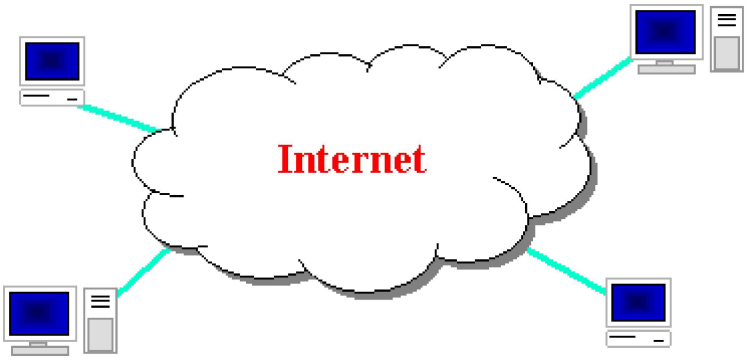
# Introduction to Web Programming

## Lecture 1: Internet/WWW

## 1.1: The Internet

- **1.1: The Internet**
  - (What is it?)
- 1.2: The World Wide Web (WWW)

# The Internet




- Wikipedia: http://en.wikipedia.org/wiki/Internet
- a connection of computer networks using the Internet Protocol (IP)
- layers of communication protocols: IP → TCP/UDP → HTTP/FTP/POP/SMTP/SSH...
- 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 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; Alibaba April 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

# Question

Who "runs" the internet? Who is responsible for overseeing it?

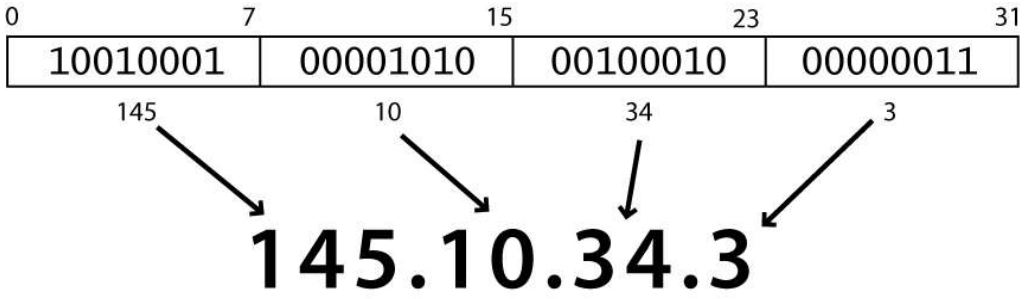
# People and organizations

- 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




# 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)

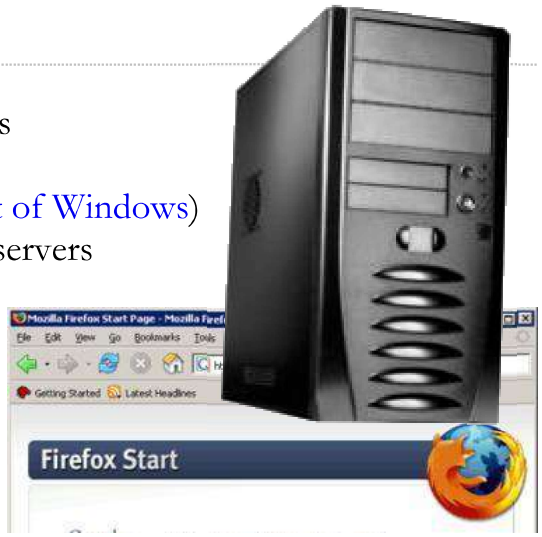
# Transmission Control Protocol (TCP)

- 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)

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

# Web servers and browsers




- **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)

- a set of servers that map written names to IP addresses
  - Example: `www.polytech.unice.fr` → 157.169.25.20
- many systems maintain a local cache called a hosts file
  - Windows: `C:\Windows\system32\drivers\etc\hosts`
  - Mac: `/private/etc/hosts`
  - Linux: `/etc/hosts`

# Question

What is a URL?

- What is an example of a URL you type a lot?
- Why are URLs useful?

# Uniform Resource Locator (URL)

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

  http://users.polytech.unice.fr/~gaetano/tiei/iwp/index.php

  protocol host path

- upon entering this URL into the browser, it would:
  - ask the DNS server for the IP address of users.polytech.unice.fr
  - connect to that IP address at port 80
  - ask the server to GET ~gaetano/tiei/iwp/index.php
  - display the resulting page on the screen

# 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.polytech.unice.fr 80
Trying 157.169.25.20...
Connected to 157.169.25.20 (157.169.25.20).
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 | |

# Web languages / technologies

- 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