Web Programming Step by Step

Chapter 1
The Internet and World Wide Web

1.1: The Internet

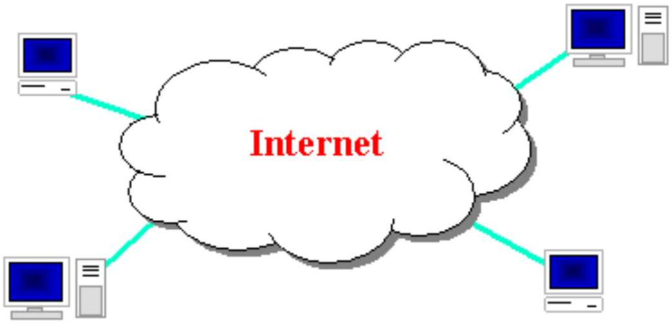
- **1.1: The Internet**
- 1.2: The World Wide Web (WWW)

What is the Internet?

- A "series of tubes" (explanation)

- How many internets are there, anyway? Is The Google one of them?

The Internet

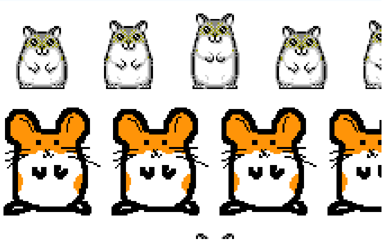


- Wikipedia: http://en.wikipedia.org/wiki/Internet
- a connection of computer networks using the Internet Protocol (IP)
- What's the difference between the Internet and the World Wide Web (WWW)?

- 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 (1.1.1)

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

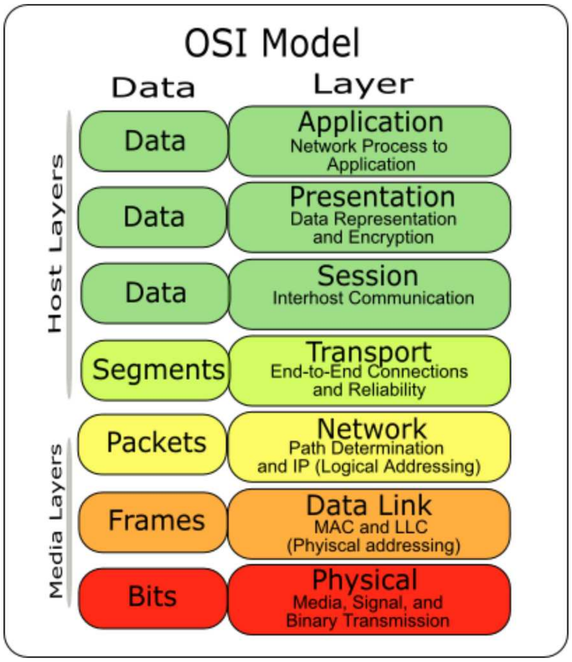
People and organizations (1.1.2)

- Internet Engineering Task Force (IETF): internet protocol standards
- Internet Corporation for Assigned Names and Numbers (ICANN): decides top-level domain names
- World Wide Web Consortium (W3C): web standards

Layered architecture (1.1.3)

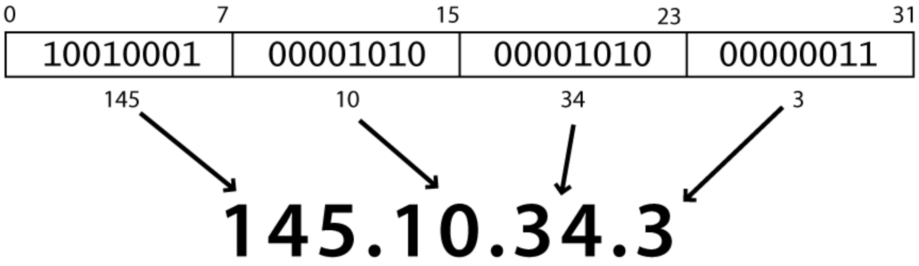
The internet uses a layered hardware/software architecture (also called the "OSI model"):

- *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)


OSI Model
Data
Layer
Host Layers
Media Layers
Data
Application
Network Process to Application
Data
Presentation
Data Representation and Encryption
Data
Session
Interhost Communication
Segments
Transport
End-to-End Connections and Reliability
Packets
Network
Path Determination and IP (Logical Addressing)
Frames
Data Link
MAC and LLC (Phyiscal addressing)
Bits
Physical
Media, Signal, and Binary Transmission

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)



- find out your internet IP address: whatismyip.com
- 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 443 for secure browsing)
  - port 25: email
  - port 22: ssh
  - port 5190: AOL Instant Messenger
  - more common ports
- some programs (games, streaming media programs) use simpler UDP protocol instead of TCP

1.2: The World Wide Web (WWW)

Web servers and browsers (1.2.1)

Mozilla Firefox Start Page - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www
Getting Started
Latest Headlines
Firefox Start

- **web server**: software that listens for web page requests
  - Apache
  - Microsoft Internet Information Server (IIS) (part of Windows)
- **web browser**: fetches/displays documents from web servers
  - Mozilla Firefox
  - Microsoft Internet Explorer (IE)
  - Apple Safari
  - Google Chrome
  - Opera

Domain Name System (DNS) (1.2.2)

- 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 hosts file
  - 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`**

  - fetches `index.html` then jumps down to 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`**

  - parameter `q` is set to `"miserable+failure"`
  - parameter `start` is set to `10`

Hypertext Transport Protocol (HTTP) (1.2.3)

- 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

- when something goes wrong, 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	

Internet media ("MIME") types

- sometimes when including resources in a page (style sheet, icon, multimedia object), we specify their type of data

MIME type	file extension
`text/html`	`.html`
`text/plain`	`.txt`
`image/gif`	`.gif`
`image/jpeg`	`.jpg`
`video/quicktime`	`.mov`
`application/octet-stream`	`.exe`

- Lists of MIME types: by type, by extension

Web languages / technologies (1.2.4)

- Hypertext Markup Language (HTML): used for writing web pages
- Cascading Style Sheets (CSS): stylistic info for web pages
- PHP Hypertext Processor (PHP): dynamically create pages on a web server
- JavaScript: interactive and programmable web pages
- Asynchronous JavaScript and XML (Ajax): accessing data for web applications
- eXtensible Markup Language (XML): metalanguage for organizing data
- Structured Query Language (SQL): interaction with databases