

Module 1:

INTRODUCTION TO INTERNET

Dr. L.M. Jenila Livingston

Professor

VIT Chennai

Web based Information System

- Disseminating all kind of Information to anybody, anytime, anywhere or to all, always, all over world
- 24/7 availability creates high impact in all fields
- Online teaching and learning,
- Virtual office,
- Marketing, selling, getting customer reviews,
- e-banking, payment portals,
- e-shopping
- ticket reservations,
- e hospitals etc.



Website Current Trends and Needs

- Websites..
 - Customized websites (Amazon ,Flipkart, based on search)
 - Personalized websites (online courses based on area of interest, profile)
 - Interactive websites (Games, Designing product eg Cyclemen)
 - Navigational websites (Google Map, Google earth, Open Streetview)
 - Regional Language (Tamil, Hindi...)
- Platforms – Desktop, Phone, Tablets
- Browsers – Google, Internet Explorer, Mozilla with various versions

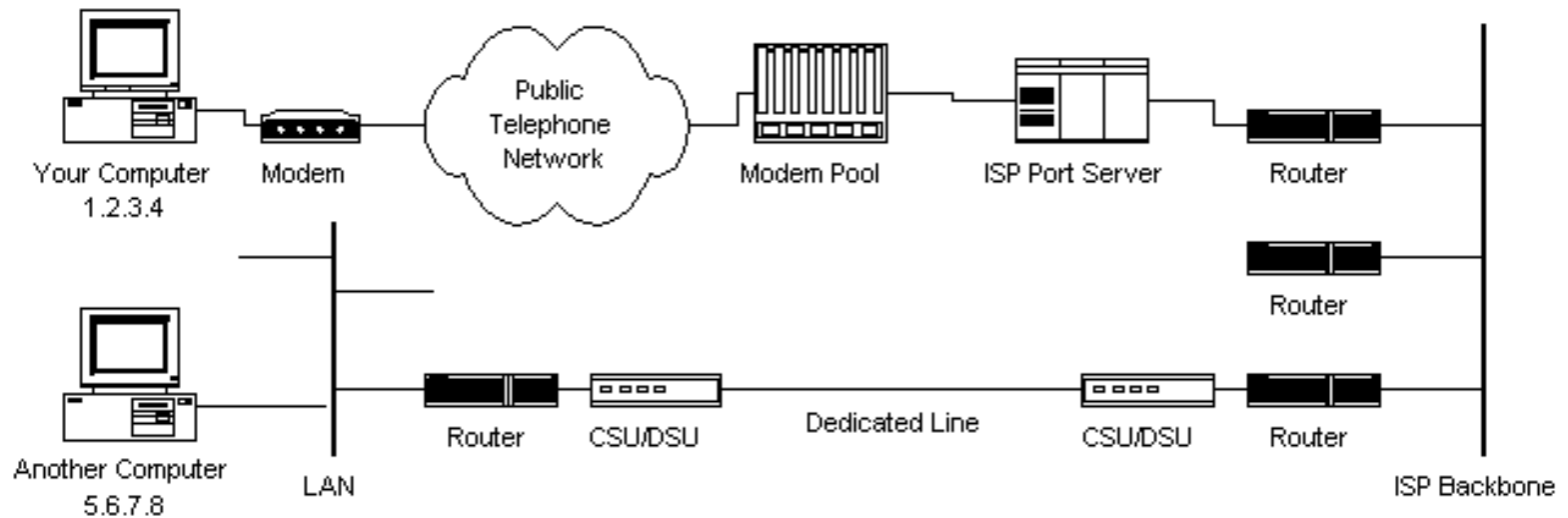
Internet

- Internet - hardware parts

Computers, Laptops, Mobiles through wired or wireless connections.

- Web - software parts

Websites connected through URL addresses.



Internet Vs Intranet

INTERNET

Public network

In internet, there are multiple users.

Internet is unsafe.

Internet provides unlimited information.

Unlimited Traffic

INTRANET

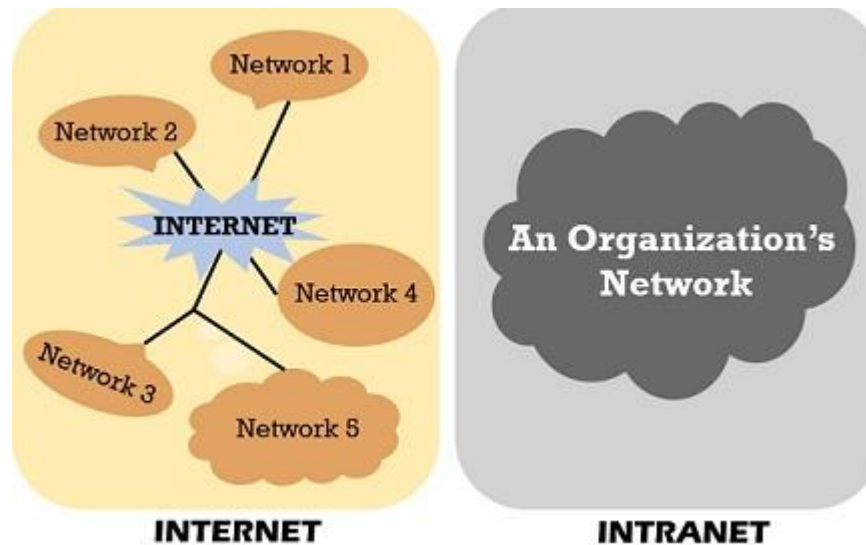
Private network

In intranet, there are limited users.

Intranet is safe.

Intranet provides limited information.

Traffic is limited



Web vs Internet

Internet	Web
Internet is network of computers	World wide Web is networks of websites
Transfer the contents	Access the information
Layered System of Protocols: 1. Application Layer – Data transfer 2. Transport Layer - Communication 3. Network Layer – Route	To link with the internet, 1. HTML, CSS, Scripting Languages – webpage creation 2. HTTP – Transfer Protocol 3. URL – Address of Resource. 4. Server- Data Storage 5. Client – Web browsers

Network

- All machines will receive short messages, called packets
- Types of Network
 - LAN - single office/ building / campus.
 - MAN - entire city / company connects its offices in a city.
 - WAN - country / continent /whole world.

Sockets

- Sockets, or ports, are a very low level software construct that allows computers to talk to one another
- When you send information from one computer to another, you send it to a port on the receiving computer

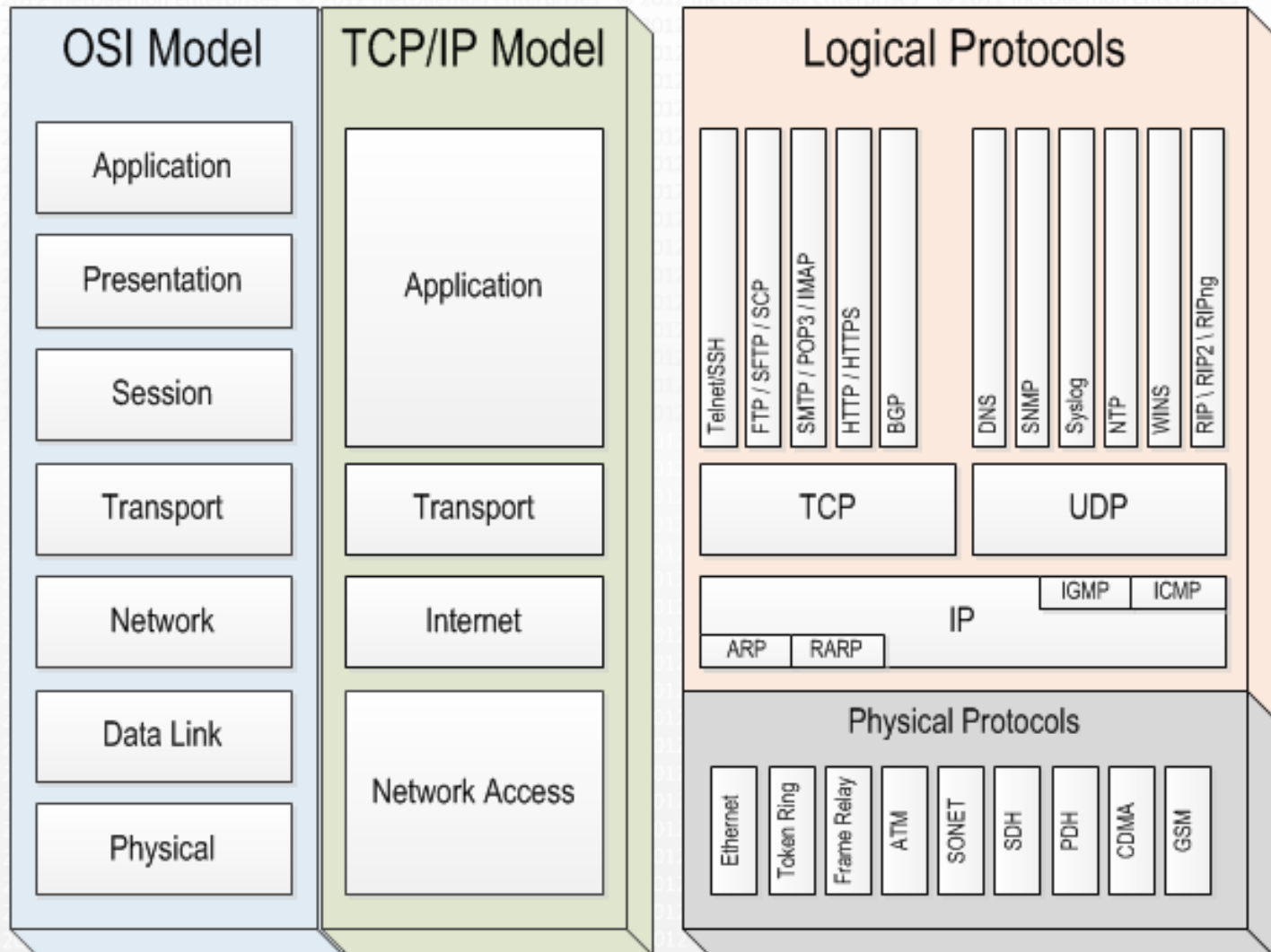
Services – Port Numbers

- Common port numbers are **80** (for web pages), **23** (for telnet) and **25** and **110** (for mail)
- Port numbers above **1024** are available for other kinds of communication between our programs
- Common (well known) ports:
 - FTP 21
 - TELNET 23
 - SSH 22
 - SMTP 25
 - HTTP 80
 - DNS – 53
 - NTP 123
 - Finger 79
 - DHCP – 67 & 68
 - POP – 110

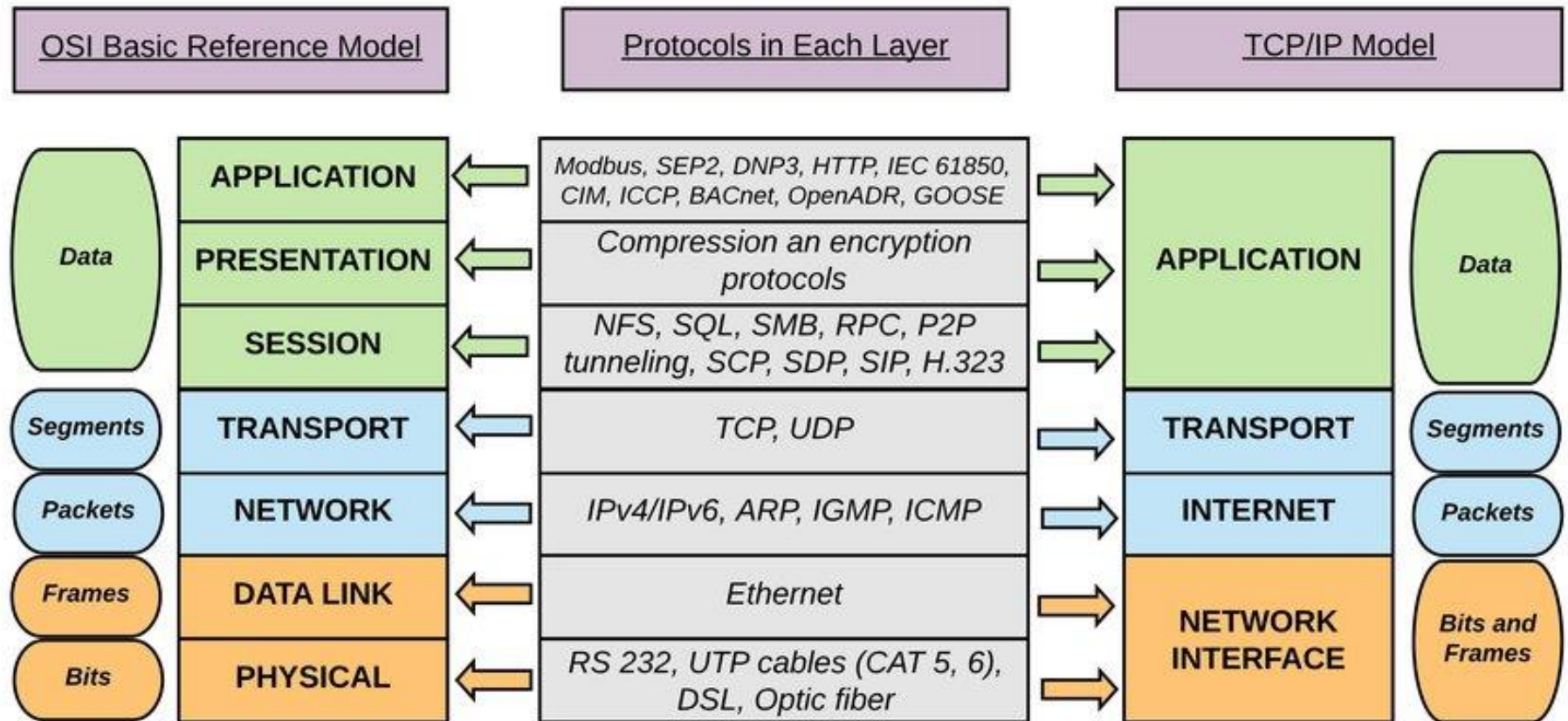
Protocols

- In order for computers to communicate with one another, they must agree on a set of rules for who says what, when they say it, and what format they say it in
- This set of rules is a **protocol**
- Different programs can use different protocols
- Protocols may be in ASCII (characters) or in binary
- Some common protocols are **HTTP** (for web pages), **FTP** (for file transfer), and **SMTP** (Simple Mail Transfer Protocol)

NETWORK MODELS



OSI Vs TCP/IP



Application Type	Application-layer protocol	Transport Protocol
Electronic mail	Send: Simple Mail Transfer Protocol SMTP [RFC 821]	TCP 25
	Receive: Post Office Protocol v3 POP3 [RFC 1939]	TCP 110
Remote terminal access	Telnet [RFC 854]	TCP 23
World Wide Web (WWW)	HyperText Transfer Protocol 1.1 HTTP 1.1 [RFC 2068]	TCP 80
File Transfer	File Transfer Protocol FTP [RFC 959]	TCP 21
	Trivial File Transfer Protocol TFTP [RFC 1350]	UDP 69
Remote file server	NFS [McKusik 1996]	UDP or TCP
Streaming multimedia	Proprietary (e.g., Real Networks)	UDP or TCP
Internet telephony	Proprietary (e.g., Vocaltec)	Usually UDP

Application Layer Protocols	Usage
HTTP: Hyper Text Transport Protocol	Transfer web pages
HTTPS: Hyper Text Transport Protocol Secure	
DNS: Domain Name System	Translates website names to IP addresses
NBNS: NetBios Name Service	Translates localhost names to IP addresses
DHCP: Dynamic Host Configuration Protocol	Assigns IP addresses
TelNet: Telephone Network	Interactive login; communication between computers
FTP: File Transfer Protocol	Commonly used for copying files to and from other computers
TFTP: Trivial File Transfer Protocol	Transfer small amount of data
SMTP: Simple Mail Transport Protocol	Sends email messages
POP3: Post office protocol	Extracts and retrieves email
IMAP: Internet Mail Access Protocol	Access email on a remote mail server
SNMP: Simple Network Management Protocol	Manages Network devices
	clock synchronization between computer systems over packet-switched
NTP: Simple Network Time Protocol	Provides time of day

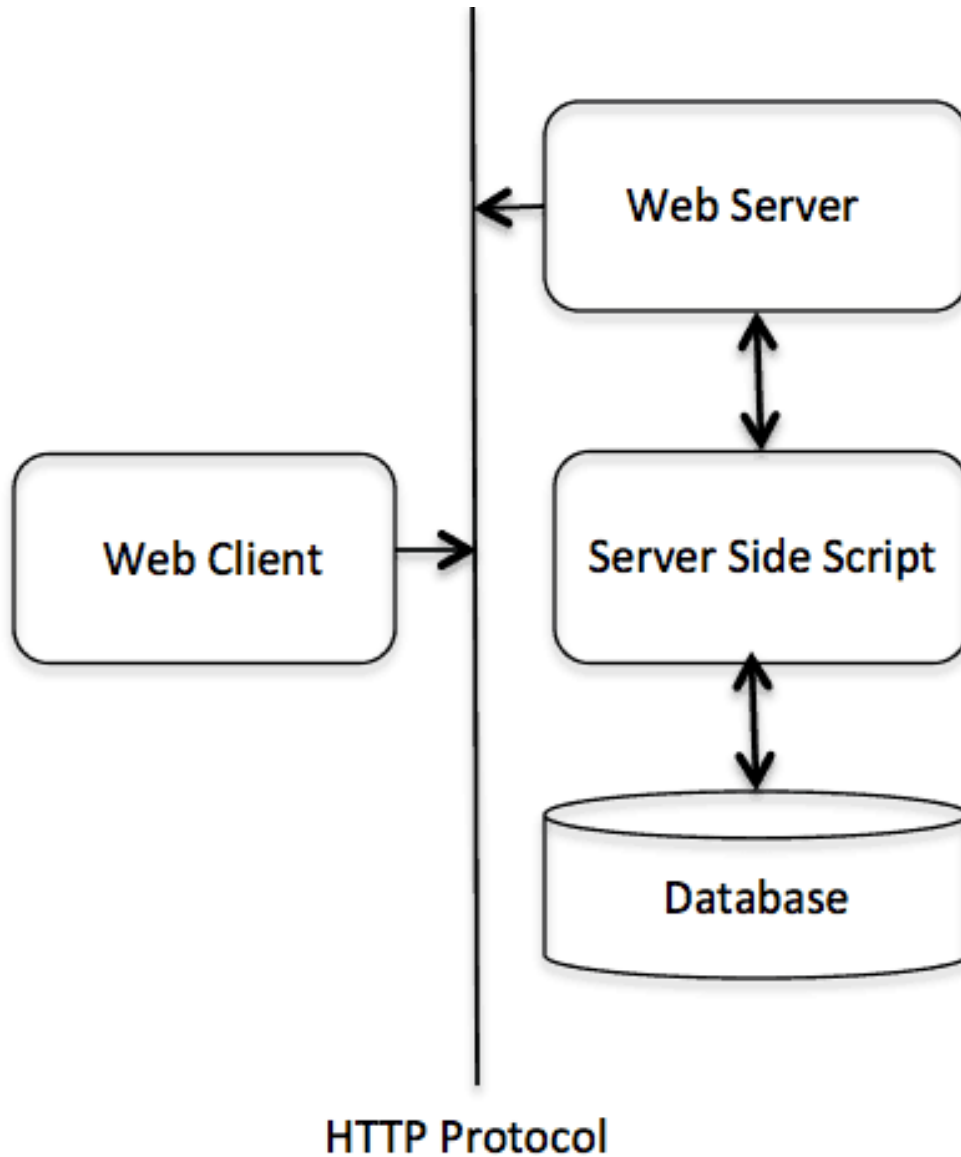
HTTP

- The Hypertext Transfer Protocol (HTTP) is an application-level protocol for distributed, collaborative, hypermedia information systems.
- This is the foundation for data communication for the World Wide Web (i.e. internet) since 1990.



HTTP Features

- **HTTP is connectionless :** The HTTP client initiates an HTTP request and after a request is made, the client disconnects from the server and waits for a response. The server processes the request and re-establishes the connection with the client to send a response back.
- **HTTP is media independent:** It means, any type of data can be sent by HTTP as long as both the client and the server know how to handle the data content.
- **HTTP is stateless:** The server and client are aware of each other only during a current request. Afterwards, both of them forget about each other.



HTTP Client Server

■ Client

The HTTP client sends a request to the server in the form of a **request method**, URI, and protocol version, followed by a MIME-like message containing request modifiers, client information, and possible body content over a **TCP/IP** connection.

■ Server

The HTTP server **responds** with a status line, including the message's protocol version and a success or error code, followed by a MIME-like message containing server information, entity meta information, and possible entity-body content.

HTTP Messages

- HTTP makes use of the Uniform Resource Identifier (URI) to identify a given resource and to establish a connection.
- Once the connection is established, **HTTP messages** are passed in a format similar to that used by the Internet mail [RFC5322] and the Multipurpose Internet Mail Extensions (MIME) [RFC2045].
- These messages include **requests** from client to server and **responses** from server to client which will have the following format:

```
HTTP-message = <Request> | <Response> ; HTTP/1.1 messages
```

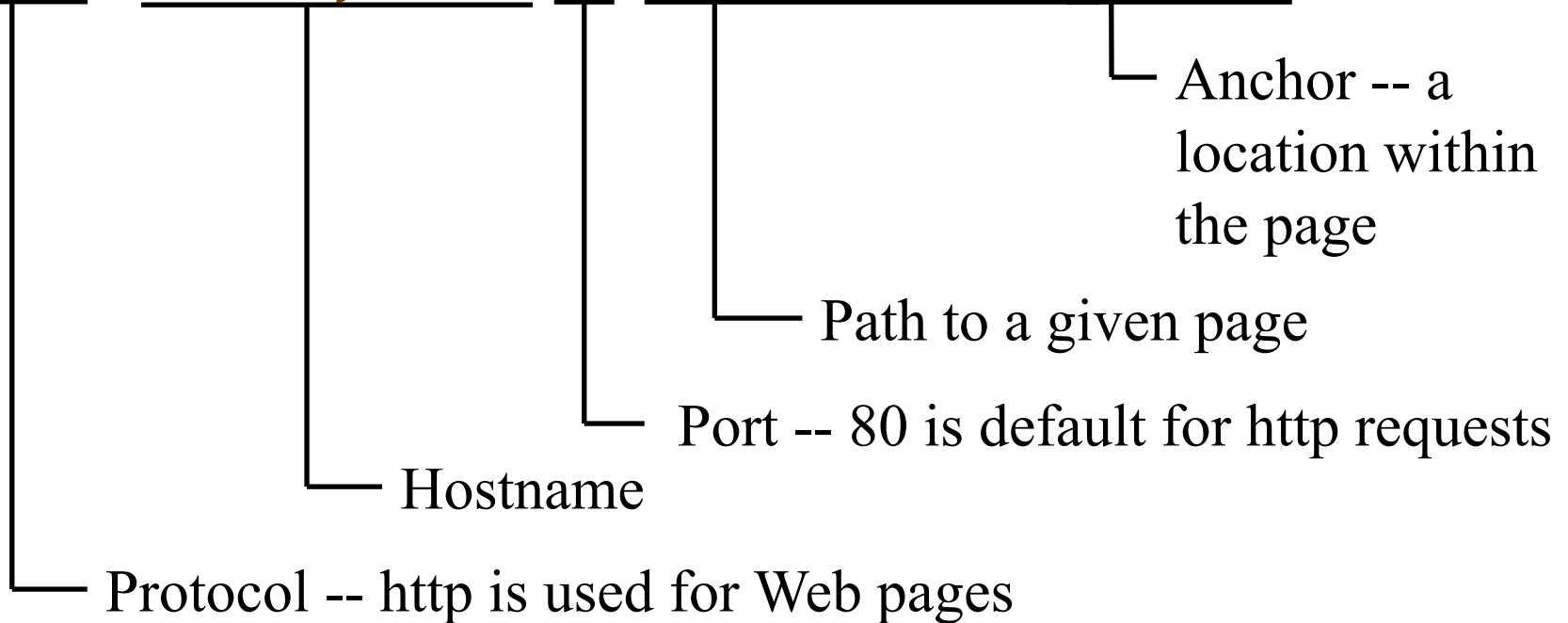
URLs

- A **URL, Uniform Resource Locator**, defines a location on the Web

```
URI = "http:" "://" host [ ":" port ] [ abs_path [ "?" query ] ]
```

- A URL has up to five parts:

http://www.xyz.com:80/ad/index.html#specials



Hostnames

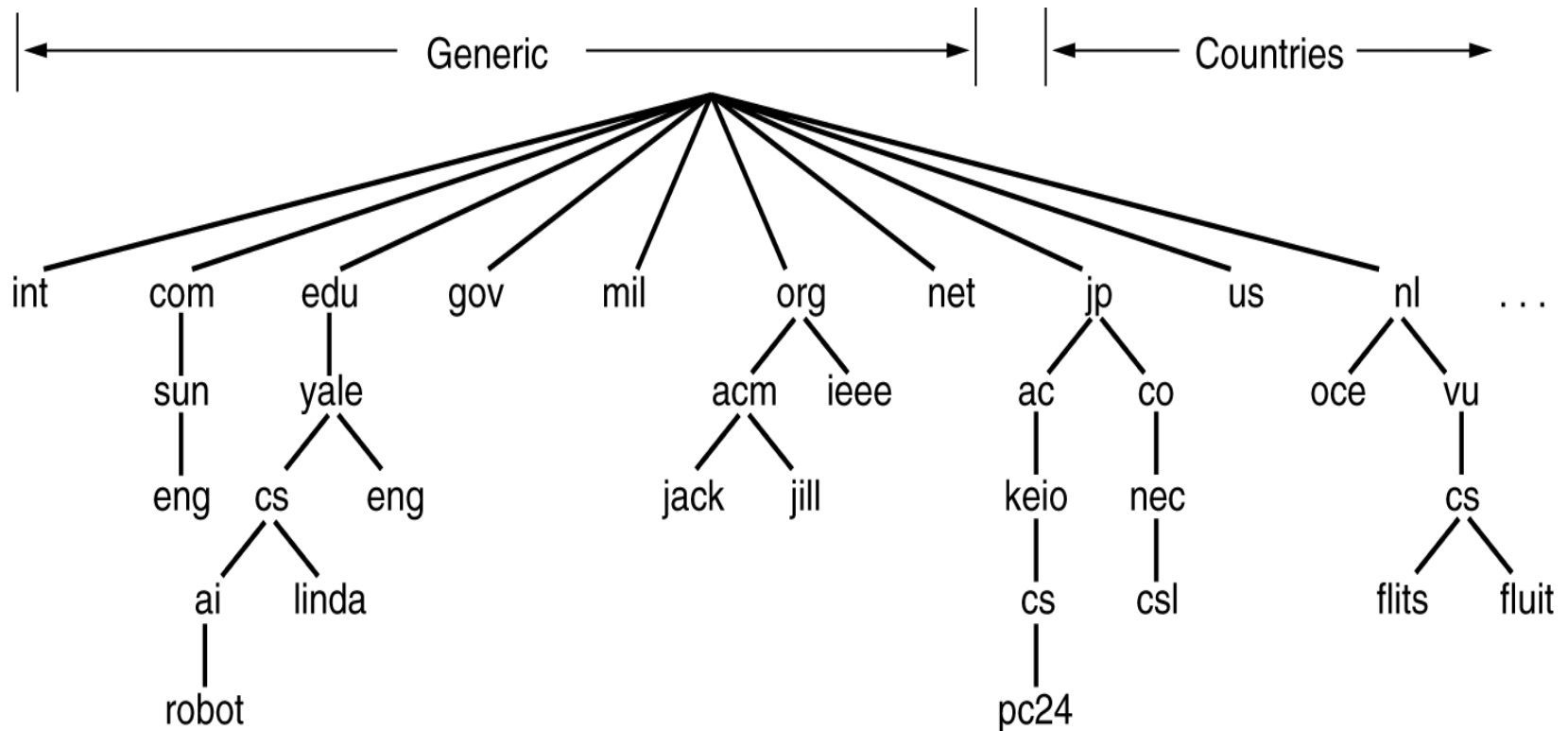
- The “real” name of a computer on the internet is its four-byte IP address
- People, however, don’t like to remember numbers, so we use **hostnames** instead
- For example, the hostname www.cis.upenn.edu is **158.130.12.9**

Domain Name System - DNS

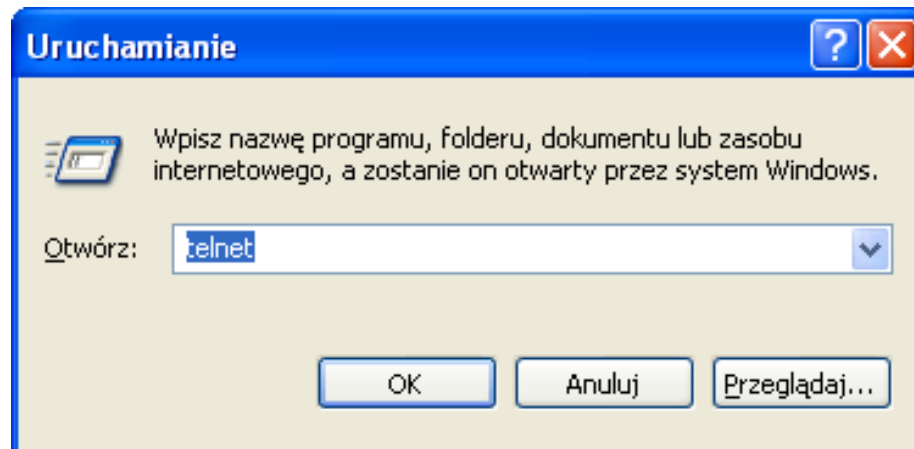
- IP addresses can be used to identify a host machine on the Internet
 - As those machines move around, the addresses need to be changed accordingly as well
- ASCII names have been used to decouple host names and their IPs to provide more flexibility
- The network itself still understands only numerical addresses
- The DNS was invented to manage and resolve host names into IP addresses

DNS Name Space

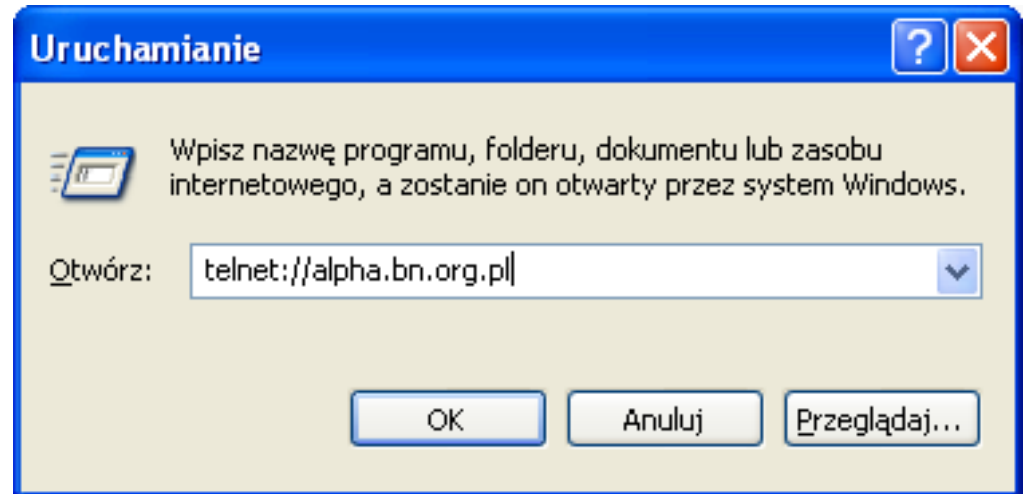
- A portion of the Internet domain name space



TELNET

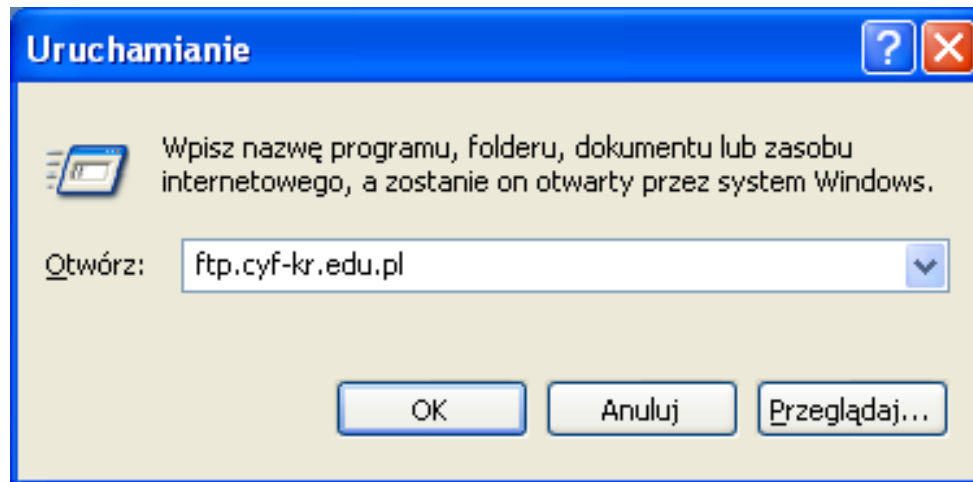
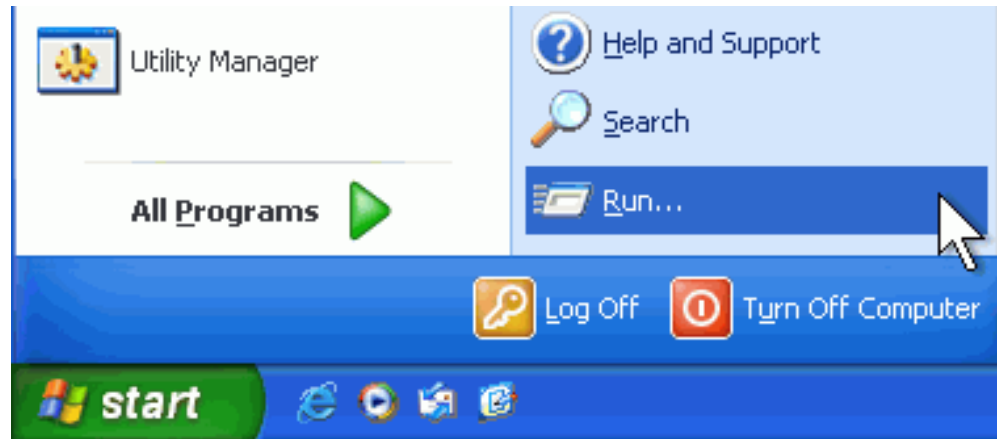


telnet://alpha.bn.org.pl

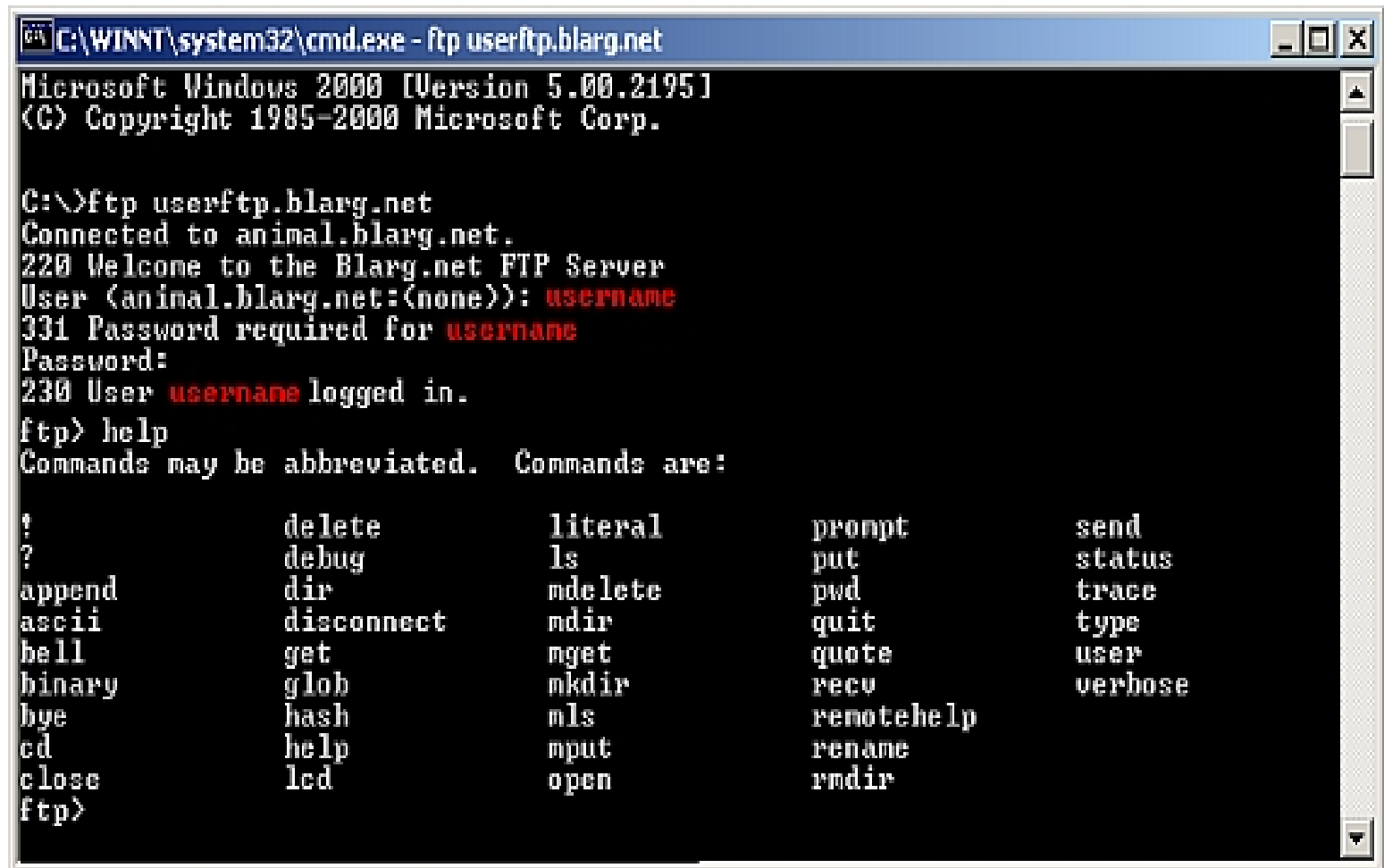


National Library

ftp.cyf-kr.edu.pl



FTP help - ?



```
C:\WINNT\system32\cmd.exe - ftp userftp.blarg.net
Microsoft Windows 2000 [Version 5.00.2195]
(C) Copyright 1985-2000 Microsoft Corp.

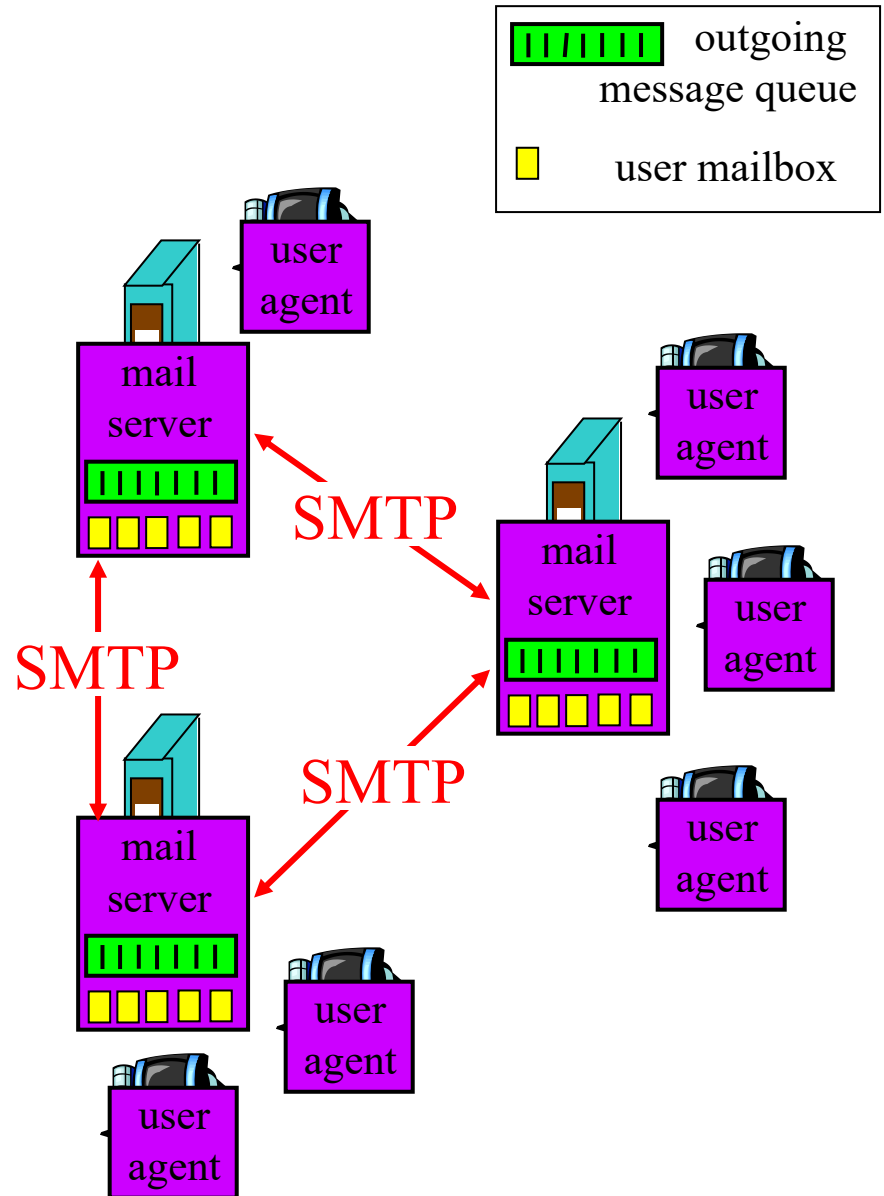
C:\>ftp userftp.blarg.net
Connected to animal.blarg.net.
220 Welcome to the Blarg.net FTP Server
User (animal.blarg.net:(none)): username
331 Password required for username
Password:
230 User username logged in.
ftp> help
Commands may be abbreviated.  Commands are:

?          delete      literal    prompt     send
?          debug       ls         put        status
append    dir         mdelete   pwd        trace
ascii     disconnect  mdir      quit       type
bell      get         nget      quote      user
binary    glob        mkdir     recv       verbose
bye       hash        nls       remotehelp
cd        help        nput      rename
close     lcd         open      rmdir
ftp>
```

Electronic Mail

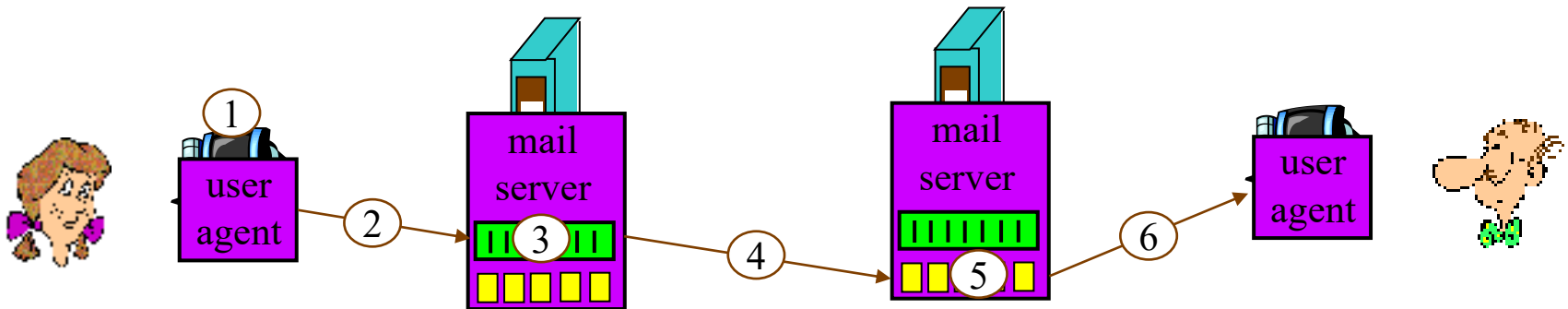
Three major components:

- user agents
 - mail servers
 - simple mail transfer protocol: SMTP
- The protocol is very simple
 - SMTP is a push protocol, information is pushed to a remote site
 - Uses port 25
 - All files are ascii text
 - Syntax: [user@host.domain](#)



Scenario: Alice sends message to Bob

- 1) Alice uses UA to compose message and “to”
`bob@someschool.edu`
- 2) Alice’s UA sends message to her mail server; message placed in message queue
- 3) Client side of SMTP opens TCP connection with Bob’s mail server
- 4) SMTP client sends Alice’s message over the TCP connection
- 5) Bob’s mail server places the message in Bob’s mailbox
- 6) Bob invokes his user agent to read message



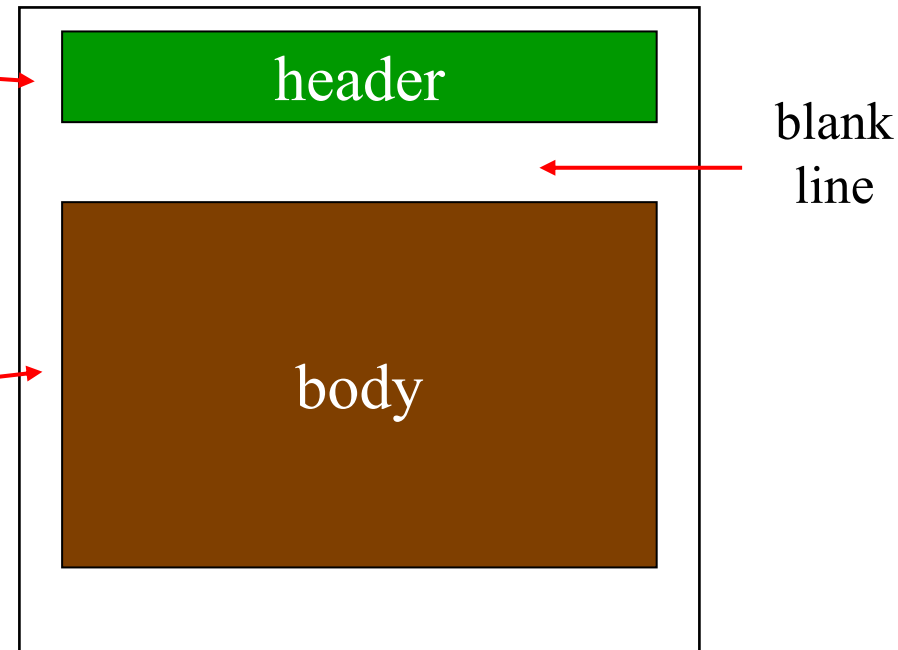
Mail message format

SMTP: protocol for exchanging email msgs

RFC 822: standard for text message format:

- header lines, e.g.,
 - To:
 - From:
 - Subject:*different from SMTP commands!*

- body
 - the “message”,
ASCII characters only

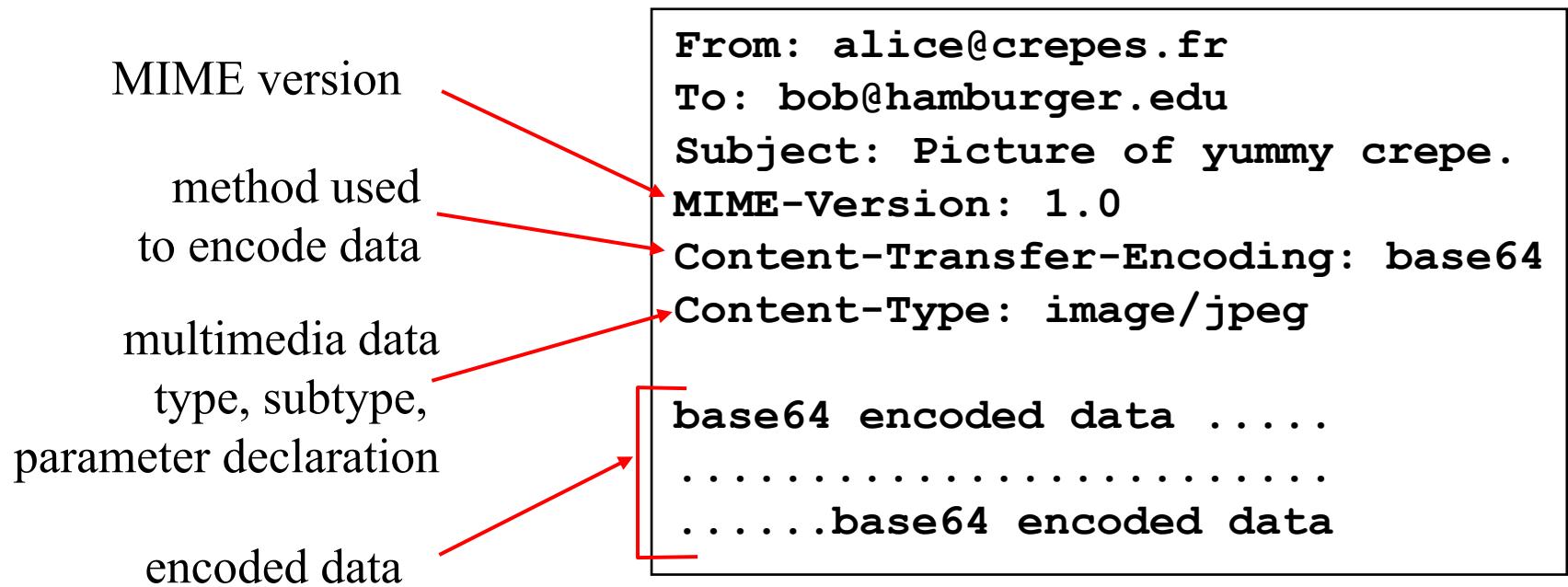


MIME Types/Encoding

- Multipurpose Internet Mail Extensions
- Handles Non-ASCII data in an ASCII transfer medium.
- Imbeds the data stream into the email message

Message format: multimedia extensions

- MIME: multimedia mail extension, RFC 2045, 2056
- additional lines in msg header declare MIME content type



POP3

- Post office protocol
- Mail access client
- Uses port 110
- Messages are downloaded to client but can be stored on server.
- Does not easily allow multiple clients

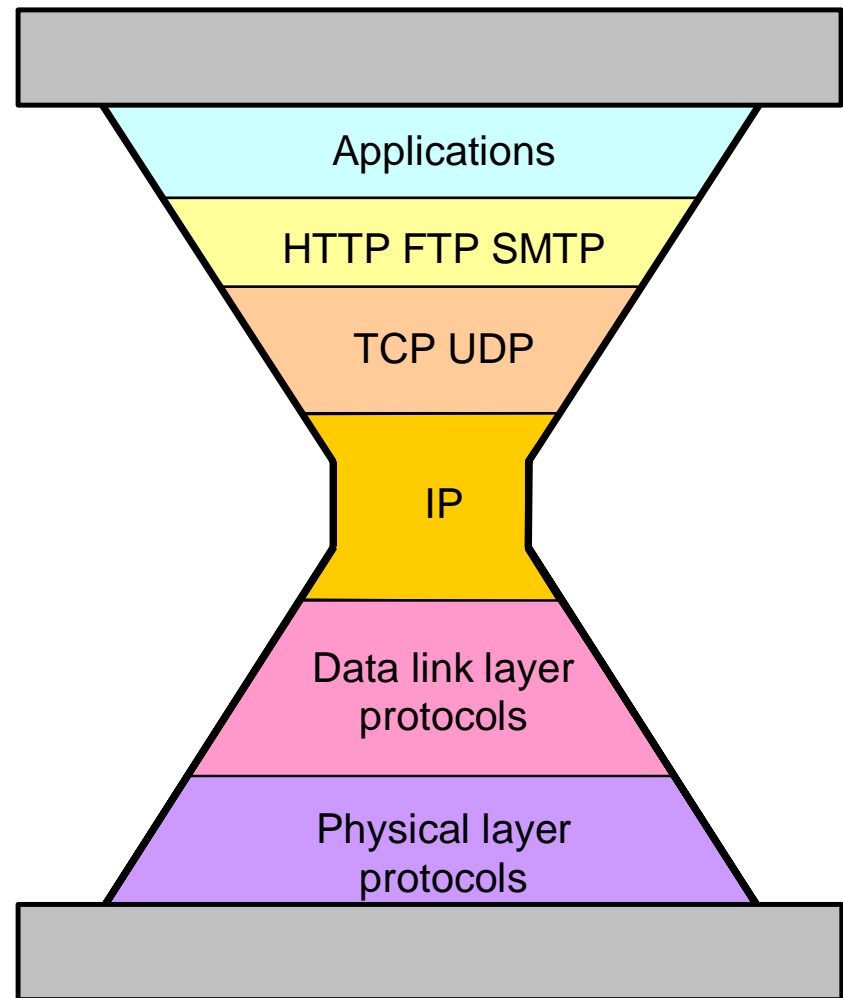
IMAP

- Internet Mail Access Protocol
- Improved POP3
- Automatically assigns folders
- Keep all messages in one place: the server
- Leaves mail on server
- Only transfers as much as needed per message (headers, subject only on list)
- Allows user to organize messages in folders
- IMAP keeps user state across sessions:
 - names of folders and mappings between message IDs and folder name

Internet Protocol

The waist of the hourglass

- IP is the waist of the hourglass of the Internet protocol architecture



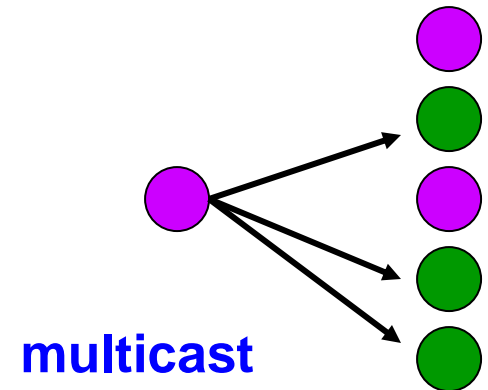
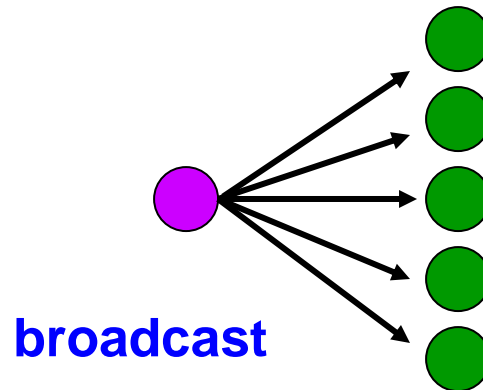
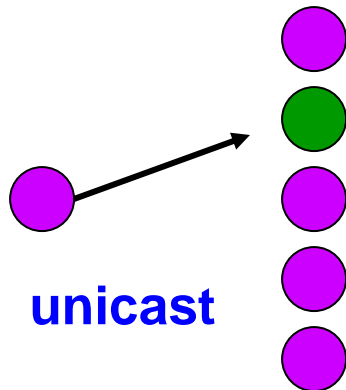
IP Service

- Delivery service of IP is minimal
- IP provide provides an **unreliable connectionless** best effort service (also called: “datagram service”).
 - **Unreliable:** IP does not make an attempt to recover lost packets
 - **Connectionless:** Each packet (“datagram”) is handled independently. IP is not aware that packets between hosts may be sent in a logical sequence
 - **Best effort:** IP does not make guarantees on the service (no throughput guarantee, no delay guarantee,...)
- Consequences:
 - Higher layer protocols have to deal with losses or with duplicate packets
 - Packets may be delivered out-of-sequence

IP Service

- IP supports the following services:

- one-to-one (unicast)
- one-to-all (broadcast)
- one-to-several (multicast)



- IP multicast also supports a many-to-many service.
- IP multicast requires support of other protocols (IGMP, multicast routing)

IP Address

- An IP address is a 32 bit sequence of 1s and 0s.
- A way to identify machines on a network
 - A unique identifier
 - A numerical label
- IP address consist of four sections, (0 to 255)
- Five classes of IP Addresses : A, B, C, D, E
 - Class A - reserved for Governments
 - Class B - reserved for medium companies
 - Class C - reserved for small companies
 - Class D - reserved for multicasting
 - Class E - reserved for future use

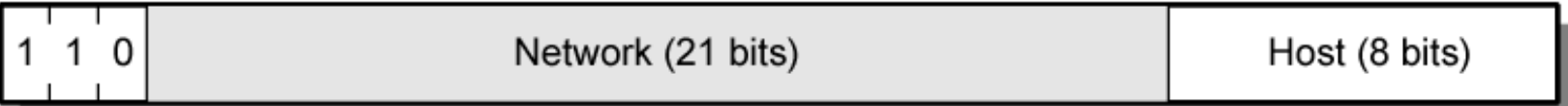
IPv4 Address Formats



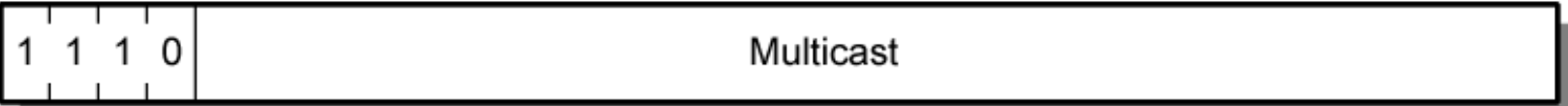
Class A



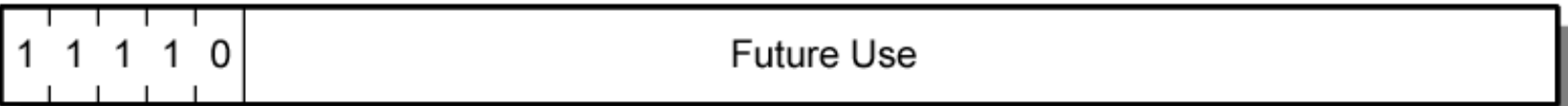
Class B



Class C



Class D



Class E

TCP/IP

- The Internet (and most other computer networks) are connected through **TCP/IP** networks
- TCP/IP is actually a combination of two protocols:
 - **IP, Internet Protocol**, is used to move **packets** (chunks) of data from one place to another
 - Places are specified by **IP addresses**: four single-byte (0..255) numbers separated by periods
 - Example: **192.168.1.1**
 - **TCP, Transmission Control Protocol**, ensures that all necessary packets are present, and puts them together in the correct order
- TCP/IP forms a “wrapper” around data of *any* kind
- The data uses its own protocol, for example, FTP

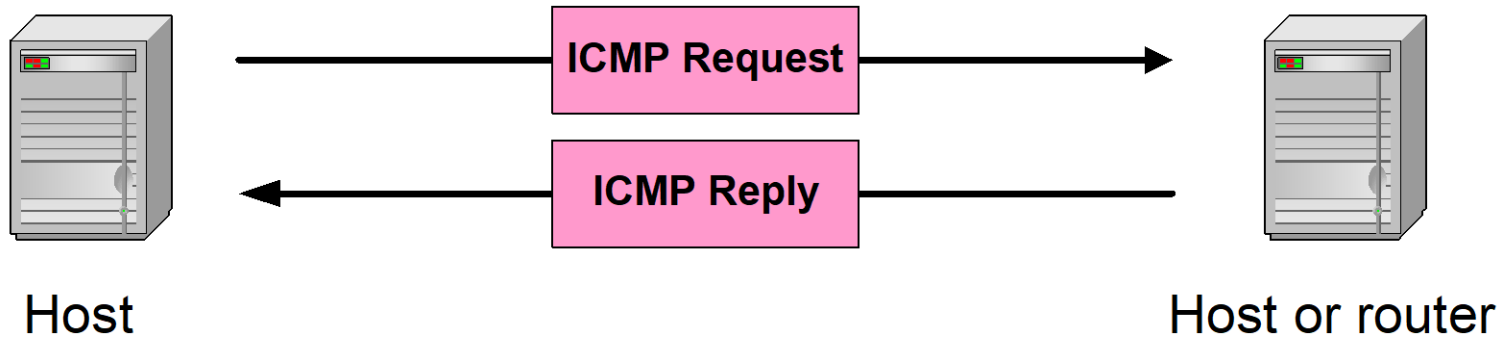
ICMP

- Internet Control Message Protocol - RFC 792
 - Sister protocol of IP
 - All IP implementations should also implement ICMP
- Transfer of (control) messages from routers-to-hosts and hosts-to-hosts
- Feedback about problems
 - e.g. datagram discarded, router's buffer full
- Some simple applications can be implemented using ICMP
 - e.g. Ping
- Encapsulated in IP datagram
 - Thus not reliable

ICMP

- The **ICMP** is a helper protocol that supports IP with facility for
 - Simple queries
 - Error reporting

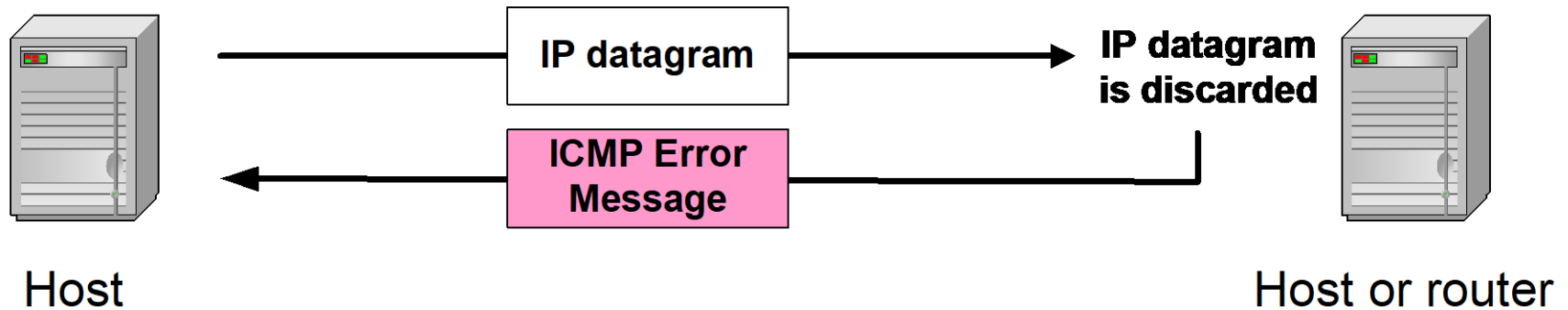
ICMP Query message



ICMP query:

- **Request** sent by host to a router or host
- **Reply** sent back to querying host

ICMP Error message

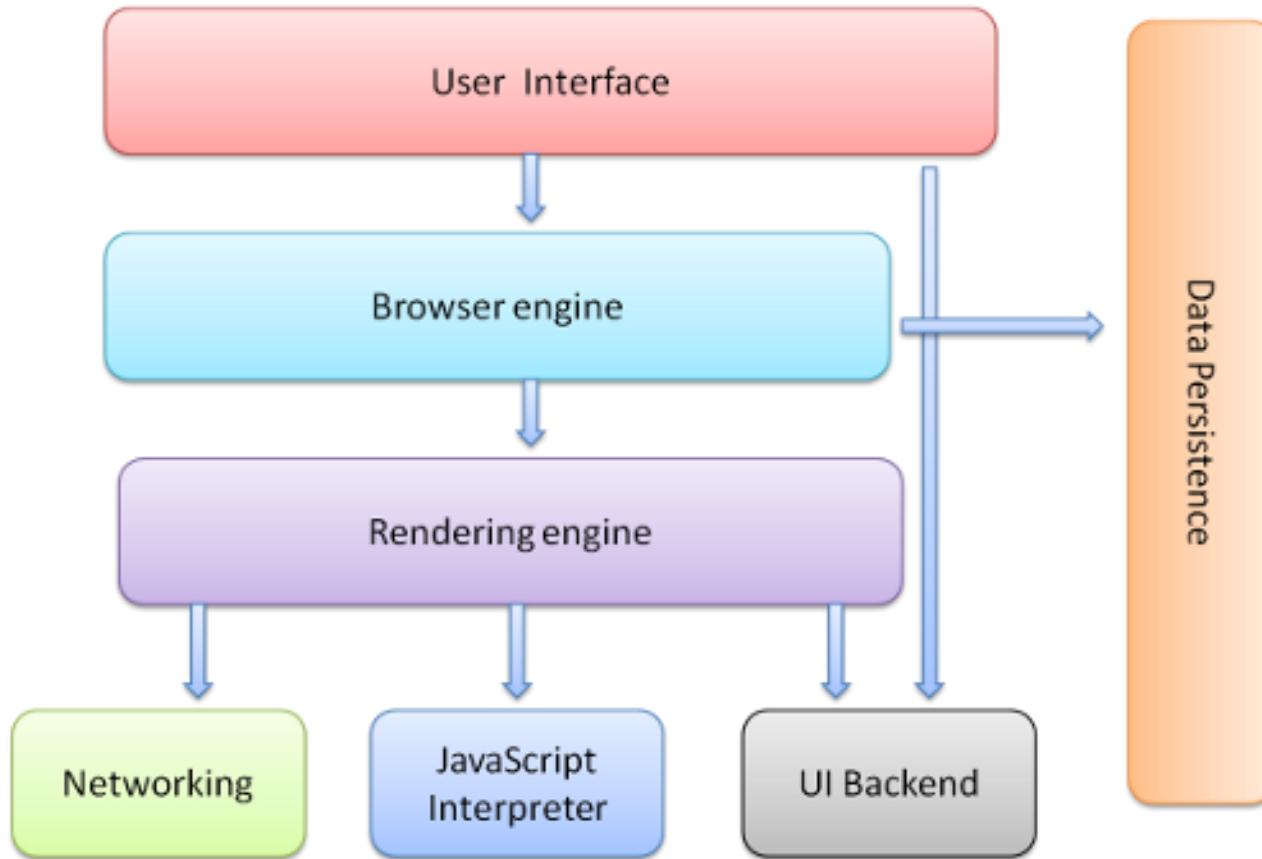


- **ICMP error messages report error conditions**
- **Typically sent when a datagram is discarded**
- **Error message is often passed from ICMP to the application program**

Web Browsers

- A software that allows users to access and view web pages.
- It is mainly used for retrieving and presenting / displaying information on web sites.
- **1990** – The WorldWideWeb this was the only browser and the only way to access the web.
- **1992** – Lynx was a texted-based browser that couldn't display any graphic content.
- **1993** – first browser to allow images embedded in text -“the world's first most popular browser”.
- **1994** – A noticeable improvement to Mosaic came Netscape Navigator.
- **1995** – Internet Explorer made its debut as Microsoft's first web browser by Microsoft.
- **1996** – Opera started as a research project in 1994 that finally went public two years later.
- **2003** – Apple's Safari browser was released specifically for Macintosh computers
- **2004** – Mozilla launched Firefox as Netscape Navigator faded out.
- **2008** – Google Chrome appeared to soon take over the browser market.
- **2011** – Opera Mini was released to focus on the fast-growing mobile browser market.
- **2015** – Microsoft Edge was born to combat Google.

Browser's main components



https://www.html5rocks.com/en/tutorials/internals/howbrowserswork/#The_rendering_engine

Browser's main components

- **The user interface:** this includes the address bar, back/forward button, bookmarking menu, etc.
- **The browser engine:** marshals actions between the UI and the rendering engine.
- **The rendering engine :** responsible for displaying requested content.
- **Networking:** for network calls such as HTTP requests
- **UI backend:** used for drawing basic widgets like combo boxes and windows.
- **JavaScript interpreter.** Used to parse and execute JavaScript code.
- **Data storage.** The browser may need to save all sorts of data locally, such as cookies.

Web Servers

- To dispense the web page when it is requested by web clients (browser)
- Computers on which this web server program run are usually called as servers , for hosting web.
- In web hosting, the web servers enable the hosting providers to handle multiple domains (or multiple websites) on a single server.
- Main uses of web server:
 - To own an website, maintains the data need to be displayed on web page
 - To use server side technologies such as PHP, json

How Web servers work?

1. Obtaining the IP Address from domain name:

- By searching in its cache.
- By requesting one or more DNS

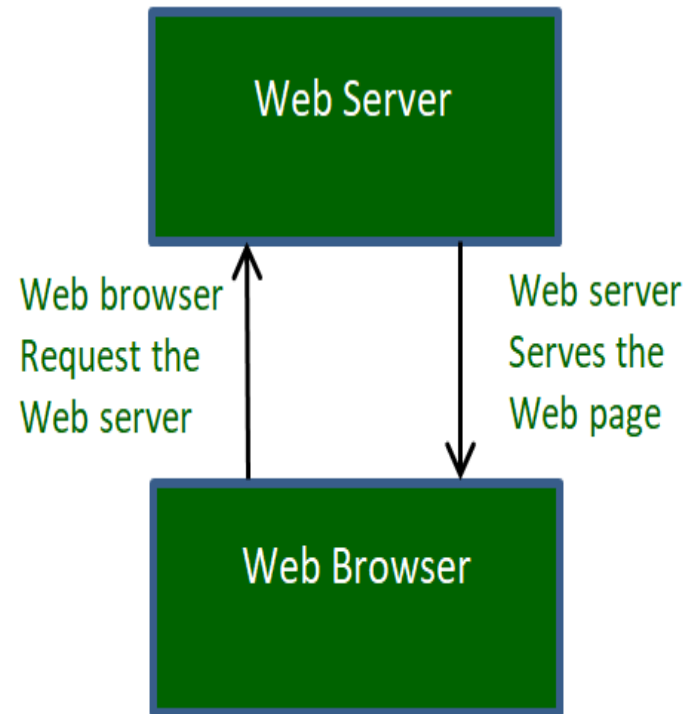
2. Browser requests the full URL

After knowing the IP Address, the browser now demands URL from the web server.

3. Web server responds to request:

by sending the desired pages, or error messages

4. Browser displays the web page:



Popular Web Servers

- **Apache HTTP Server:** most popular web server available and is widely used.
which makes it free and open source, available for Windows, Mac OS X, Unix, Linux, Solaris,
- **Microsoft Internet Information Service (IIS)**
 - second most popularly used web server,
 - IIS server has all the features just like Apache, but it is not open source.
 - It is developed, and maintained by Microsoft, it works all Windows operating system platforms.

Web Browser vs Web Server

WEB BROWSER	WEB SERVER
Web Browser is an Application program that displays a World wide web document	Web server is a program or the computer that provide services to web browsers
The Web browser requests the server for the web documents and services.	The Web server accepts, approve and respond to the request made by the web browser.
The web browser act as an interface between server and client and displays a web document	The web server is a software or a system which maintain the web applications, generate response and accept clients data.
The web browser sends an HTTP request and gets an HTTP response.	The web server gets HTTP requests and send HTTP responses.

Website vulnerability

- A **website vulnerability** is a weakness or misconfiguration in a **website** or **web** application code
 - allows an attacker to gain some level of control of the site, and hosting server.
 - Most **vulnerabilities** are exploited through **vulnerability** scanners and botnets
- Cybercriminals create specialized tools that scour the internet for certain platforms, like WordPress or Joomla,
 - to steal data, distribute malicious content, or inject defacement and spam content

Website Security Vulnerabilities

1. SQL INJECTIONS

- to use application code to access or corrupt database content.
- to create, read, update, alter, or delete data stored in the back-end database.

2. CROSS SITE SCRIPTING (XSS)

- usually a client-side script such as JavaScript, into a web application's output.
- to execute scripts in the victim's browser which can hijack user sessions, deface websites or redirect the user

3. BROKEN AUTHENTICATION & SESSION MANAGEMENT

- maintaining the identity of a user.
- hijack an active session and assume the identity of a user.

Website Security Vulnerabilities

4. INSECURE DIRECT OBJECT REFERENCES

Insecure direct object reference - hackers can manipulate it to gain access to a user's personal data.

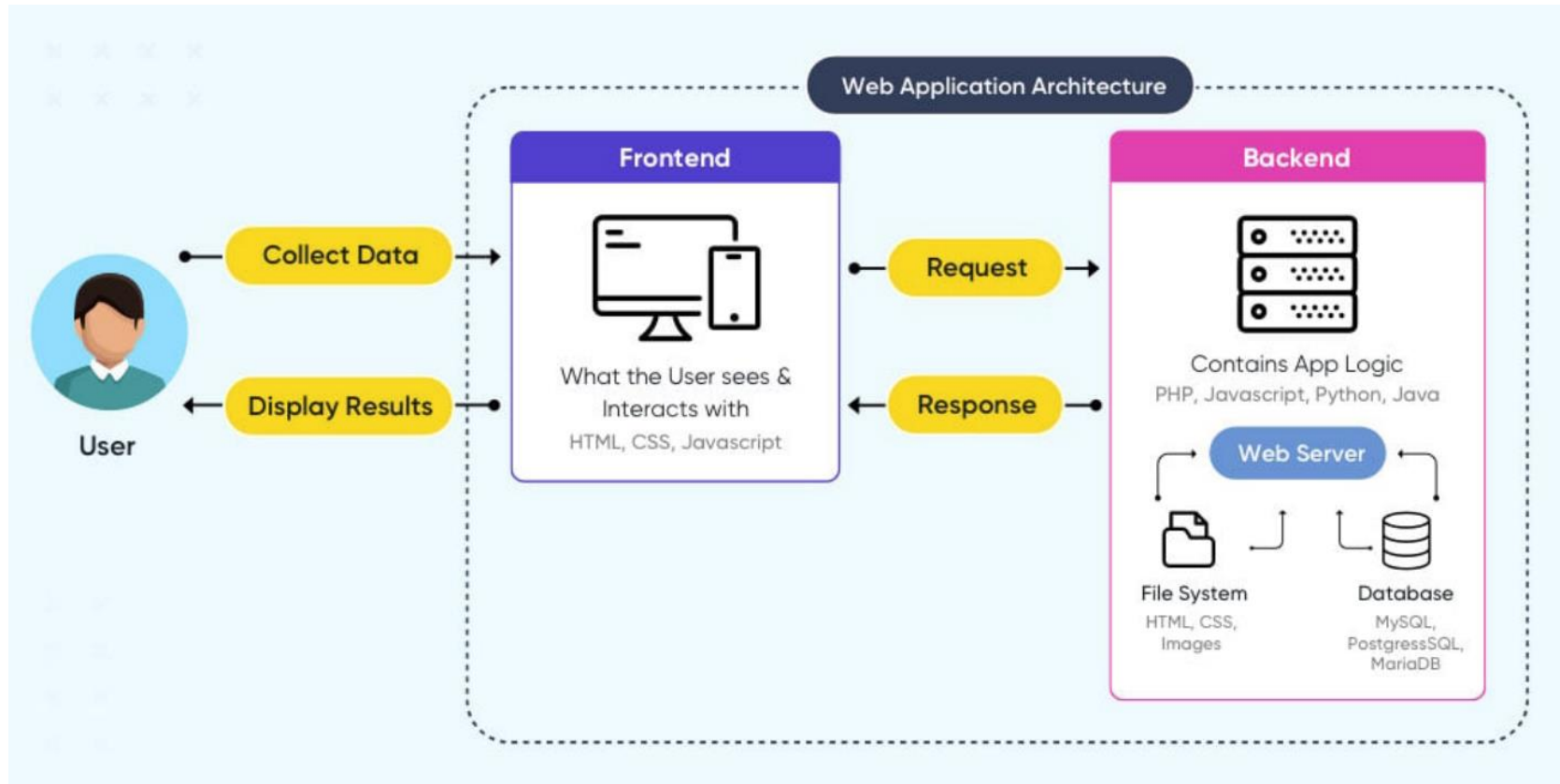
5. SECURITY MISCONFIGURATION

- lack of maintenance or a lack of attention to the web application configuration.
- hackers access to private data or features and can result in a complete system compromise.

6. CROSS-SITE REQUEST FORGERY (CSRF)

- A third-party website will send a request to a web application
- attacker can then access functionality via the victim's already authenticated browser.

Web System Architecture



Static Web: HTML, CSS

- **HTML** stands for **H**yper**T**ext **M**arkup **L**anguage
 - It is a text file containing small markup tags that tell the Web browser how to display the page
- **CSS** stands for **C**ascading **S**tyle **S**heets
 - It defines how to display HTML elements

Client-Side Programmability

- Scripting language: a lightweight programming language
- Browser scripting: **JavaScript**
 - Designed to add interactivity to HTML pages
 - Usually embedded into HTML pages
 - What can a JavaScript Do?
 - Put dynamic text into an HTML page
 - React to events
 - Read and write HTML elements
 - Validate data before it is submitted to a server
 - Create cookies

Server-Side Programmability

- The requests cause the response to be generated
- Server scripting:
 - **PHP:** Open source, strong database support (*.php)
 - **Node.JS** – Server side JavaScript
 - ASP: Microsoft product, uses .Net framework (*.asp)
 - Java via JavaServer Pages (*.jsp)
 - CGI/Perl: Common Gate Way Interface (*.pl, *.cgi)
 - ...

Transmitting DATA in Web Application

- MySQL
- XML
- JSON
- ...

Thank You