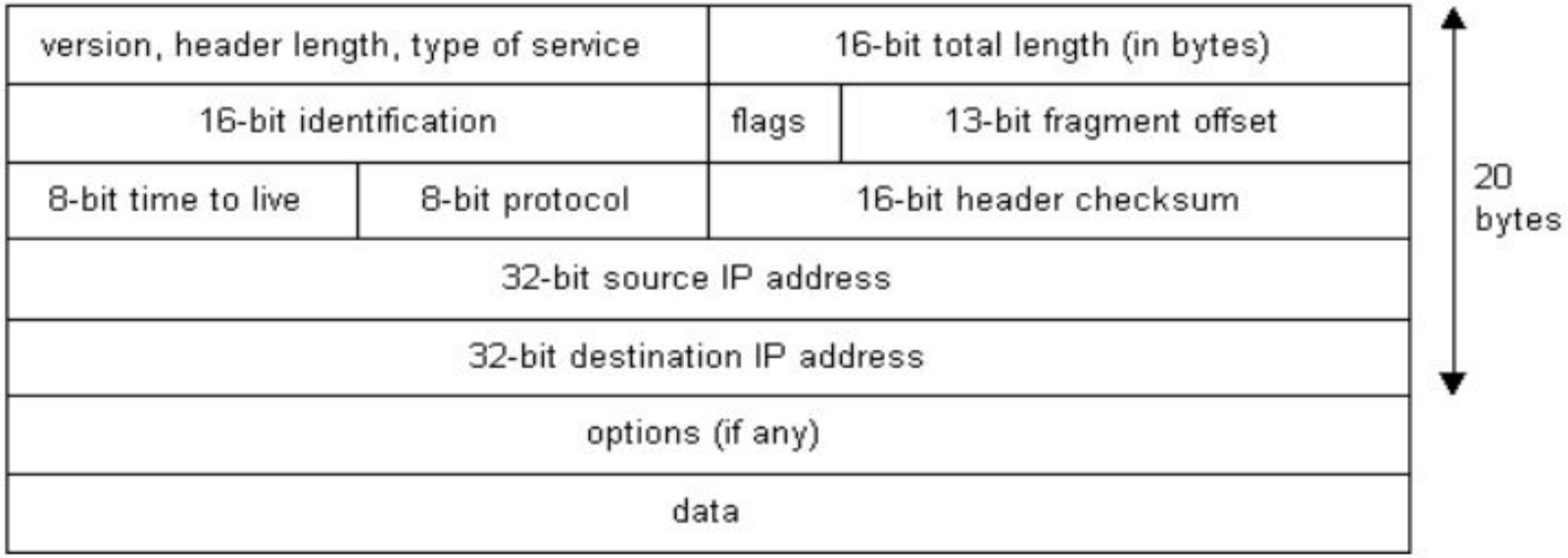
- **TCP works like this:**
  - When the TCP layer receives the application layer protocol data from above, it segments it into manageable **'chunks'** and then **adds a TCP header** with specific TCP information to each 'chunk'. The information contained in the TCP **header includes the port number** of the application the data needs to be sent to.
  - When the TCP layer receives a packet from the IP layer below it, the TCP layer strips the TCP header data from the packet, does some data reconstruction if necessary, and then sends the data to the correct application using the port number taken from the TCP header.
- This is how TCP routes the data moving through the protocol stack to the correct application.
- TCP is not a textual protocol. **TCP is a connection-oriented, reliable, byte stream service**.
- Connection-oriented means that two applications using TCP must first establish a connection before exchanging data.
- TCP is reliable because for each packet received, an acknowledgement is sent to the sender to confirm the delivery.
- TCP also includes a checksum in it's header for error-checking the received data.



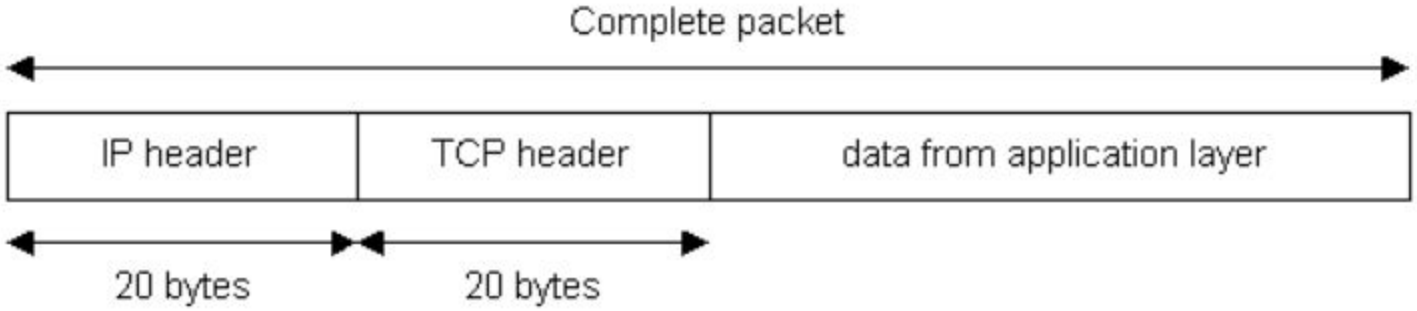
Notice that there is no place for an IP address in the TCP header. This is because TCP doesn't know anything about IP addresses. TCP's job is to get application level data from application to application reliably. The task of getting data from computer to computer is the job of IP.

Internet Protocol

- Unlike TCP, **IP is an unreliable, connectionless protocol**. IP doesn't care whether a packet gets to it's destination or not. Nor does IP know about connections and port numbers.
- **IP's job is too send and route packets to other computers**. IP packets are independent entities and may arrive out of order or not at all. It is TCP's job to make sure packets arrive and are in the correct order.
- About the only thing IP has in common with TCP is the way it receives data and adds it's own IP header information to the TCP data.
- **The IP header looks like this:**



- In previous slides, we see the IP addresses of the sending and receiving computers in the IP header.
- Below is what a packet looks like after passing through the application layer, TCP layer, and IP layer. The application layer data is segmented in the TCP layer, the TCP header is added, the packet continues to the IP layer, the IP header is added, and then the packet is transmitted across the Internet.



Internet Addressing scheme

- Machine Addressing (IP Address)
- E-mail Address
- Resource Addresses

Chapter-1

INTERconnected NETwork: Internet

Internet and the Web

- Both are different terms, not synonymous.
- Internet: Huge network of networks connecting millions of computers together.
- Web: One of the popular services offered on the internet. It is one of the ways in which information is disseminated and accessed over the medium of internet.

Internet: The Giant WAN

- It started as a federal funded research project ARPANET initiated by the Advanced Research Project Agency (ARPA), Department of Defense of the US Government in 1969.
- The experiments were initially to link researchers with a remote computer centers, allowing them to share hardware and software resources such as computer disk space and databases. Later, it was renamed Internet.
- As the internet evolved it become challenging to allow communication with different networks.
- The problem was solved by developing the Internet Protocol (IP) which truly created networks of networks, the current architecture of the internet.

Packet Switching and TCP

- It operates with a technique: Packet Switching.
- In packet switching, digital data is transmitted in small bundles called packets.
- These packets contains information about the address, error control and the sequence in which the packets are to be sent.
- There was no centralized control. If a portion of a network failed the remaining working portion will still route packets.
- ARPANET used Transmission Control Protocol known as TCP for communication.
- TCP ensured the messages were properly routed from sender to receiver.

Classification of Networks

Based on transmission media

- Wired: Unshielded Twisted-Pair cables, shielded twisted-pair cables, coxial cables, fibre-optics cables
- Wireless

Based on physical topology (layout of the network)

- Point to point (PTP)
- Multi-access: Ring, Star, or bus

Based on application architecture

- Peer to Peer
- Client Server

Based on Geographical Area Covered:

- LAN (Local Area Networks)
- MAN (Metropolitan Area Network)
- WAN (Wide Area Network)

LAN

- Used for communicating among computer devices usually within an office building or home.
- Spanning a few hundred meters, and no more than a mile.
- It is fast, with speeds from 10 Mbps to 10 Gbps
- Minimal infrastructure requirements, low cost and high security.
- Can be either wired or wireless.
- Nodes in LAN are linked together with certain topology.

MAN

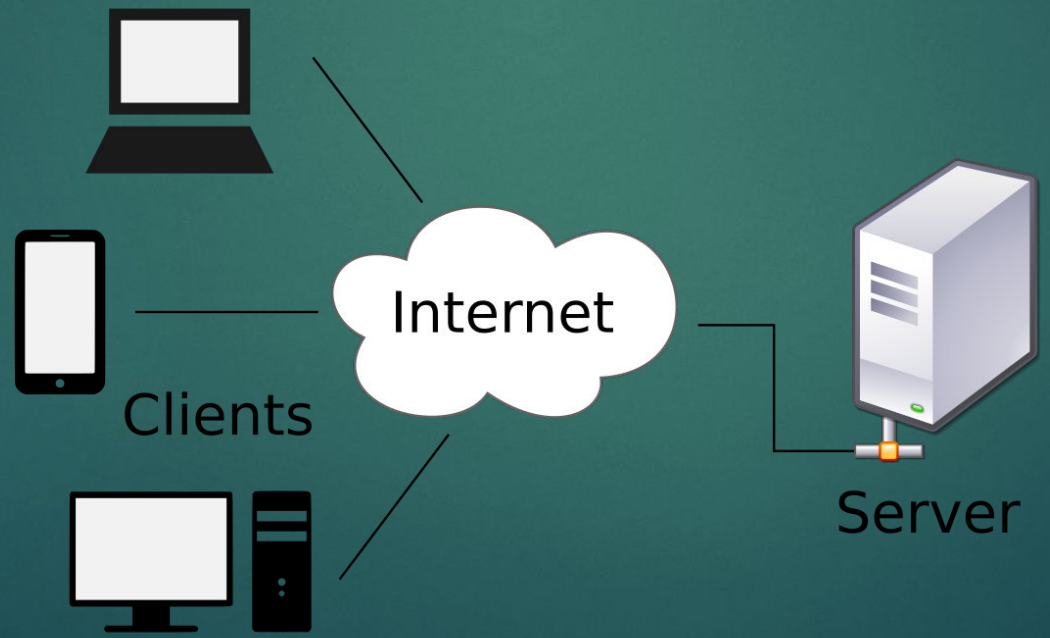
- Optimized for a larger geographical area than a LAN
- Ranging from 5 to 50km diametrically.
- Allows sharing of regional resources.
- Might be owned or operated by single organization.
- Speed achieved is as high as in LAN, it requires high speed connections, such as fiber optics.
- High installation and operational costs.

WAN

- Covers large areas such as country, continent or even whole world.
- Two or multiple LANs connected together using devices such as bridges, routers or gateways.
- To cover distance, WANs may transmit data over leased high-speed phone lines or wireless links such as satellites.

Client/Server Network

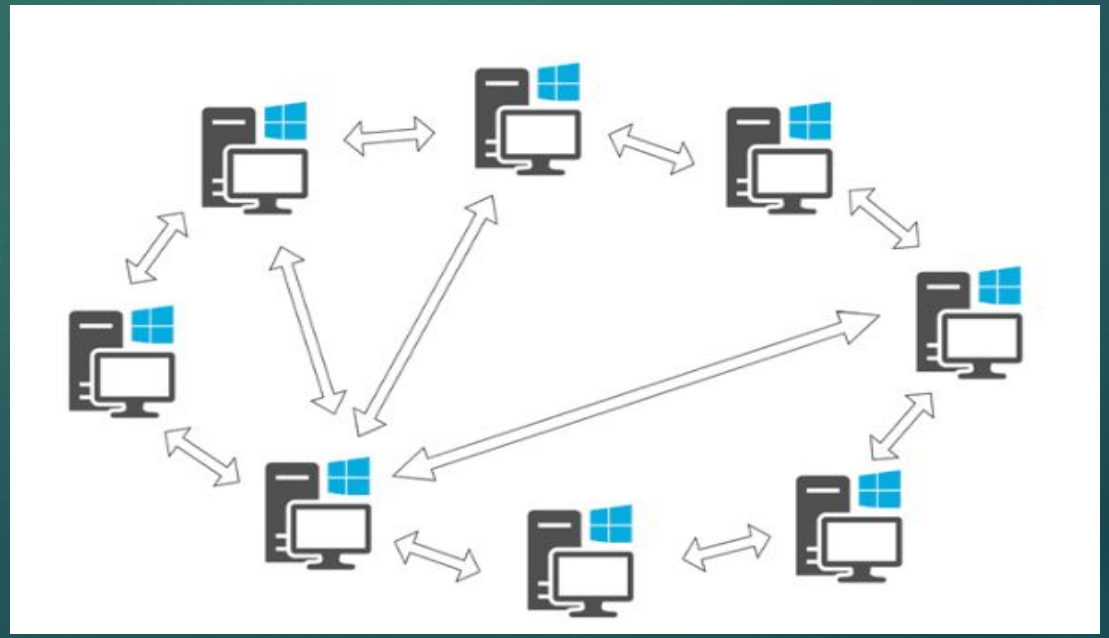
- All end systems are divided into clients and servers each designed for specific purpose.
- A central computer, such as a workstation or a server is a common source that provides shared services with other machines and manages resources in the network.



- Clients initiates the communication by sending requests to server.
- Clients communicate with servers only.
- Servers have passive role and respond to their clients by acting on each request and returning results.
- One server generally supports numerous clients.

Peer-to-Peer Network

- All the end systems have equivalent capabilities and responsibilities (equal peers).
- Either party can initiate a communication session.
- No central location for authenticating users, storing files, or accessing resources.
- The expected maximum number of peers that can operate on a peer to peer network is ten.



Types of P2P networks

- Pure P2P network: No central service of any kind i.e the entire communication occurs among connected peers without any assistance. E.g.- Workgroup in Microsoft, Freenet.
- Hybrid P2P Network: Depends partially on servers or allocate selected functions to a subset of dedicated peers. Dedicated peers directly control information among the other peers. E.g.- Skype, BitTorrent

OSI Model

Approved as an international standard for communications architecture in 1984. It divides the problem of moving information between computers over a network medium into seven layers. Each layer provides a service to the layer above it in the protocol specification.

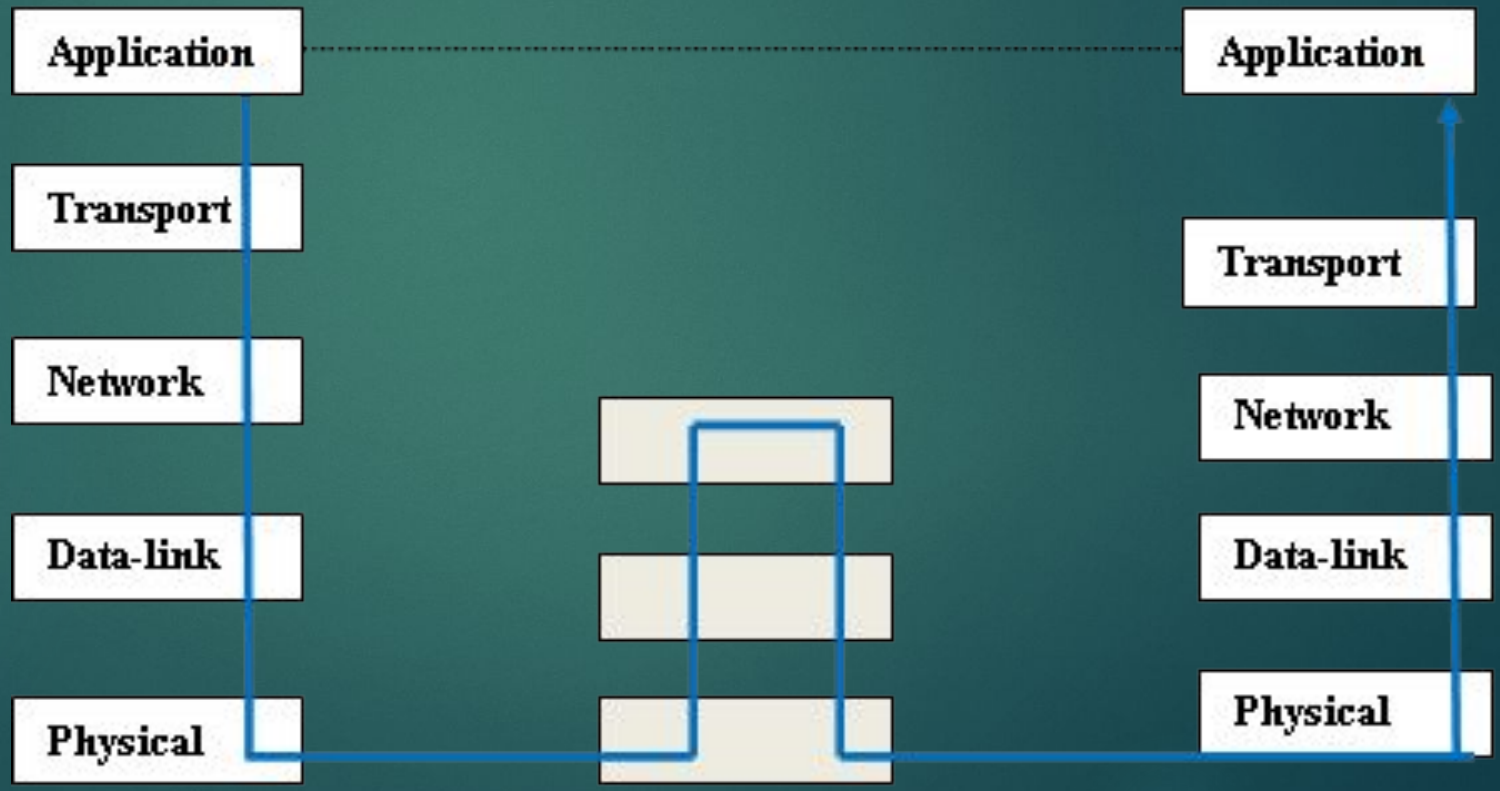
1. Application
2. Presentation
3. Session
4. Transport
5. Network
6. Data Link
7. Physical

TCP/IP Model

OSI REFERENCE MODEL	TCP/IP CONCEPTUAL LAYERS
7. APPLICATION	4. APPLICATION
6. PRESENTATION	
5. SESSION	
4. TRANSPORT	3. TRANSPORT
3. NETWORK	2. INTERNET (NETWORK)
2. DATA LINK	1. NETWORK INTERFACE
1. PHYSICAL	

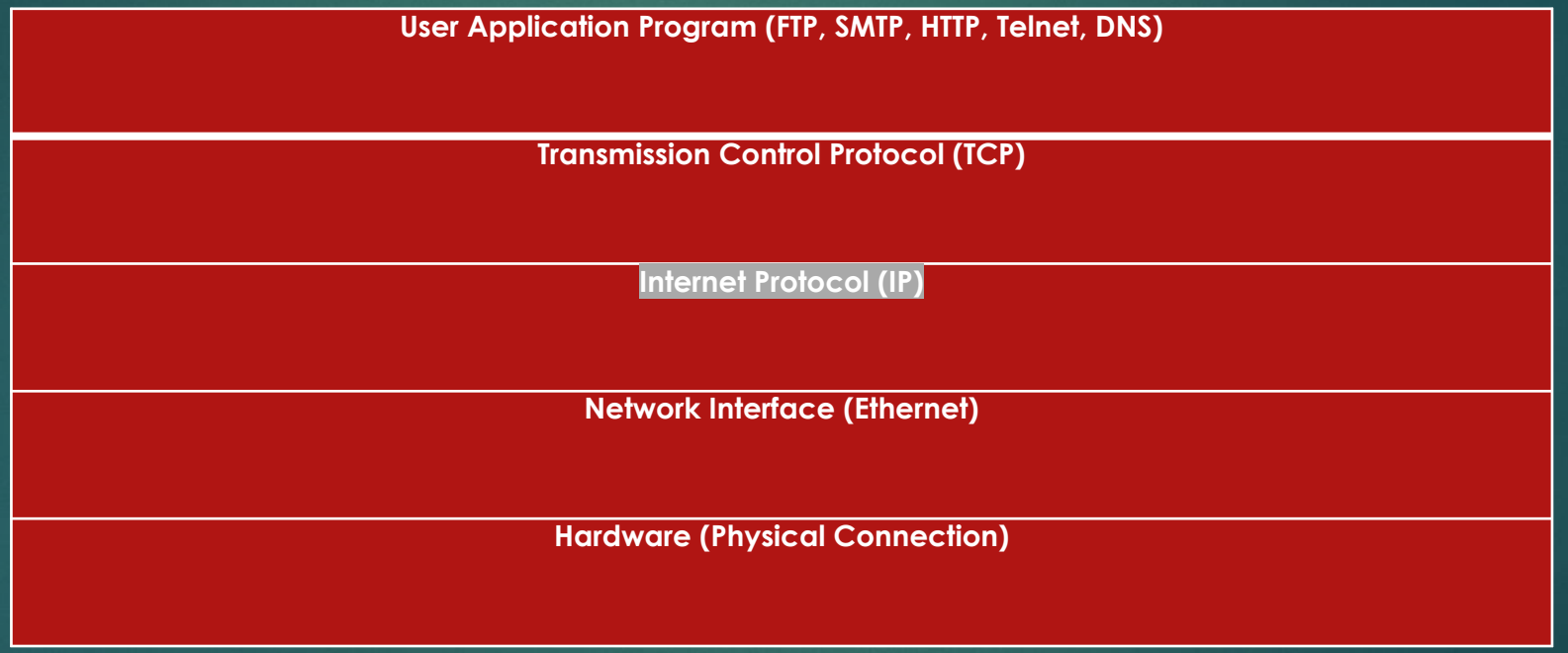
Protocol Layering

To communicate using the internet system, a host must implement a layered set of protocols comprising the Internet protocol suite.



Internet Protocol Stack

The protocol layers used in the Internet architecture are as follows:



Role of each layer

Application Layer: It combines the functions of the two top layers- Presentation and Application of the OSI model. There are two categories of application layer protocols: user protocols that provide service directly to users and the support protocols that provide common system function. Most common internet user protocols are: Telnet, FTP, SMTP, HTTP. Support protocols are used for host name mapping, booting and management include Simple Network Management Protocol (SNMP).

Transport Layer: It is responsible for reliable source to destination delivery of the entire message. There are two primary Transport Layer protocol- Transmission Control Protocol(TCP) and User Datagram Protocol (UDP). TCP is reliable connection-oriented protocol whereas UDP is a connectionless protocol.

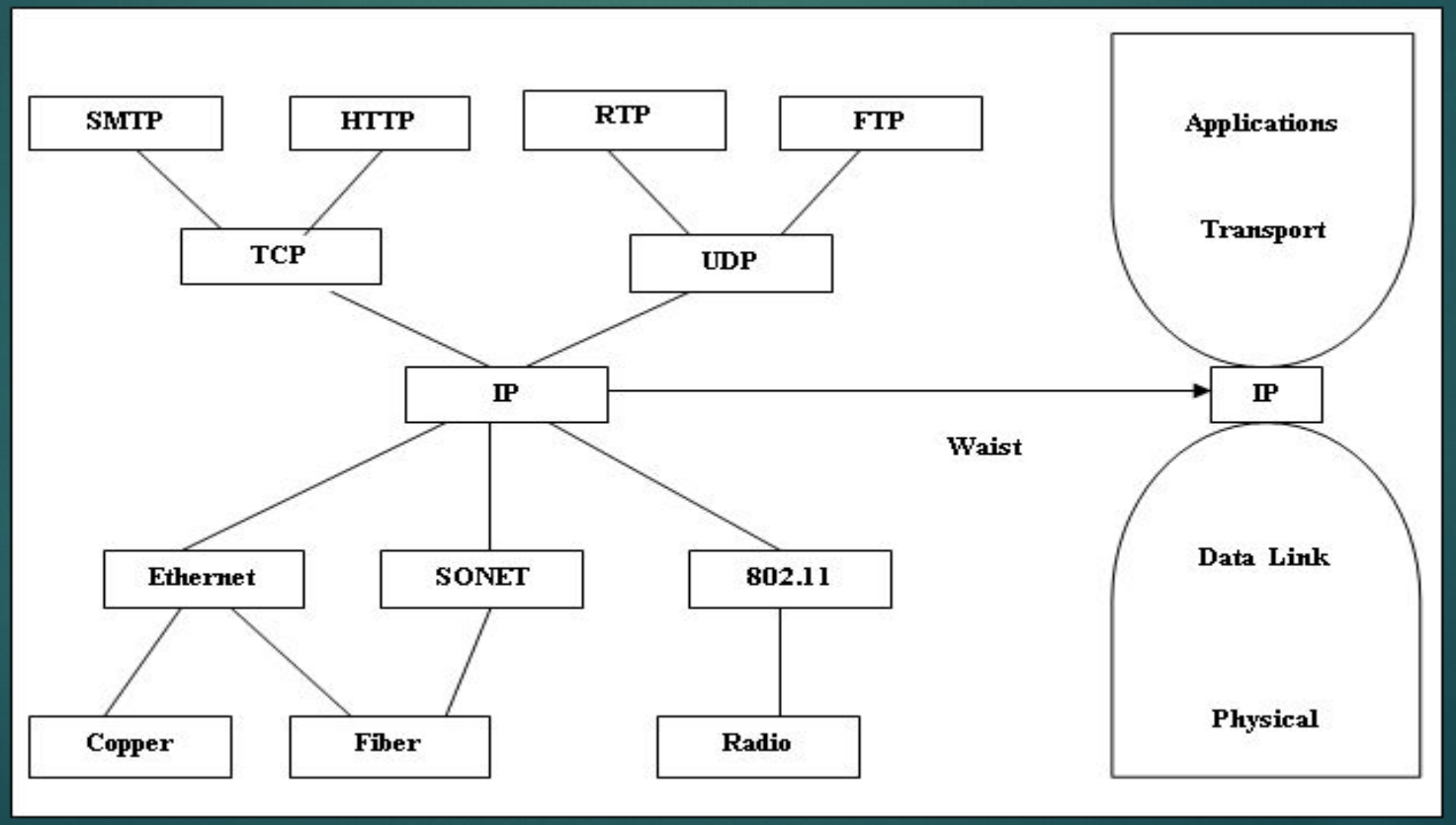
Internet Layer: All Internet transport protocols use the Internet Protocol (IP) to carry data from source host to destination host. IP is connectionless service with no end-to-end delivery guarantee. The Internet Control Message Protocol (ICMP) provides error reporting, congestion reporting , and first-hop router redirection.

Link Layer (Network Interface): To communicate on a directly connected network, a host must implement the communication protocol used to interface to that network. The link layer protocol provides this interface. It specifies how to organise data into frames and how to deliver a frame over a network. Its responsibility is correct delivery of message.

Physical Layer: Provides physical interface for transmission of information. Defines rules by which bits are passed from one system to another on a physical communication medium. It covers all mechanical, electrical, functional and procedural aspects for physical communication.

The Hourglass Model

Depicts IP as the waist of the hourglass of the Internet protocol architecture with multiple higher-layer protocols, multiple lower-layer protocols and only one protocol at the network layer.



Curse of Narrow Waist: As the number of service interfaces doubles, changes are desired below and above the waist to avoid interoperability problems. Moreover, below the waist bulge is apparent too as the lower layers mostly seem to just make IP's harder with cells, circuits, QoS, multicasts, large clouds, opaque clouds. Many reserachers have suggested a promising fitness goal to trim down from an hourglass to wineglass architecture.

Internet Addressing – IP Address

- IP address is a unique global address for a network interface. It is 32-bit logical address, composed of four 8-bit fields, called octets. Each octet represents a decimal number in the range 0-255. For eg- 17.112.152.32
- It is divided into two parts: a prefix and a suffix.
- Prefix: Identifies the physical network to which the host is attached.
- Suffix: Identifies a specific computer(host/node) on the network.
- A single address in never assigned to more than one computer.
- Network number (prefix) assignments must be coordinated globally whereas suffixes are assigned locally without global coordination.

Uniform Resource Locators (URL)

- It specifies the internet address of any resource.
- They are translated into numeric addresses using a Domain Name Server (DNS).
- When we type a URL, this application-layer service called DNS translates the human friendly URL into the computer friendly IP address. DNS Performs this translation by consulting the databases maintained by the Domain Name Registrars.
- DNS is the “phone book” of the internet: look up a name, and find its number. DNS associates host names to their equivalent IP address.
- The domain name or IP address gives the destination location for the URL.

The generic anatomy of a URL is:

protocol://domain name/path/filename

Protocol specifies the transfer protocol that will be used for the retrieval of desired resource. Example- http, https, ftp.

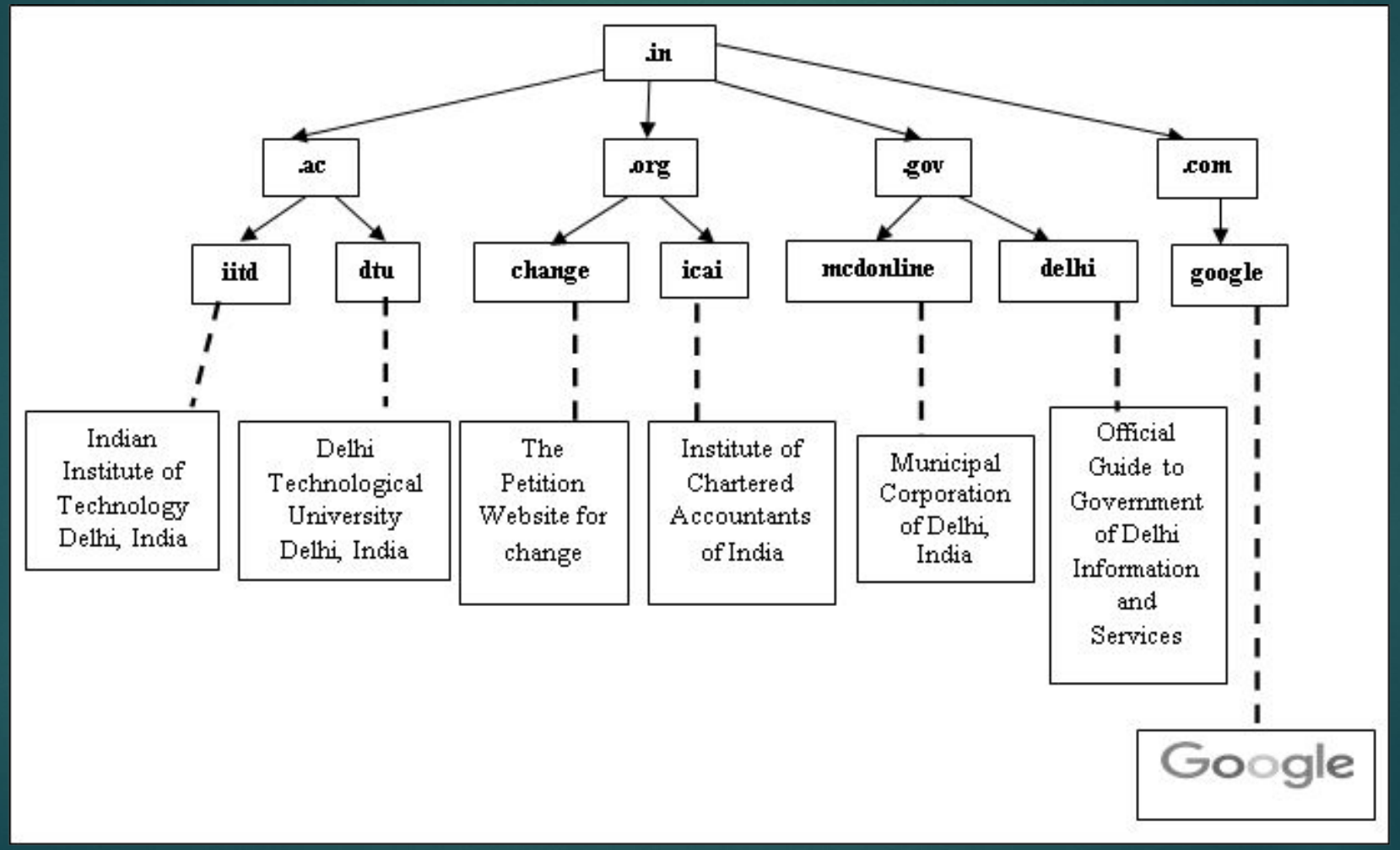
Domain Name is also known as website name or host name. It is divided into three parts: www (optional), second level name (dtu.ac in www.dtu.ac.in), and top level name. It can be organizational 3-character code (.com, .org, .edu) or geographical 2 character code (.uk, .au).

Path is the directory or the folder on the server. File name is the file name within that directory, with an extension as html or php etc.

DNS

- It is implemented as a distributed system and has a hierarchical structure (can be represented as a tree).
- Each node represents a DNS name. Each branch below a node is a DNS domain. DNS domain can contain hosts or other domains. The top level domains are the root of the tree followed by the top-down sub domains.

Sample DNS Hierarchy



A URL is now considered to be a subset of Uniform Resource Identifier (URI). A URI is a string of characters used to identify a name or a resource on the Internet and is recognised as a more general form of URL. It identifies a resource either by location, or a name or both.

URL and URN (Uniform Resource Name) are two specializations of URI.

URL: Contains information about how to fetch a resource from its location.

URN: Identifies a resource by a unique and persistent name.

URI: Encompasses URLs, URNs, and other ways to identify a resource.

Internet Configuration

- Every computer connected to the Internet accesses the Internet through an Internet Service Provider(ISP).
- The ISP, in turn, may connect to a larger network such as a Network Service Provider (NSP) that provides backbone services to the ISP.
- These connections are collectively known as Internet Backbone.

Basic Building Blocks of the Internet

- An Internet backbone is a collection of routers (nation or worldwide) connected by high-speed point-to-point networks.
- A Network Access Point (NAP) is a router that connects multiple backbones (sometimes referred to as peers).
- Regional networks are smaller backbones that cover smaller geographical areas.
- A point of pressure (POP) is a machine that is connected to the internet.
- Internet Service Providers (ISPs) provide dial-up or direct access to POPs.

Internet service provider offers Internet accounts to configure the network connections either by using a dial-up access, high-speed access or wireless access.

Dial-up Access: Dial-up connection works over an ordinary phone line, using analog modems establishing the Point-to-Point (PPP). As the personal computers are mostly equipped with analog modems, there is usually no additional money needed for hardware.

High-speed Access: A High-speed connection also knows as the broadband connection comprises of divergent options such as Digital Subscriber Line (DSL), Integrated Services Digital Network (ISDN) Lines, leased lines and cable Internet connections.

Wireless Access: Wireless access uses Wi-Fi technology. A Wi-Fi enabled device such as a PC, game console, cell phone, MP3 player or PDA can connect to the Internet when within range of a wireless network connected to the Internet.

Web Browser

A web browser is a software program that retrieves, presents, and traverses information resources on the Web. Primary function of a browser is to identify the URL and bring the information resource to user.

To identify a Web pages' exact location, Web browsers rely on Uniform Resource Locator (URL).

URL is a four-part addressing scheme that tells the Web browser:

- What transfer protocol to use for transporting the file.
- The domain name of he computer on which the file resides.
- The pathname of the folder or directory on the computer on which the file resides.
- The name of the file.

All major browsers allow users to access multiple information resources at the same time in different windows or in tabs. They also include pop up blockers to open windows with users consent.

Examples of web browsers: Internet Explorer, Netscape Navigator, Opera, Google Chrome, Safari etc.

Basic Functions of a Web Browser

- Interpret HTML markup and present documents visually.
- Support hyperlinks in HTML documents so the clicking on such a hyperlink can lead to the corresponding HTML file being downloaded from the same or another web server and presented.
- Use HTML form and HTTP protocol to send requests and data to web applications and download HTML documents.
- Maintain cookies (name-value pairs) deposited on client computers by a web application and send all cookies back to a web site if they are deposited by the web application at that web site.

Internet Organizations

No one actually owns the Internet, and no single person or organization controls the Internet in its entirety. The Internet is more of a concept than an actual tangible entity, and it relies on a physical infrastructure that connects networks to other networks. A number of loosely coupled organizations are concerned with governing the development of the Internet.

ISOC: Internet Society, concerned with the long-tern coordination of the Internet development.

IETF: Internet Engineering Task Force, concerned with producing high quality technical documents for improving the Internet's quality and performance.

ICANN: Internet Corporation for Assigned Names and Numbers, responsible for IP address space allocation, gTLD (generic Top Level Domain) and ccTLD(country code TLD), DNS management, Root server system management, ad Protocol identifier assignment.

IANA: Internet Assigned Numbers Authority, manages different duties of ICANN, namely the TLD, protocol number, IP address and AS number management.

IAB: Internet Architecture Board, responsible for the Internet architecture as a whole and protocol development.

IESG: Internet Engineering Steering Group, carries out the technical management of IEFT activities and the Internet standards process.

IRFT: Internet Research Task Force, conducts research on protocols, applications, architecture and technology.

IRSG: Internet Research Steering Group, responsible for steering IRFT and provides good conditions for research carried out by IRTF.

W3C: World Wide Web Consortium, develops web technology standards.

Registries: RIRs are responsible for the management and allocation of Internet number resources, namely IP addresses and AS numbers.

Cyber Ethics

- Ethics which should be practiced to be good "cyber citizens" include:
- Communicating, sharing and contributing to e-society.
- To be respectful and courteous in communication
- Avoid harming others (do not spread pictures, viruses, gossip, information about others or impersonate others).
- Sharing network resources, being honest and trustworthy.
- Honor property rights and copyrights

Internet Applications

Internet Services

The internet can be viewed from two perspectives. One way is to describe the "nuts and bolts" of the internet, and the other is the "service" view of the internet.

Nuts and Bolts Description

Devices connected together by communication links are called host or end-systems. They are indirectly connected to each other through intermediate switching devices known as packet switches.

Different links can transmit data at different rates, with the transmission rate of a link measured in bits/second. End systems, packet switches, and other pieces of the Internet, run protocols that control the sending and receiving of information within the Internet. **Transmission Control Protocol (TCP) and the Internet Protocol (IP)** are two of the most important protocols in the Internet.

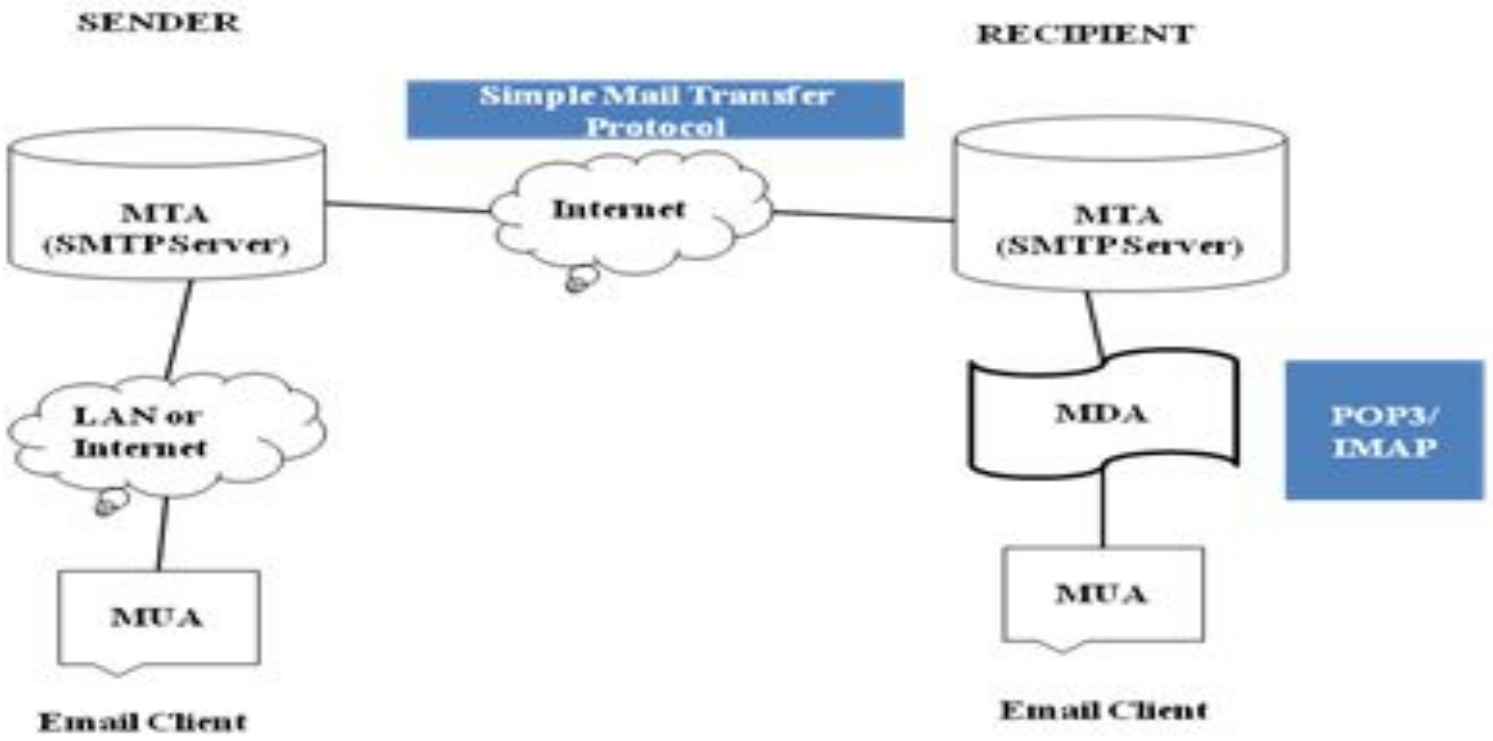
Service View Description

The Internet allows distributed applications such as **e-mail, file transfer, real-time user communication and Web**, running on its end systems to exchange data with each other. It provides two services to its distributed applications, namely a **connection-oriented reliable service and a connectionless unreliable service**. The following sections discuss these distributed applications provided as services by the Internet.

Electronic Mail (E-mail)

- It’s a method of sending a message from user at a computer to a recipient on another computer.
- Based on Store and Forward model.
- It’s a message that may contain text, files, images or other attachments.
- Based around the use of electronic mailboxes.
- The first email was sent by Ray Tomlinson in 1971.

The Store and Forward Model of E-Mail



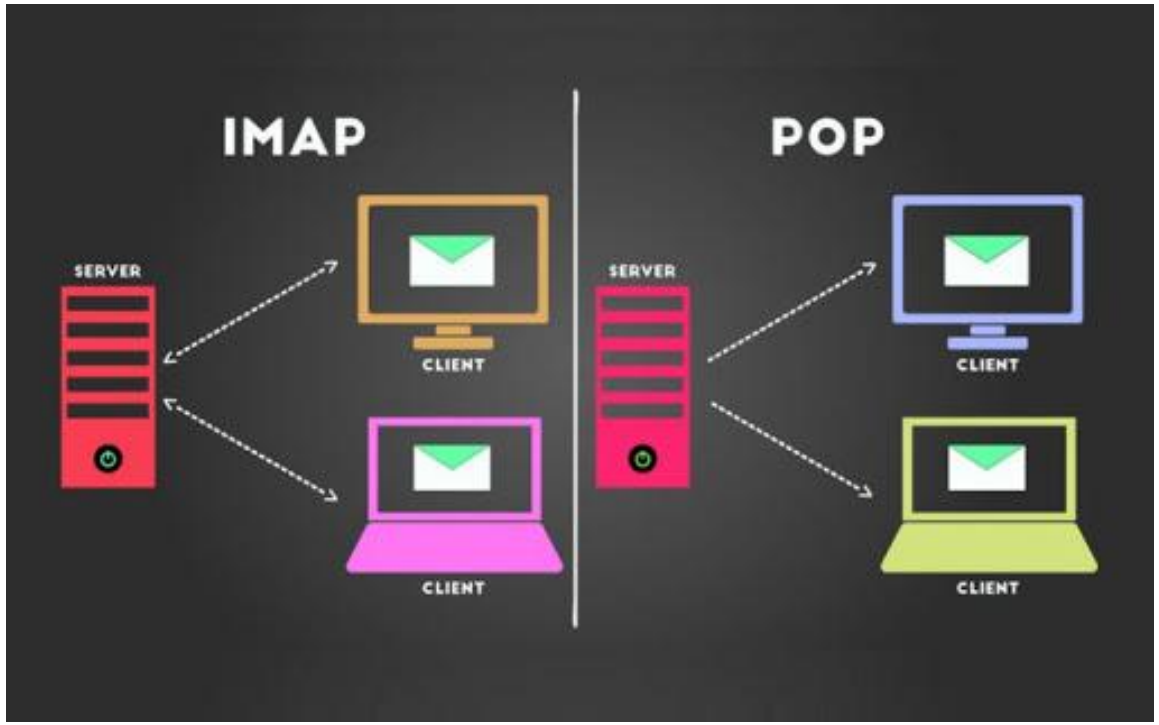
Working of an E-mail

- When an email is sent, the message is sent to the mail server tasked with transporting emails (called the **MTA**, for *Mail Transport Agent*) to the recipient's MTA. On the Internet, MTAs communicate with one another using the protocol SMTP *(Simple Mail Transfer Protocol),* and so are logically called **SMTP servers** (or sometimes *outgoing mail servers*).

- The recipient's MTA then delivers the email to the incoming mail server (called the **MDA**, for *Mail Delivery Agent*), which stores the email as it waits for the user to accept it.

There are two main protocols used for retrieving email on an MDA.

- **POP3*(Post Office Protocol*)**
- **IMAP*(Internet Message Access Protocol*).**



MUA (Mail User Agent)

- Retrieving mail is done using a software program called an **MUA** (*Mail User Agent*). When the MUA is a program installed on the user's system, it is called an **email client** (such as Mozilla Thunderbird, Microsoft Outlook, Eudora Mail, or Lotus Notes).When it is a web interface used for interacting with the incoming mail server, it is called **webmail** (Gmail, Hotmail, Yahoo mail).

- To use a real-world analogy, MTAs act as the post office ,while MDAs act as mailboxes, which store messages until the recipients check the box. This means that it is not necessary for recipients to be connected in order for them to be sent email.

Email Protocols:
Simple Mail Transport Protocols (SMTP)

- It is a Client-server protocol where client is the sending mail server and server is the receiving mail server
- It allows reliable data transfer built on top of TCP (on port 25)
- It is a Push protocol where sending server pushes the file to the receiving server rather than waiting for the receiver to request it
- It is a Command/response interaction where commands are ASCII text and responses are three digit status code and phrase.
- It is Synchronous as sender awaits response before issuing the next command.
- It has three phases of transfer: Handshaking, Transfer of messages and Closure.

Mail Access Protocols:
POP & IMAP

POP: When using a POP server, email messages are downloaded by email client applications. By default, most POP email clients are automatically configured to delete the message on the email server after it has been successfully transferred, however this setting usually can be changed. The most current version of the standard POP protocol is POP3.

- It does not handle multiple mailboxes easily, it is designed to put user's incoming e-mail in one folder
- It is poor in handling multiple-client access to mailbox which is increasingly important as users have home PC, work PC, laptop, cyber café computer, friend's machine, etc.
- It has high network bandwidth overhead, it transfers all of the e-mail messages, often well before they are read.

IMAP: When using an IMAP mail server, email messages remain on the server where users can read or delete them. IMAP is particularly useful for those who access their email using multiple machines. IMAP, like POP, is fully compatible with important Internet messaging standards, such as MIME, which allow for email attachments.

- Multiple clients can connect to mailbox at once. It detects changes made to the mailbox by other clients and server keeps state about message.
- Allows access to MIME parts of messages & partial fetch as clients can retrieve individual parts separately. E.g., text of a message without downloading attachments.
- Facilitates multiple mailboxes on the server where client can create, rename, and delete mailboxes and also can move messages from one folder to another.
- Enables server-side searches, i.e., search on server before downloading messages.

Vulnerability of Emails

- Although e-mail is a very practical tool for exchanging information, it is also vulnerable and users can encounter problems such as interception, identity theft and monitoring.
- One way to ensure confidentiality is by using email client such as Thunderbird instead of webmail or as an alternative emails can use an encryption program which provides cryptographic privacy and authentication.
- Pretty Good Privacy (PGP) is one such protocol which encrypts the emails end-to-end and only the addressee is able to decrypt it, thus excluding any possibility of interception.

FTP (File Transfer Protocol)

- FTP is a widely used network protocol for transferring files between computers over a TCP/IP-based network, such as the Internet.
- FTP lets people and applications exchange and share data within their offices and across the Internet.
- FTP was one of the first technologies developed to solve this common need, and it remains, with several generations of enhancements, the second most popular protocol used today (after HTTP or the "World Wide Web").

Applications of FTP

FTP finds application in many day-to-day business operations that span business-to-business and peer-to-peer data transfer use cases, including:

- Organizations use FTP to allow employees to share files across different locations and branch offices.
- Employees use FTP to securely share files with coworkers and external business partners.
- IT teams use FTP to transfer data back to DR (disaster recovery) sites.
- Webmaster teams use FTP to transfer Web pages, Web application files, and images to their Web server.

How File Transfer Happens with FTP

- FTP works in a **client-server model** where an FTP server and FTP client perform the file transfer operation.
- An FTP server is configured in the network, and a **specific file storage location (folder/system)** is identified to become the shared storage, which will host the files we want to share. The end-users will access this file server via FTP to start copying the files to their local folder/system.
- FTP requires a **TCP/IP network** to function, and relies on the use of one or more FTP clients.
- The FTP client acts as the communication agent to interact with the server to download or upload files. In other words, the FTP client sends out connections to the FTP server. Upon listening to the request from the client to either upload or download a file, the FTP server performs the file transfer operation.

FTP Client

- Previously FTP clients were just a **command line interface (CLI).**
- They now come in easy-to-use, intuitive interfaces to facilitate and simplify file transfers.
- FTP clients are used for **desktops, servers, and mobile devices**, and are available as standalone apps, Web clients, and simple extensions to Web browsers.

FTP Server

- The FTP server can support both **active and passive connections** with the FTP client.
  - In an active FTP connection, the client opens a port and listens while the server actively connects to it.
  - In a passive connection, the server opens a port and listens passively, which allows clients to connect to it.
- A passive connection is **more secure** and also preferred by IT admins because data connections are made from the FTP client to the FTP server. This is a more reliable method, and it avoids inbound connections from the Internet back into individual clients.
- Passive mode is also known as "firewall-friendly" mode. The more secure file transfer protocols (such as **FTPS, SFTP**) that the FTP client supports, the more secure it becomes.

- *FTPS*
  - The fastest and the most widely implemented option is FTPS, or FTP over SSL. FTPS secures files being transmitted through FTP with **transport socket layer security (TLS).** TLS is also sometimes referred to as SSL (it predecessor Secure Sockets Layer). FTPS requires separate control and transmission channels.

- *SFTP*
  - SFTP (SSH File Transfer Protocol) provides file transfer and administration over a single channel (typically the **SSH-2 protocol {TCP port 22}**). It includes some extra features that allows resuming interrupted file transfers and the ability to remove files remotely. SFTP expects the underlying protocol (like SSH) to provide authentication and security.

TELNET

- Telnet is a **client-server protocol** based on **character-oriented** data exchange over TCP connections.
- Telnet enables remote control of computers via **text-based inputs and outputs.**
- A client-server connection is established as a default via the TCP protocol and port 23, where the remote-controlled device acts as a server and waits for commands.

- The Telnet client, the controlling instance in this process (**also referred to as remote access or login**), can be installed on a particular device, as well as on an ordinary computer. However, the presentation of the transmitted information differs, depending on the device.
- This protocol can also be used to manage applications that do not have a graphical interface.

Use of TELNET

- One of the main reasons for developing the remote protocol was that computer systems at that time were still really expensive and not easily accessible for everyone.
- Another reason was their extreme size, which meant that they were bound to specific locations. In order to make computer resources available at **universities and companies**, Unix was developed at the end of the 1960s as a suitable operating system and **Telnet as an appropriate protocol service.**
- This meant that any user that had the right authorization, could **start, manage, and use applications** on the powerful large computers for their own personal use.

- **Access to databases**
- **Interaction with programs on application servers**
- **Administration of networks and servers**

The Telnet standard commands

- telnet hostname

- the connection setup is started, which is completed after the username and password have been entered. The prerequisite for using Telnet is that the control device has user recognition. The entry hostname represents the device name in the network, and you can also specify the correct IP address and connection request through the desired port:

- telnet ip-address port number

Command	Description
?	Calls up the help menu
close	Ends the Telnet session
display argument	Displays the various parameters for the current connection (port, terminal type, etc.)
environ argument	Defines variables for the respective operating system environment
logout	Ends the current Telnet session as long as the remote host supports the logout option
mode type	Specifies the transmission type (text file, binary file)
open hostname	Builds an additional connection to the selected host on top of the existing connection
quit	Ends the Telnet client connection including all active connections
send argument	Sends selected, typical Telnet character strings to the host
set argument	Changes the connection parameters
unset	Loads the pre-defined connection parameters

The advantages and disadvantages of the Telnet protocol

Advantages	Disadvantages
Telnet client is versatile	Unencrypted data exchange
Can be used cross-platform	Full access makes it easier for hackers
Unlimited access to target resources	Only few servers can be reached via Telnet

File Transfer Protocol

- To Transfer files between two computers, an application layer protocol is used, File Transfer Protocol (FTP). It is a standard, interactive, connection-oriented protocol based on client-server architecture that relies on TCP for transferring files.

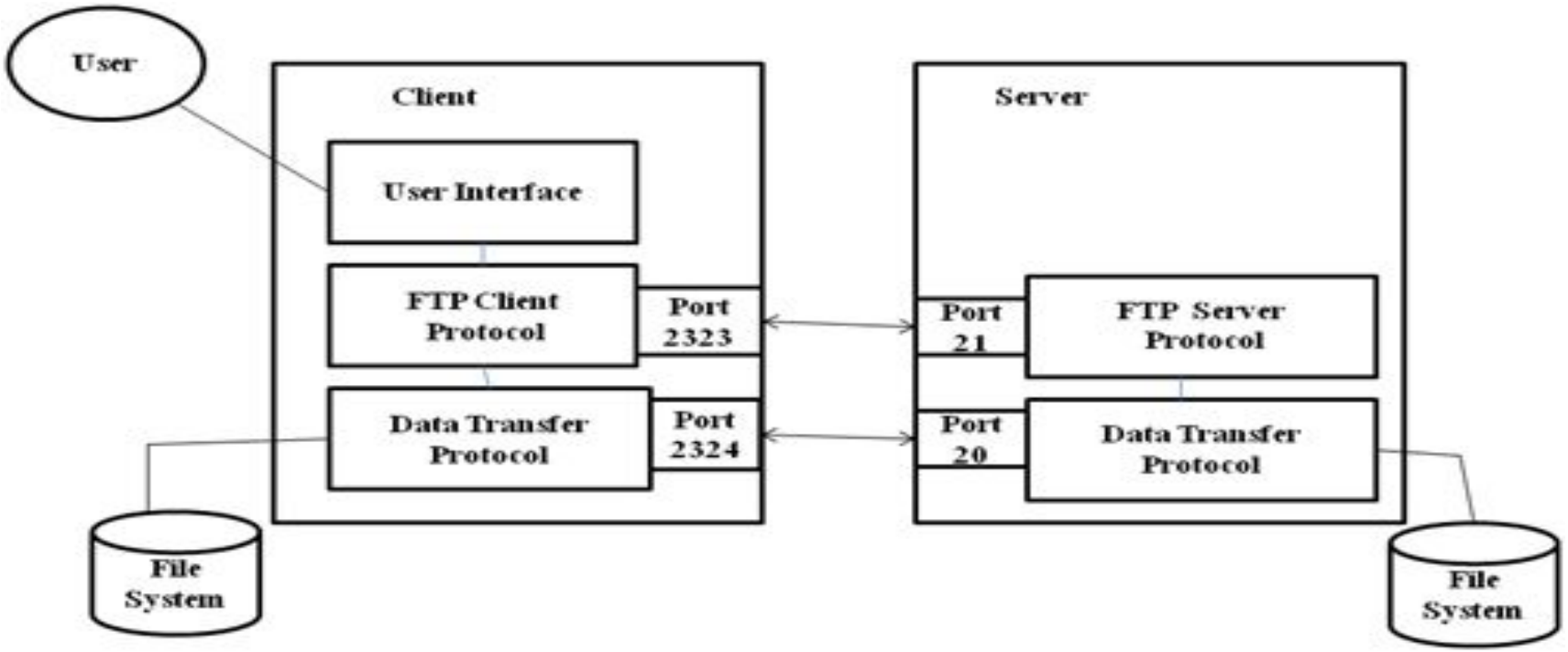
- It's not really transfer, as in moving a file from one location to another. It's "file copying" -- copying files from one computer to another. FTP is the standard mechanism provided by TCP/IP for copying a file from one host to another.

Some Problems to deal with

- Although transferring files from one system to another seems simple and straightforward, some problems must be dealt with first.

  - 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 of these problems have been solved by FTP in a very simple and elegant approach. FTP addresses and resolves these problems related to heterogeneous systems that use different operating systems, character sets, naming conventions, directory structures, file structures and formats.



FTP Client Commands

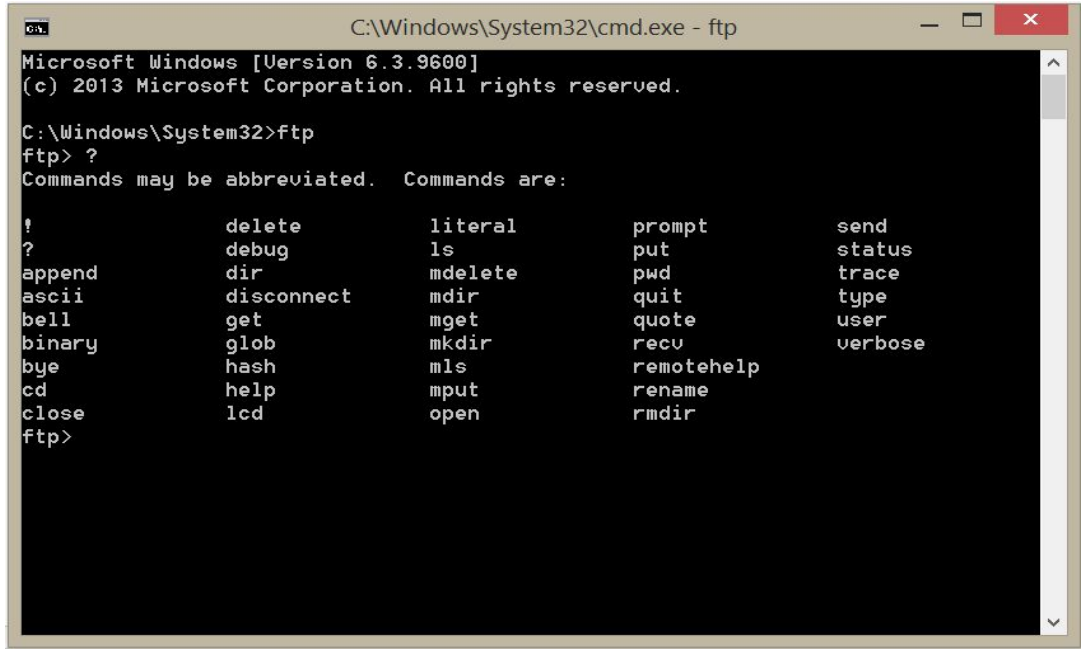
To use FTP, you need:

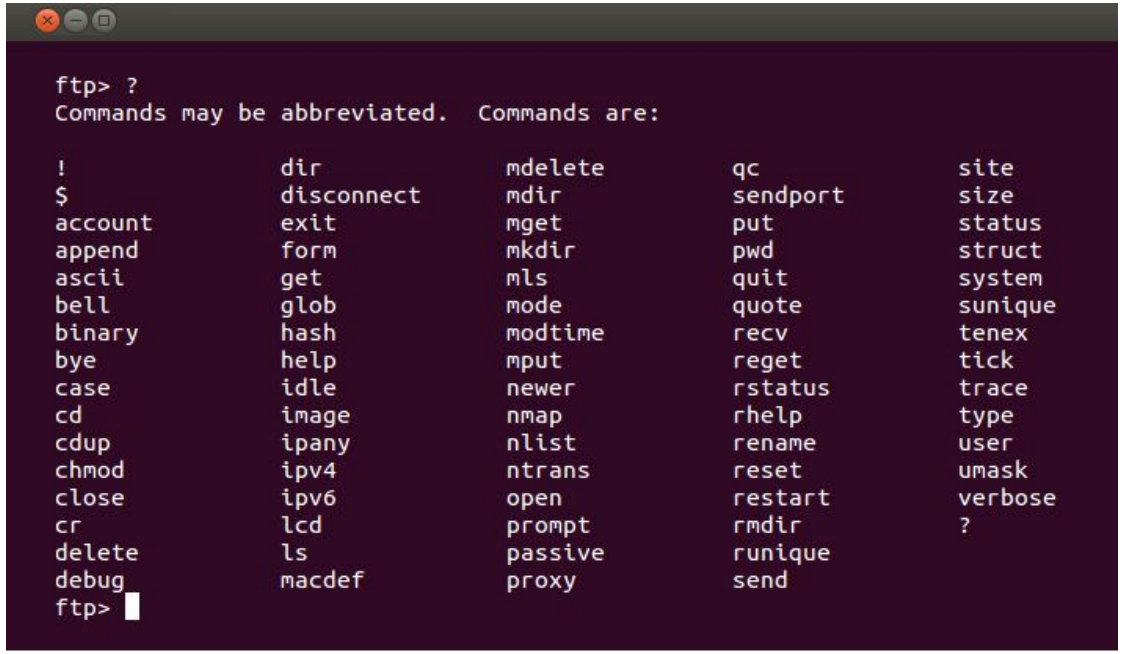
- FTP client software
- FTP server.
- Server address
- Username
- Password
- Port number

Login Information	Definition
site:ftp.example.com	This is the site address of the FTP server you're connecting to.
login: abc.xyz	Login or the USER command is the username used for logging into FTP.
pass: P@ssworD	PASS is the password.
port:21	PORT is the COMMAND port number you are using to connect to the server. The most common port number is port 21.

Built-in Commands

Windows®, Mac OS® X, and Linux® operating systems have built-in command-line clients that can be used for establishing an FTP connection. To initiate an FTP connection from Windows, type ftp at the command prompt, and press enter.





FTP Connection Types

FTP connects using two TCP ports for all communications between the server and user.

- ***COMMAND Port:*** This is the main TCP port created upon a session is connected. It is used for passing commands and replies. Port 21 (unsecured) or 990 (secured) are the default command ports used.
- ***DATA Port:*** Each time when files or directories are transferred between server and client, a random TCP data connection is established and data transfer commences over the connection. Once data transfer is complete, the connection is closed. Subsequent data connections are established and terminated as required. Data connections are never left open.

FTP Modes

Connection Modes – ASCII and Binary

FTP transfers files between systems by using one of these two modes – ASCII and binary. The mode is determined at the initial stage of all FTP transactions by the server. The FTP client will automatically switch to the mode. ASCII mode is used exclusively to transfer text and HTML. Binary mode transfers zip files, images or executable files in binary form. Binary files cannot be sent via ASCII mode and vice versa as corruptions will occur.

Transfer Modes – Passive and Active

Active and passive are the two modes that FTP can run in. FTP uses two channels between client and server as described previously, the command channel and the data channel. The command channel is for commands and responses; the data channel is for actually transferring files. During the address/port negotiation phase, the client should issue either the PORT command (when initiating Active Mode) or the PASV command (when initiating Passive Mode).

Active Mode

- The client issues a PORT command to the server signaling that it will “actively” provide an IP and port number to open the Data Connection back to the client.
- Client opens up command channel from client port 2000 to server port 21.
- Client sends PORT 2001 to server and server acknowledges on command channel.
- Server opens up data channel from server port 20 to client port 2001.
- Client acknowledges on data channel.

Passive Mode

- The client issues a PASV command to indicate that it will wait “passively” for the server to supply an IP and port number, after which the client will create a Data Connection to the server.
- Client opens up command channel from client port 2000 to server port 21.
- Client sends PASV to server on command channel.
- Server sends back (on command channel) PORT 1234 after starting to listen on that port.
- Client opens up data channel from client 2001 to server port 1234.
- Server acknowledges on data channel.

Real-time User Communication

Internet Telephony: Voice/Video communication Internet

- It uses the Voice over Internet Protocol (VoIP) application layer protocol in the TCP/IP protocol stack
- Used to transmit voice over the Internet.
- The voice is first converted into digital data which is then organized into small packets.
- Packets are stamped with the destination IP address and routed over the Internet.
- At the receiving end the digital data is reconverted into voice and fed into the user's phone.
- Allows you to make phone calls over a broadband internet connection
- Some VoIP services require a computer or a dedicated VoIP phone, while others allow you to use your landline phone to place VoIP calls through a special adapter.
- Examples include Xbox Voice, Windows messenger, AOL Instant Messenger, Motorola Phone Adapter (Vonage), Cisco Phone, Skype.

VoIP Components

- Phones – End-point devices, including both analog and IP phones.
- Gateways – allows a non-VoIP (analog) device to communicate with the VoIP network, or a VoIP device to communicate with an analog network.
- Application Servers – provides required applications to VoIP phones.
- Gatekeepers – maps phone numbers to IP addresses, and grants permission for call setup
- Call Agents – handles call routing and setup.

Benefits of VoIP

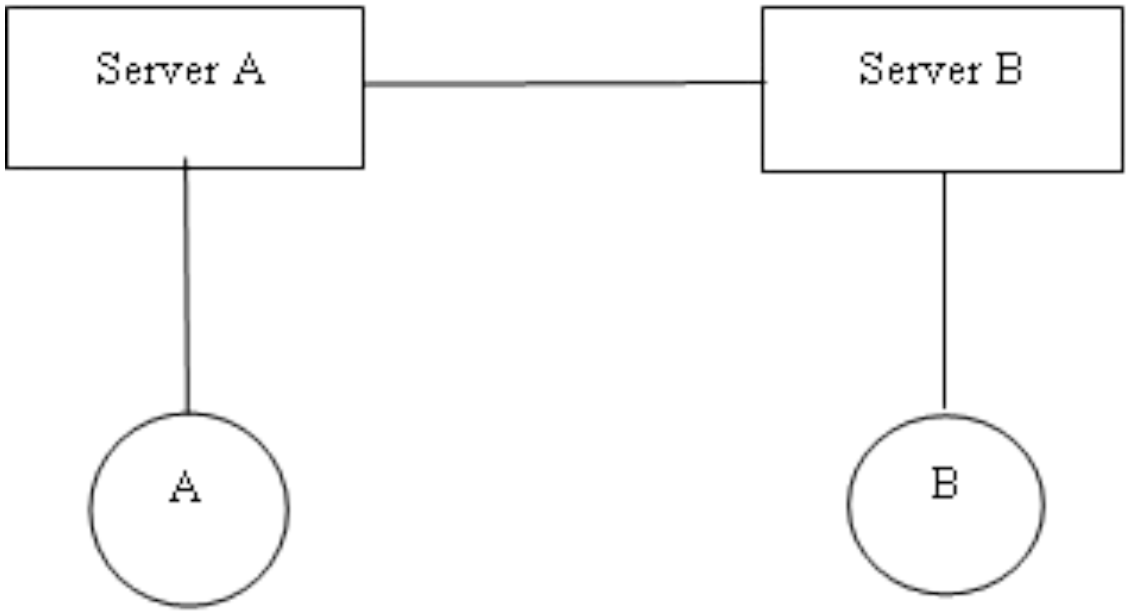
- Better use of bandwidth - Traditional voice requires a dedicated 64- Kbps circuit for each voice call, while VoIP calls can use considerably less. Additionally, no bandwidth is consumed when no call is being made.
- Single form of cabling – Reduces implementation and maintenance costs by having a standardized and consolidated cabling and equipment infrastructure.
- Cost savings from integration into the data network – Toll charges for inter-office voice communication can be avoided by routing voice traffic across existing data lines.
- Integration into devices beyond telephones.

Internet Relay Chat (IRC)

- It's a text-based interactive service using TCP/IP on the Internet to exchange text messages.
- It is a multi-user, multi-channel system teleconferencing system.
- Works on client/server architecture
- IRC clients are computer programs that a user can install on their system.
- Clients communicate with chat servers to transfer messages to other clients. IRC clients and servers communicate by sending plain ASCII *messages* to each other over TCP.
- Mainly designed for group communication in discussion forums, called channels
- Also allows one-on-one communication via private messages as well as chat and data transfer, including file sharing.
- The only configuration allowed for IRC servers is that of spanning tree where each server acts as a central node for the rest of the network it sees.

IRC Network: Steps

Locating Client: The two clients must be able to locate each other. Upon connecting to a server, a client registers using a label which is then used by other servers and clients to know where the client is located. Servers are responsible for keeping track of all the labels being used.

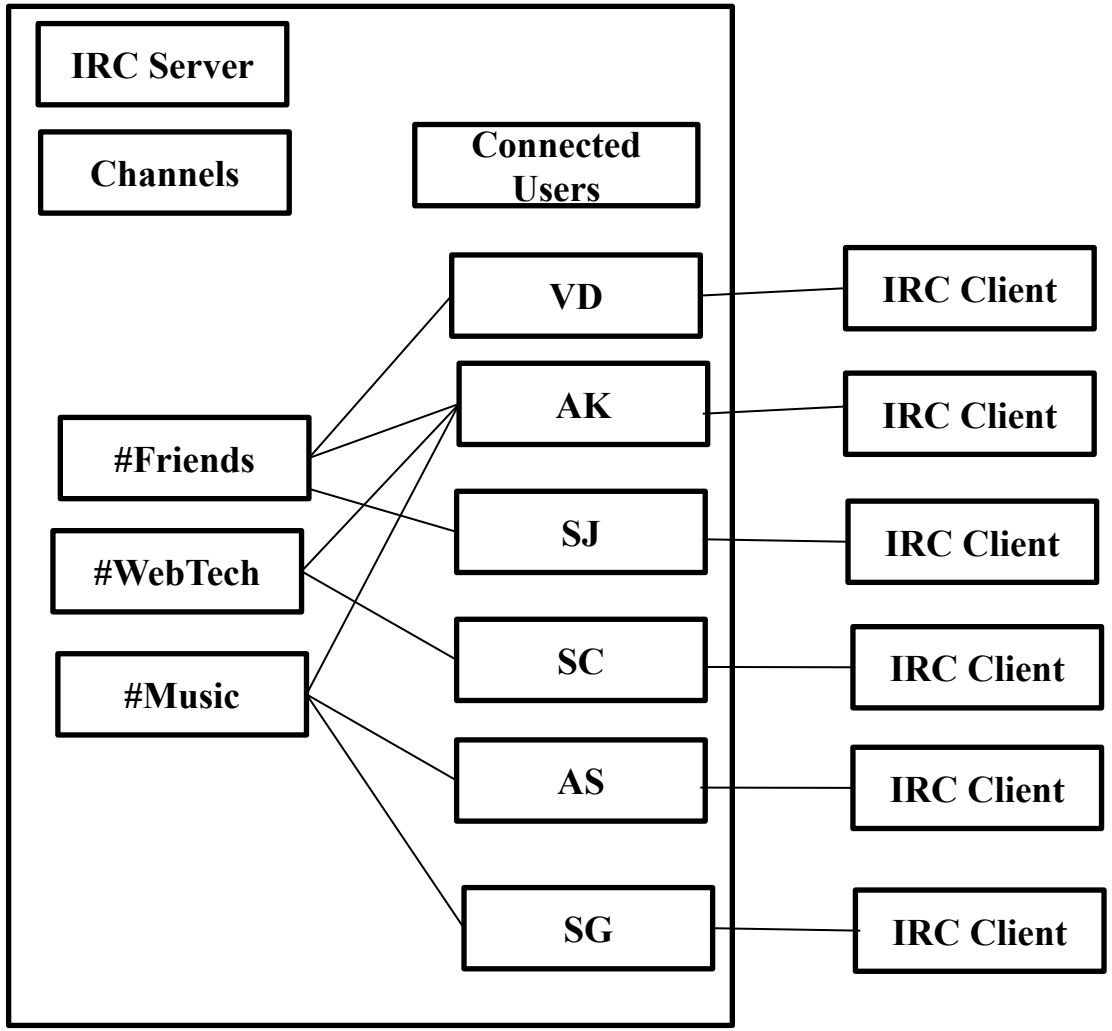


Steps

- ***Relaying Message:*** The IRC protocol provides no mean for two clients to directly communicate, all communication between clients is relayed by the server(s)

- ***Channel Hosting and* Management:** IRC supports the creation of chat rooms (discussion forums) called *channels*. A channel is a named group of one or more users which will all receive messages addressed to that channel. Channels provide a mean for a message to be sent to several clients. Servers host channels, providing the necessary message multiplexing. Server are also responsible for managing channels by keeping track of the channel members. The messages in an Internet Relay Chat may be delivered either using a one-to-one communication, one-to-many or one-to-all.

Architecture



Architectural Issues of IRC

- **Scalability:** This protocol does not scale sufficiently well when used in a large arena.

- **Reliability:** As the only network configuration allowed for IRC servers is that of a spanning tree, each link between two servers is an obvious and quite serious point of failure.

- **Network Congestion:** Due to the spanning tree architecture the IRC protocol is extremely vulnerable to network congestions. This problem is endemic (local), and should be solved for the next generation, i.e., if congestion and high traffic volume cause a link between two servers to fail, not only this failure generates more network traffic, but the reconnection (eventually elsewhere) of two servers also generates more traffic.

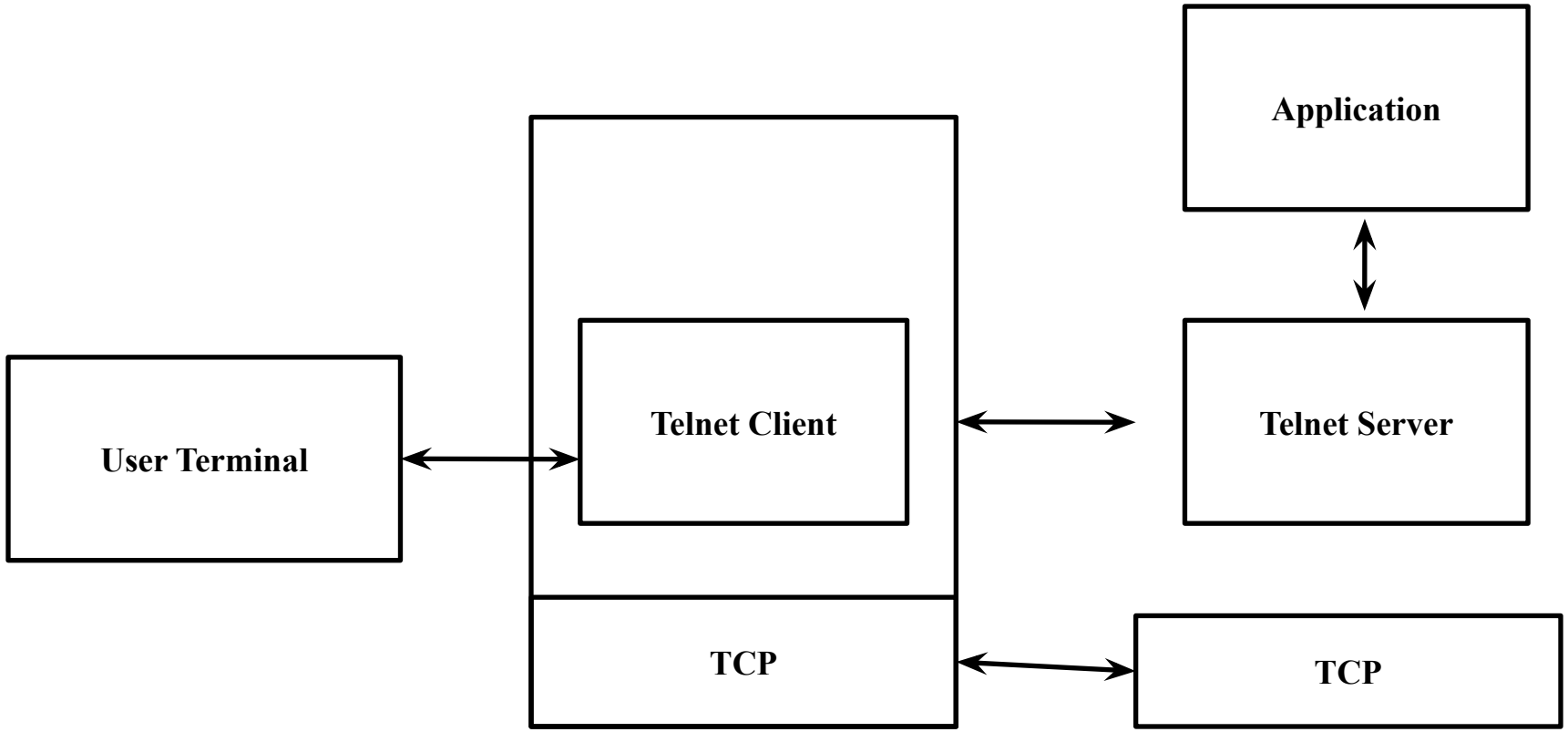
- **Privacy:** Besides not scaling well, the fact that servers need to know all information about other entities, the issue of privacy is also a concern.

Telnet

- A standard TCP/IP protocol for virtual terminal service
- 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
- Abbreviation for TErminaLNETwork.
- Offers a user the possibility to connect and log on to any other hosts in the network from user's own computer by offering a remote log on capability.
- Telnet uses the TCP protocol.
- The Telnet service is offered in the host machine's TCP port 23.

- The user at the terminal interacts with the local telnet client.
- The Telnet client acts as a terminal accepting any keystrokes from the keyboard, interpreting them and displaying the output on the screen
- The client on the computer makes the TCP connection to the host machine's port 23 where the Telnet server answers.
- The Telnet server interacts with applications in the host machine and assists in the terminal emulation.
- As the connection is setup, the both ends of the Telnet connection are assumed to be originated and terminated at the network virtual terminal (NVT). The NVT is a network wide terminal which is host independent so that both the server and the client in the connection may not need to keep any information about each other's terminal's characteristics as both sees each other as a NVT terminal.
- The Telnet has a set of options and these options can be negotiated through a simple protocol inside the Telnet. The negotiation protocol contains commands DO, WILL, WON'T and DON'T.

Telnet Model



Archie

To download or upload files, you use the File Transfer Protocol (FTP). But how do you find the files you want to transfer?

One way is to use a client program called Archie. Telnet, FTP, and Archie are interrelated. The Archie database is made up of the file directories from hundreds of systems. When you search this database on the basis of a file's name, Archie can tell you which directory paths on which systems hold a copy of the file you want. To use Archie, you must Telnet to an Archie server. You can do that by keying in a command such as **telnet://archie.internic.net** to get to the Archie server at that address and log on by keying in archie when prompted to do so. Once you do your Archie search, you must then go get the file using FTP, the Internet File Transfer Protocol.

Usenet

- Acronym for USErNETwork.
- It’s a way of sharing information.
- Initially it was limited to two sites but today there are thousands of Usenet sites involving millions of people.
- Usenet is a globally distributed discussion group where people can share views on topic of their interest. It is a collection of special interest groups, called newsgroups.
- Each newsgroup is devoted to a certain topic. Under each newsgroup, there are many messages called news articles. The article posted to a newsgroup becomes available to all readers of the newsgroup.
- The Network News Transport Protocol (NNTP) is used in transferring news articles between news clients and news servers. NNTP is the protocol that is used for posting/ distributing/ retrieving USENET news articles among news server.

Step-by-Step Working

- You use a news reader to read news, post news, follow-up a piece of news, etc.
- Your newsreader interrogates with a news server.
- A news server negotiates with other servers to transfer certain newsgroups between each other. A news server holds the news articles for a certain pre-set period (controlled by the server's administrator) and eventually discards them at their expiry date.

Basic Terminology used

- *News administrator* – A person who is in charge of running a news server.
- *News server* – A computer that saves, forwards and manages news articles. Usually, a news server is running in one domain.
- *News reader* – A program that allows user to read/post/subscribe/unsubscribe a newsgroup.
- *Newsgroup* – An on-line forum that allows users from the Internet to join the discussion on a specific topic.
- *Usenet* – A collection of newsgroups.

Newsgroup Classification

- The articles that users post to Usenet are organized into topical categories called newsgroups, which are logically organized into hierarchies of subjects. For instance, sci.math and sci.physics are within the *sci.** hierarchy, for science.

- There exist a number of newsgroups distributed all around the world. These are identified using a hierarchical naming system in which each newsgroup is assigned a unique name that consists of alphabetic strings separated by periods. The leftmost portion of the name represents the top-level category of the newsgroup followed by subtopic.

- The major set of worldwide newsgroups is contained within nine hierarchies, eight of which are operated under consensual guidelines that govern their administration and naming.

Current Big Eight

- *comp.** – computer-related discussions (*comp.software*, *comp.sys*)
- *humanities.** – fine-arts, literature, and philosophy (humanities.*classics, humanities.design. misc*)
- *misc.** – miscellaneous topics (*misc.education*, *misc.forsale*, *misc.kids*)
- *news.** – discussions and announcements about news (meaning Usenet, not current events) (*news.groups*, *news.admin*)
- *rec.** – recreation and entertainment (*rec.music*, *rec.arts.movies*)
- *sci.** – science related discussions (*sci.psychology*, *sci.research*)
- *soc.** – social discussions (*soc.college.org*, *soc.culture.african*)
- *talk.** – talk about various controversial topics (*talk.religion*, *talk.politics*, *talk.origins*)

World Wide Web

- The services provided by the Internet can be classified into two categories, namely, the communication services & the information services.
- The most important communication services on the Internet are electronic mail (and some derived services) and Usenet
- Major information services are file transfer, telnet and World Wide Web (WWW).
- The WWW offers a view on one virtually unified but decentralized information space. The Web is the most sophisticated and most exciting new Internet service. It is essentially a huge client-server system with millions of servers distributed worldwide. Each server maintains a collection of documents; each document is stored as a file.
- All communication in the Web between clients and servers is based on the Hypertext Transfer Protocol (HTTP). HTTP is a relatively simple client-server protocol; a client sends a request message to a server and waits for a response message.