Session 2

Background

Lecture Objectives

- Understand how an Internet resource is accessed
- Understand the high level structure of the "Internet cloud"
- Understand the high level structure of the TCP/IP protocols
- Understand how a computer finds the IP address of a host using DNS
- Know the structure of MIME type standards

Reference

- Mime types
  http://en.wikipedia.org/wiki/Internet_media_type

What is a URL?

- A short string that identifies a resource on the Web
- Stands for Uniform Resource Locator
  - Uniform - varied and new types of resources
  - Resource - Anything that has identity (e.g., image)
- Reduces the tedium of connecting to a host, selecting a path, selecting a resource, etc. into a single string that:
  - Can be saved as a bookmark in your browser
  - Can be saved as an object in your Java code

URL Example

http://www.cs.sunysb.edu

Protocol identifier *Resource name*

- Protocol Identifier - Indicates the name of the protocol to be used to fetch the resource
- Resource name is the complete address of the resource
- Resource name may be appended with a fragment / reference / named anchor (denoted by #) or include a query string (denoted by ?)

5

URL Resource Name

- For http, the name includes: host name, path name to the file, port number (optional), and location within the resource (optional)
- The resource is not necessarily a file, it can be generated dynamically

important concept

- A trailing slash (www.sun.com/) is shorthand for the file named /index.html

6

Relative URLs

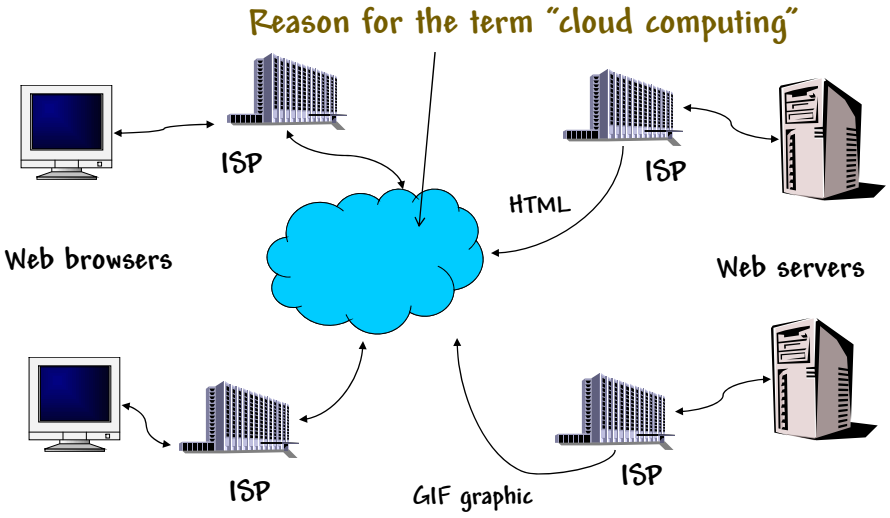
- Contains only enough information to reach the resource relative to (in the context of) another URL
- Used within HTML files

```
<td> <a href="LectureCode.html" >Intro</a></td>
```

- specified relative to the location of the file in which they are contained

7

Web Architecture



8

9

Internet Origins

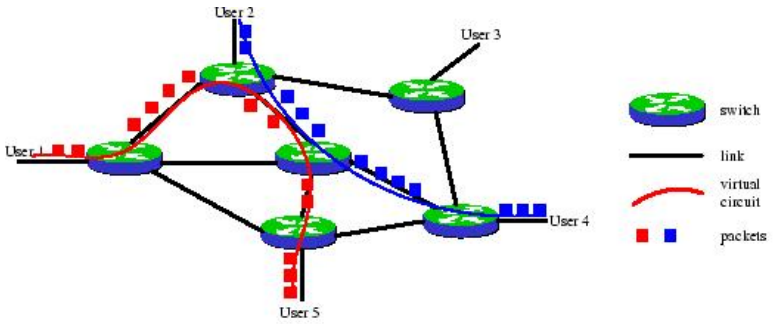
- Late 1960s
  - ARPA (Advanced Research Projects Agency)
  - Funded research project
  - ARPANet - redundant network, connecting research labs and US Government installations
- 1970s - 1980s
  - Enormous growth in US sponsored Internet
- 1990s
  - US Government turns Internet over to independent global agencies

DARPA has changed its name (back and forth) to ARPA

10

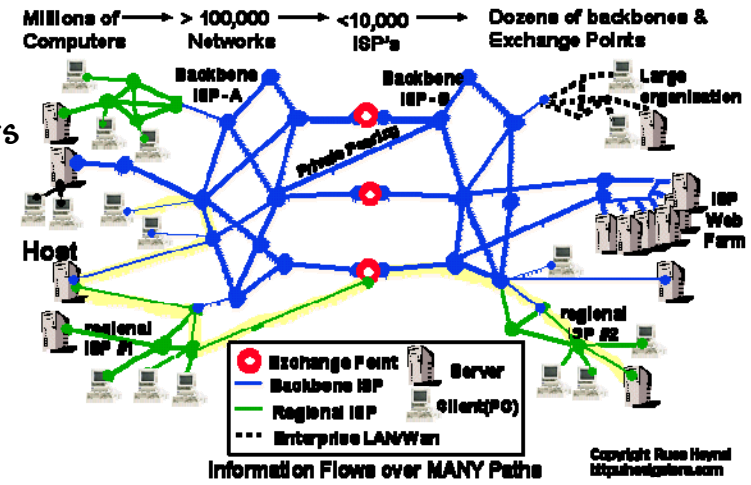
ARPANet Principles

- Free and open information sharing
- Datagram message packets - data carried in chunks (not streams)
- Uniform protocol for communications between dissimilar computers

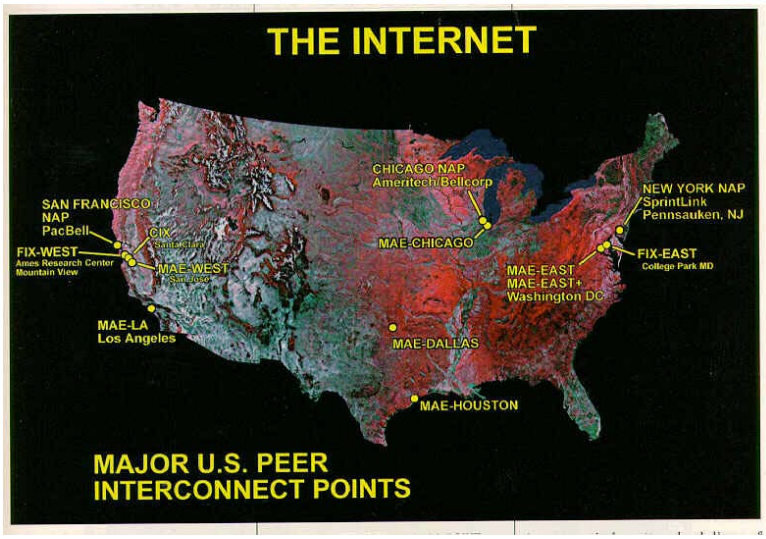


Internet Physical Layer

- Multiple backbones
- Multiple ISPs
- Multiple exchange points



Internet Interconnect Points



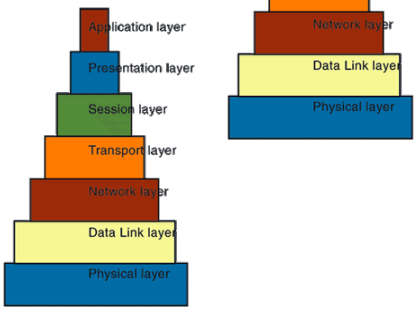
13

Terms

- Protocols - control the sending and receiving of information over the Internet
- Physical media - cable, copper wire, fiber, radio spectrum
- Routers - intermediate switching devices
- Route (or path) - sequence of physical media and routers to complete end-to-end communication
- Packet switching - decomposing a message into packets and routing the packets to a destination

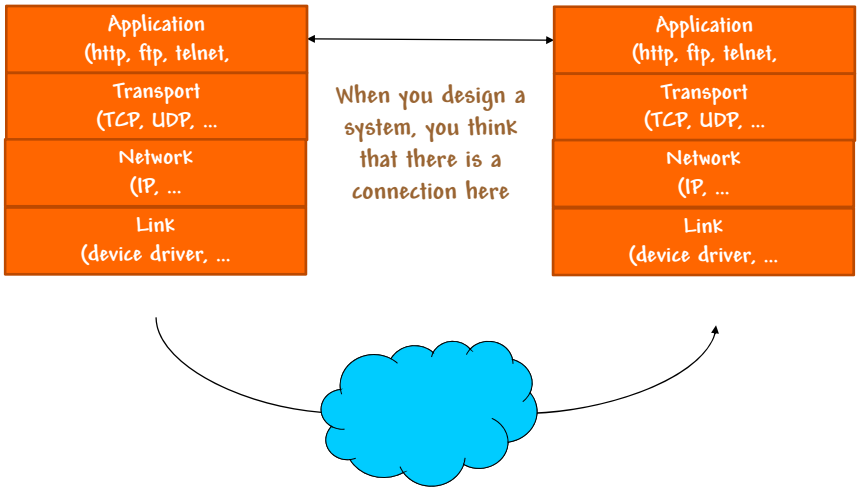
Protocol Stacks

- Makes interfaces portable and maintainable
- Each layer of the stack only interfaces with the adjacent layers
- Protocol conversion occurs between the same layer on different computers
- Examples
  - ISO/OSI



14

Internet Protocol Stacks

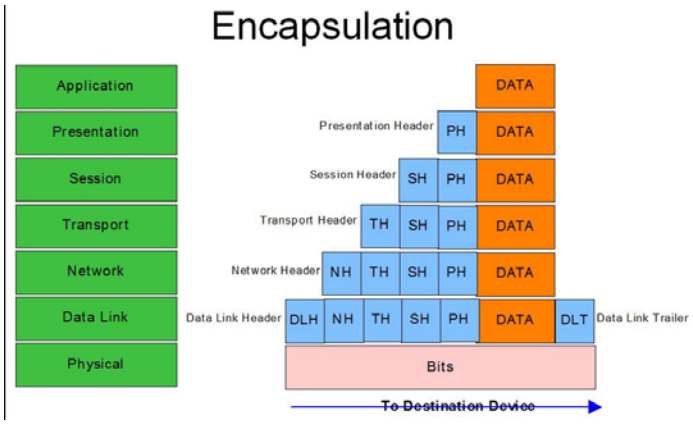


15

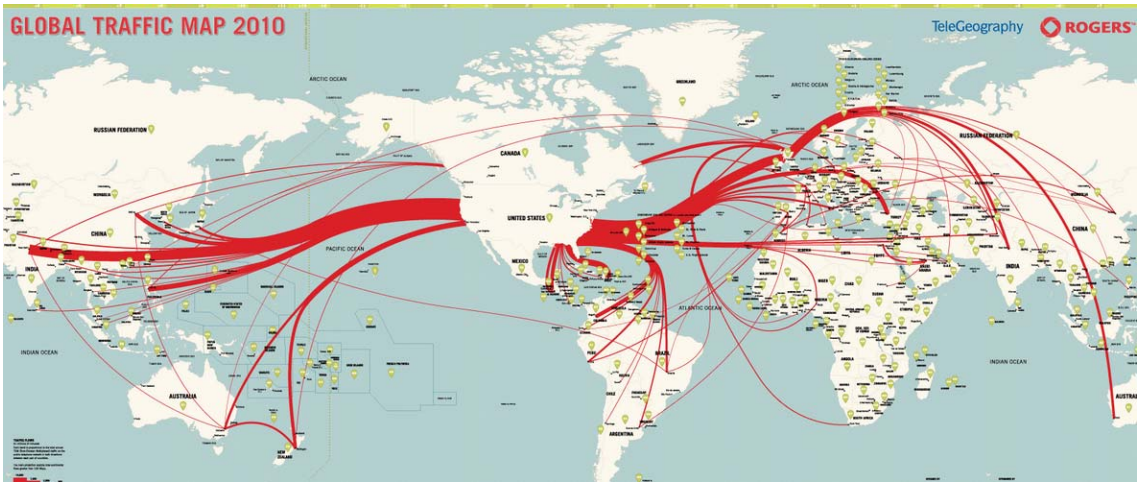
16

Data Encapsulation

- Layer to layer interface usually involves encapsulation of data
- Encapsulation - wrapping the data with information that describes the data and helps route the data



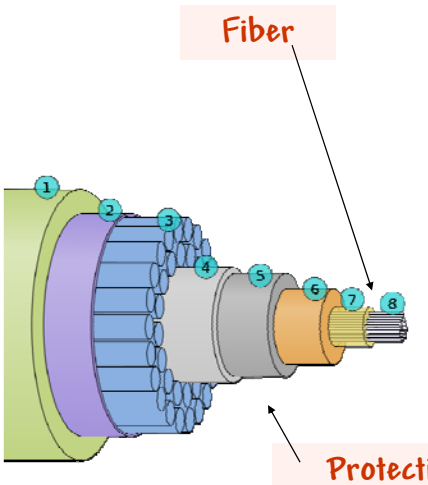
Global Traffic

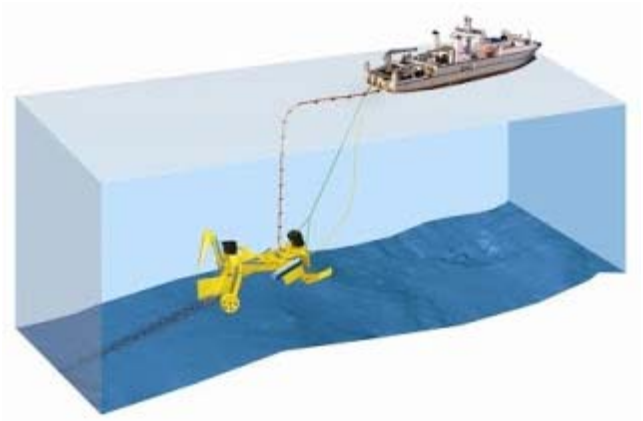


17

18

Transoceanic Cables





Asia Backbone

19

Transport Protocols

- Internet connection-oriented service is implemented through buffers at the sending and receiving end

- TCP
- Transmission Control Protocol
- Connection based
- Reliable flow of data between two computers

- UDP
- User Datagram Protocol
- Connectionless service
- Order of delivery is not guaranteed

20

TCP/IP

- IP (Internet Protocol) - network layer protocol that specifies the format of information that is sent and received among routers and end nodes

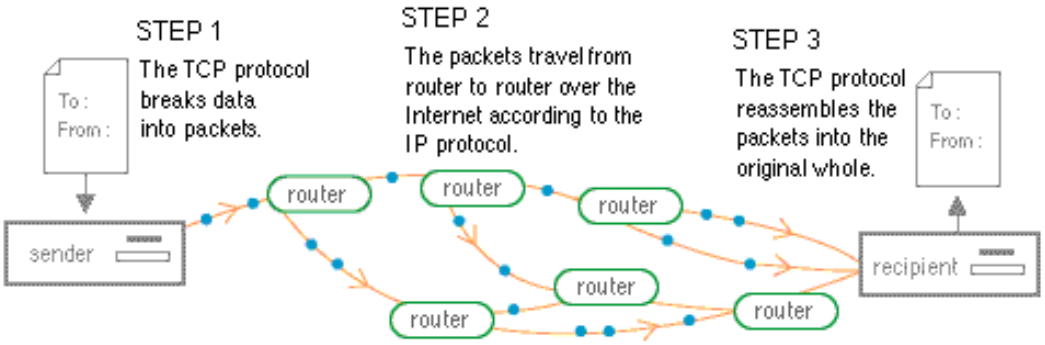
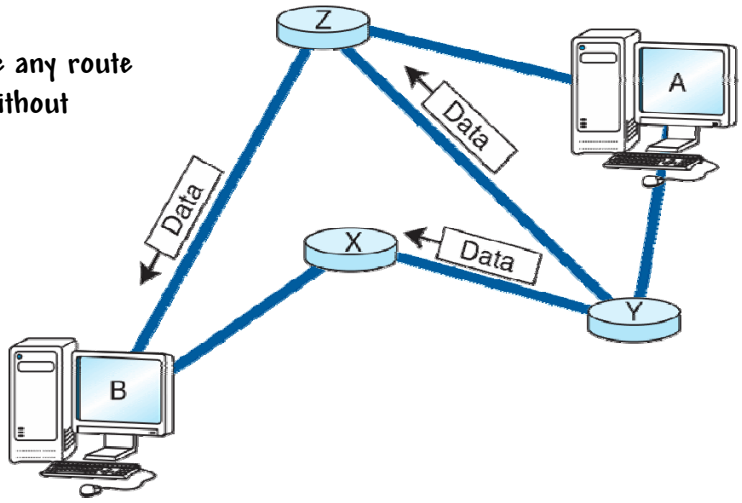


Figure 2. How data travels over the Net.

21

TCP/IP Network Architecture

- Datagrams can take any route available to them without human intervention



IPv4

- Current version of IP
- Not adequate to serve millions of network components scattered across the globe.
- Limitations
  - 32-bit addresses
  - a packet length limited to 65,635 bytes
  - all security measures are optional
- Network addresses have been assigned with little planning resulting in slow and cumbersome routing hardware and software

IP Addressing

- Hosts and other devices have network interfaces identified by an IP address
- IP (IPv4) addresses are 32-bit numbers represented as four groups of 8 bits (byte)
- Written in dotted-decimal notation
- IP address consists of
  - Network ID - portion of the IP address that defines the network to which the device is connected
  - Host ID - portion of the IP address that defines the host

stonybrook.edu

129.49.2.176

Record Type: IP Address

2^{32} possible IP addresses (not enough)

24

25

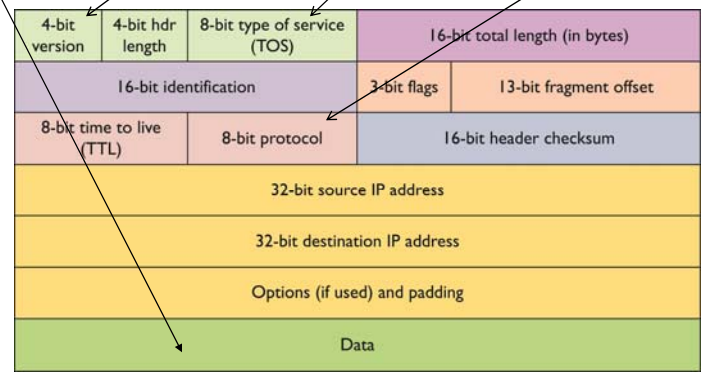
IP Packet

- Header
- Data section (payload)

IPv4 or IPv6

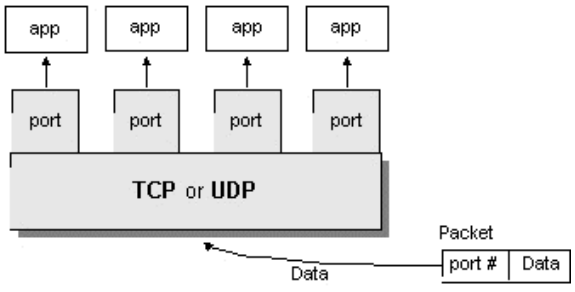
Can be used to distinguish differing levels of service

Usually TCP or UDP



Application Interface

- Interface is designated by a port number
- Socket - combination of the port number, the host ID, and the protocol designation (equivalent to a file name to the application running above TCP)

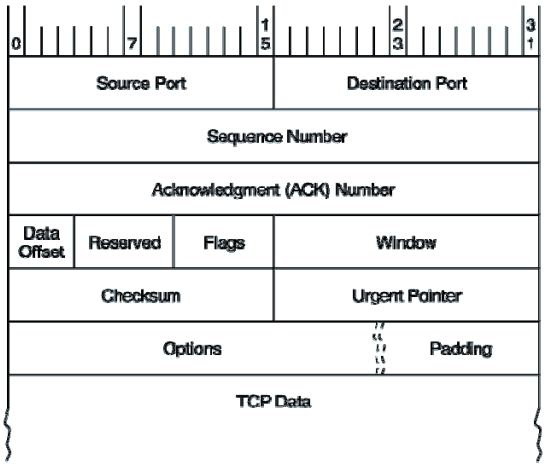


Port Numbers

- Port numbers 0 through 1023 are reserved for particular TCP applications
- Examples
  - telnet - 23
  - smtp - 25
  - http - 80

27

TCP Segment Format



28

IPv6

- In 1994, the Internet Engineering Task Force began work on what is now IP Version 6
- Motivation
  - Extend IP's address space beyond its current 32-bit limit to 128 bits for both the source and destination host addresses
  - Slow roll-out

29

How Does a Browser Work?

- It reads a named resource (usually an HTML document) on an Internet-based server
- Begins to display the page
- Finds all the URLs in the HTML
- Requests the resources associated with the other URLs (e.g., images)
- Includes the additional resources in the display of the page

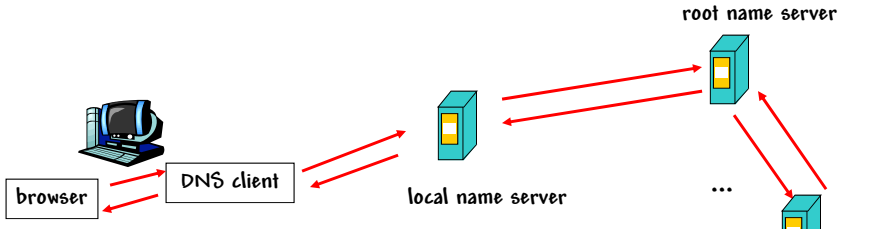
30

How Does the Internet Find a Host?

- URL contains a host name, (e.g., www.cnn.com)
- Internet routers identify hosts by their IP address
  - 4 bytes, presented in dotted-decimal notation
    - decimal numbers, separated by periods (e.g., 121.7.106.83)
    - Each number is between 0 and 255 (8-bit binary number)
  - 2^{32} possible IP addresses in IP V4 (theoretical maximum)
- DNS (Domain Name System) is a directory service that translates host names (sometimes referred to as domain names) into IP addresses

31

How DNS Works



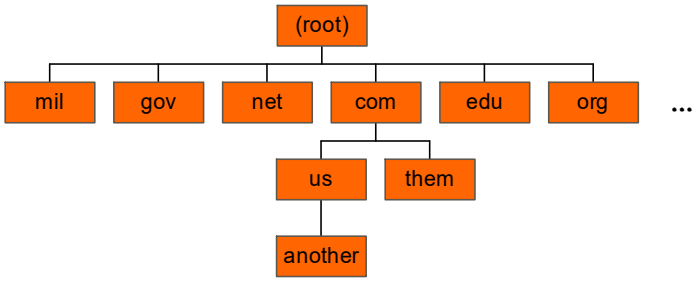
- Browser passes host name to DNS client.
- DNS client looks up IP address from the distributed database located on the DNS name servers
  - Local
  - Root
  - Authoritative

There are 13 root server identities (A, B, C, ..., M), but many of them are aliased

32

Domain Name Service

- Used to map host names (contained within a URL) into network addresses (32 bit IP address)
- Name space (partial)



33

Other DNS Services

- Host aliasing - a host can have one or more alias names (e.g., ibm.com and www.ibm.com), one of which is the canonical hostname
- Aliasing by service - a company can use the same host name for its Web server, mail server, ftp server, etc. even though these are different computers with different IP addresses
- Load distribution - a set of IP addresses can be associated with a canonical hostname. DNS will return the set of addresses, varying the order of addresses each time it is accessed

How many hosts do you think are associated with google.com?

34

DNS Concepts to Remember

- DNS provides name to address mapping
- DNS is implemented through local address caching
- DNS is a distributed database service
  - Very reliable
  - Not always fast
  - Not always up to date

35

Cyber Security Issues

- Internet is inherently insecure
- Industries and country economies are increasingly dependent on the Internet
- Many successful cyber attacks
- Examples of cyber warfare (e.g., Estonia, Georgia, 2016 US elections)
- "Open" countries are particularly vulnerable to attack
- Limited experience with response options

36

Vulnerabilities of the Internet[1]

- Routing among ISPs - Border Gateway Protocol routes packets across the Internet, but there is no checking of authenticity of messages
- No governance of the Internet beyond domain names
- Operational messages are unencrypted
- Malware can be easily propagated
- Decentralized design

An advanced packet sniffer on an Ethernet network can look at all the traffic

1. Clarke, R, Cyber War

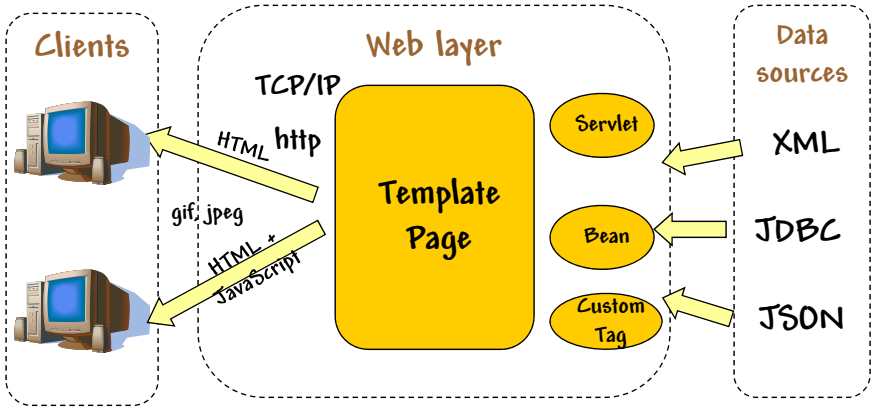
37

What is a MIME Type?

- Multipurpose Internet Mail Extensions
- Designed for the interchange of data among various e-mail systems
- Allows for universal interchange of data
- Defines naming of file types
- Organized into 8 base type categories

38

Extended Web Architecture



The host-host data transfers usually involve one MIME file

39

"Base" Media Types

- application - application and miscellaneous
- audio - audio data
- example
- image - image data
- message - news, e-mail, etc.
- model - models (e.g., geometric)
- multipart - multipart
- text - HTML, CSS, etc.
- video - video data

e.g., image/png

40

MIME Type Concepts to Remember

- Internet files have standard formats so that data can be exchanged easily between very different computers (hardware, OS, etc.)
- Sender specifies the type and receiver interprets the data accordingly, taking into account all the differences in internal data representation

41

Did You Achieve the Lecture Objectives?

- Understand how an Internet resource is accessed
- Understand the high level structure of the "Internet cloud"
- Understand the high level structure of the TCP/IP protocols
- Understand how a computer finds the IP address of a host using DNS
- Know the structure of MIME type standards

42