The Internet and World Wide Web

CSE 190 M (Web Programming), Spring 2008
University of Washington

Reading: Ch. 0: Introduction

The Internet

- Wikipedia: http://en.wikipedia.org/wiki/Internet
- a connection of computer networks using the Internet Protocol (IP)
- What's the difference between the Internet and the World Wide Web (WWW)?

- the Web is the collection of web sites and pages around the world; the Internet is larger and also includes other services such as email, chat, online games, etc.

Brief history

- began as a US Department of Defense network called <u>ARPANET</u> (1960s-70s)
- initial services: electronic mail, file transfer
- opened to commercial interests in late 80s
- WWW created in 1989-91 by <u>Tim Berners-Lee</u>
- popular web browsers released: Netscape 1994, IE 1995
- Amazon.com opens in 1995; Google January 1996

- <u>Hamster Dance</u> web page created in 1999

Key aspects of the internet

- subnetworks can stand on their own
- computers can dynamically join and leave the network
- built on open standards; anyone can create a new internet device
- lack of centralized control (mostly)
- everyone can use it with simple, commonly available software

People and organizations

- Internet Engineering Task Force (<u>IETF</u>): internet protocol standards

- Internet Corporation for Assigned Names and Numbers (<u>ICANN</u>): decides top-level <u>domain names</u>

- World Wide Web Consortium (<u>W3C</u>): web standards

Web Servers and Browsers

- **web server**: a computer running web server software that listens for web page requests on TCP port 80
- popular web server software:
    - Apache: www.apache.org
    - Microsoft Internet Information Server (IIS) (part of Windows)
- popular web browser software (fetches/displays documents from web servers):

    - Mozilla Firefox: getfirefox.com
    - Microsoft Internet Explorer (IE): part of Windows
    - Apple Safari: part of Mac OS X
    - Opera: opera.com

Layered architecture

The internet uses a layered hardware/software architecture (also called the "OSI model"):

- *physical layer* : devices such as ethernet, coaxial cables, fiber-optic lines, modems
- *data link layer* : basic hardware protocols (ethernet, wifi, DSL PPP)
- *network / internet layer* : basic software protocol (IP)
- *transport layer* : adds reliability to network layer (TCP, UDP)
- *application layer* : implements specific communication for each kind of program (HTTP, POP3/IMAP, SSH, FTP)
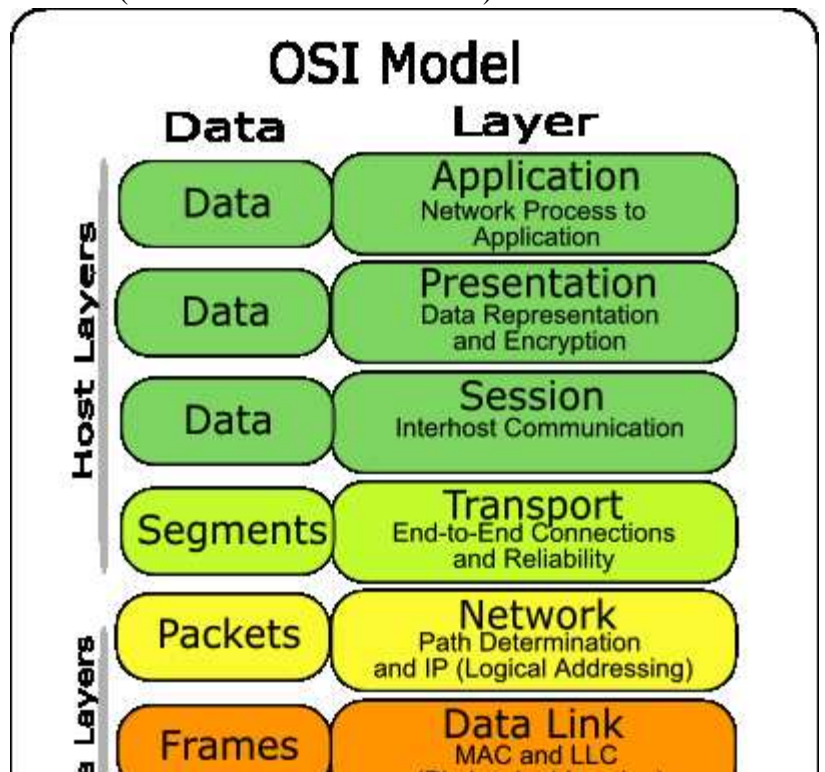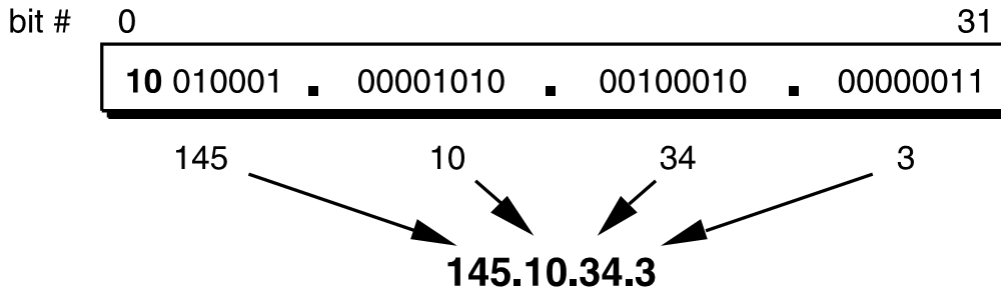
Internet Protocol (IP)

- a simple protocol for attempting to send data between two computers
- each device has a 32-bit IP address written as four 8-bit numbers (0-255)
  e.g. 145.10.34.3

bit # 0   31

10 010001 . 00001010 . 00100010 . 00000011

145   10   34   3

145.10.34.3

- find out your internet IP address: <u>whatismyip.com</u>
- find out your local IP address:
  in a terminal, type: `ipconfig` (Windows) or `ifconfig` (Mac/Linux)

Transmission Control Protocol (TCP)

- adds multiplexing, guaranteed message delivery on top of IP
- **multiplexing**: multiple programs using the same IP address
  - **port**: a number given to each program or service
  - port 80: web browser
  - port 25: email
  - port 22: ssh
  - port 5190: AOL Instant Messenger
  - <u>more common ports</u>
- some programs (games, streaming media programs) use simpler <u>UDP</u> protocol instead of TCP

Domain Name System (DNS)

- a set of servers that map written names to IP addresses
  - Example: `www.cs.washington.edu` → `128.208.3.88`
- many systems maintain a local cache called a <u>hosts file</u>
  - Windows: `C:\Windows\system32\drivers\etc\hosts`
  - Mac: `/private/etc/hosts`
  - Linux: `/etc/hosts`

Uniform Resource Locator (URL)

- an identifier for the location of a document on a web site
- a basic URL:

```
http://www.aw-bc.com/info/regesstepp/index.html
~~~~    ~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~
protocol    host                   path
```

- upon entering this URL into the browser, it would:
    - ask the DNS server for the IP address of `www.aw-bc.com`
    - connect to that IP address at port 80
    - ask the server to `GET /info/regesstepp/index.html`
    - display the resulting page on the screen

More advanced URLs

- **anchor**: jumps to a given section of a web page

  `http://www.textpad.com/download/index.html`**#downloads**

    - the above URL fetches `index.html` and then jumps downward to a part of the page labeled `downloads`
- **port**: for web servers on ports other than the default 80

  `http://www.cs.washington.edu`**:8080**`/secret/money.txt`

- **query string**: a set of parameters passed to a web program

  `http://www.google.com/search`**?q=miserable+failure&start=10**

    - the above URL asks the server at `www.google.com` to run the program named `search` and pass it two parameters:
        - `q` (set to `"miserable+failure"`)
        - `start` (set to 10)

Hypertext Transport Protocol (HTTP)

- the set of commands understood by a web server and sent from a browser
- some HTTP commands (your browser sends these internally):
    - GET *filename* : download
    - POST *filename* : send a web form response
    - PUT *filename* : upload
- simulating a browser with a terminal window:

```
$ telnet www.cs.washington.edu 80
Trying 128.208.3.88...
Connected to 128.208.3.88 (128.208.3.88).
Escape character is '^]'.
GET /index.html
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 ...">
<html>
...
```

HTTP error codes

- the web server returns a special "error code" number to the browser, possibly followed by an HTML document
- common error codes:

Number	Meaning
200	OK
301-303	page has moved (permanently or temporarily)
403	you are forbidden to access this page
404	page not found
500	internal server error

- complete list of error codes

Web programming technologies

- Hypertext Markup Language (HTML): used for writing web pages
    - XHTML: A newer, more rigidly standardized specification of HTML
- Cascading Style Sheets (CSS): supplies stylistic info to web pages
- JavaScript: allows interactive and programmable web pages
- eXtensible Markup Language (XML): metalanguage for organizing data
- Asynchronous JavaScript and XML (Ajax): allows fetching of web data in the background for enhanced interactive web apps
- PHP Hypertext Processor (PHP): allows web server to create pages dynamically
- Structured Query Language (SQL): interaction with databases