The Internet and the Web

• First Lesson: The Internet is **NOT** the Web!
• The Internet was developed as a Cold-War method of computer communication in case of failure (i.e., someone nukes a site).
• The Web was developed *much* later as an application to use the Internet for exchanging *hypertext* documents.

Timeline

**Portents and Beginnings**

• 1957
  - USSR Launches Sputnik (first artificial satellite)
  - ARPA started as a response
• 1962
  - Research on a new network started
• 1969
  - ARPAnet started with 4 machines: (UCLA, UCSB, University of Utah, Stanford Research Institute)
• 1970
  - (13 machines on the ARPAnet)
  - UNIX Operating System development starts

Timeline (Continued)

**Early Network Tools**

• 1972
  - (31 machines)
  - Email developed
  - Telnet (remote log-in) developed
• 1973
  - Email 75% of network traffic
  - FTP (File Transfer Protocol) developed
• 1978
  - First Spam sent

Timeline (Continued)

**Rise of the Internet**

• 1981
  - Rival networks BitNet, Cnet start
  - IPv4 described (Internet Protocol)
  - IBM PC released
• 1982
  - (235 machines)
  - TCP/IP (Transmission Control Protocol/Internet Protocol) formalized
  - Term “Internet” coined as a Network of Networks

Timeline (Continued)

**Growing Pains, but Signs of Maturity**

• 1984
  - (1000 machines)
  - Apple Macintosh released
  - Term “Cyberspace” coined (William Gibson)
  - Domain Name System (DNS) introduced
• 1988
  - (10,000+ machines)
  - Internet Worm released (Robert Morris)
Timeline (Continued)

Rise of the Web

• 1990  (300,000+ machines)
  – ARPAnet decommissioned
  – (Sir) Tim Berners-Lee develops first code for Web
  – First HTTP (HyperText Transport Protocol) action
• 1991
  – First World Wide Web pages available (CERN)
• 1992  (1 Million+ machines)
  – HTML (HyperText Markup Language)
  – First Graphical Browser (Mosaic)

©2014 Dr. William T. Verts

Timeline (Continued)

Internet Commerce

• 1994  (3.9 Million+ machines)
  – Amazon, Yahoo!, IMDb
• 1995
  – Ebay, Craigslist
  – Windows 95

©2014 Dr. William T. Verts

Timeline (Continued)

Reality Sets In

• 1998  (26 Million URLs/addresses)
  – IPv6 described
  – Google, PayPal, Yahoo! Groups
• 2000  (1 Billion unique URLs)
  – Dot-com bubble bursts
• 2001
  – Wikipedia
  – iPod

©2014 Dr. William T. Verts

Timeline (Continued)

Practical Social Networking

• 2003
  – LinkedIn, Myspace, Skype, iTunes store, 4Chan
• 2004
  – Facebook, Flickr, WoW
  – Firefox released
• 2005
  – YouTube, Google Earth, Reddit

©2014 Dr. William T. Verts

Timeline (Continued)

Ubiquity

• 2006
  – Twitter
• 2007
  – Google Street View, Wikileaks, Kindle
  – iPhone
  – Windows Vista, Mac Leopard
• 2008  (1 Trillion unique URLs known)
  – Dropbox

©2014 Dr. William T. Verts

Timeline (Continued)

Saturation

• 2009  (90 Trillion emails)
  – Windows 7
  – Bing, Google Docs, Kickstarter
• 2010  (1.97 Billion Internet users)
  – iPad released
  – 4.6 Billion cell phone subscriptions
  – International Space Station on Twitter
• 2011
  – IPv4 address exhaustion, cut-over to IPv6 starts
• 2012  (900,000,000+ machines)

©2014 Dr. William T. Verts
Web Addresses

- URL: Uniform Resource Locator
    - Protocol: http://
    - Host: www.cs.umass.edu/
    - Username: ~verts/
    - Folder Path: coins105/
    - Resource: coins105.html

Protocol: http://

- HyperText Transport Protocol
- Type of Internet communication required
- One protocol among many
  - http://
  - ftp://
  - telnet://
  - gopher:// (obsolete)
  - etc.

Host: www.cs.umass.edu

- Read from Right-To-Left
  - Top Level Domain (TLD): .edu
  - Network: .umass
  - “sub” Network: .cs
  - Machine Name: www
- Not all Web addresses use WWW!
  - Our Web server: http://elsrv3.cs.umass.edu

Username: ~verts

- Folder on site belonging to a particular user
- Notice the “tilde” character ~
- Most keyboards the ~ is on the same key as `.
- Not all URLs use a username

Folder Path: coins105/

- All Web files are in a special folder called public_html (notice the underscore), never shown as part of a URL.
- Users may or may not create subfolders of public_html to contain related files:
  - .../~verts/ (no subfolder)
  - .../~verts/coins105/ (one level down)
  - .../~verts/coins105/classes/ (two levels down)
  - .../~verts/cmpsci119/ (one level down)

Resource: coins105.html

- The document file actually fetched.
- If present, it is the last part of a URL.
- Types of Web files, by extension:
  - Web files: .html / .htm
  - Text files: .txt
  - Pixel-Based Graphics files: .gif / .jpg / .png
  - Sound files: .mp3 / .wav
  - Scalable Vector Graphics: .svg
  - JavaScript Program Code: .js
  - Cascading Style Sheets: .css
What if the Resource Isn’t Specified?

- If not specified, assume index.html or index.htm as the default file to fetch (.htm dates from when MS-DOS and Windows PCs supported only 3-character extensions).
- Examples:
  - http://www.cs.umass.edu/
  - http://www.cs.umass.edu/~verts/
  - http://www.cs.umass.edu/~verts/coins105/
- All look for an index.html in different folders.

Top Level Domains

- Country Codes:
  - .us, .uk, .de, .dk, .fi, .ca, .cz, .jp, .ru, etc.
  - Now in native scripts: .РФ for .ru (Russian Fed.)
- Traditional Top-Level Domains (U.S. centric):
  - .edu .net
  - .com .org
  - .gov .mil
- Newer Domains:
  - .xxx .bike etc.

Basic Networking

- You connect your computer to a network directly through a wire (an Ethernet cable)
- You connect your computer to a network through a wireless access point (WiFi):
  - 802.11b (10 Mbit/s, rare anymore)
  - 802.11g (54 Mbit/s, common)
  - 802.11n (600 Mbit/s, now mature)
  - 802.11ac (800 Mbit/s...1.7 Gbit/s, soon)

The Client-Server Model

- Your computer/browser is the client,
- Remote computer containing desired resource is the server,
- There may be many computers in between,
- Each resource is requested separately so no single client can dominate the server,
- Requests from one client are interleaved with requests from other clients.

Packets

- Resource files are split into packets,
- Packets from one resource are interleaved with packets for other resources,
- Intermediate machines send packets to machines “closer” to their desired destination,
- Packets may follow different paths (and arrive out-of-order) depending on traffic or network damage.

Packet Sniffers

- Packet Sniffers are legitimate programs that examine packets to make certain they are constructed correctly,
- “Compromised” packet sniffers may watch for sensitive information (passwords, SSNs, credit card numbers, etc.),
- Treat email as postcards readable by anyone: never send sensitive info in the clear. Encrypt!
IP Addresses

- IP address is the unique identifier for a machine,
- Used by routers to guide packets,
- IPv4 four bytes (32 bits):
  - Format: ___.___.___.___
  - Many UMass addresses are 128.119.xxx.xxx
  - 4.3×10^9 (=4 billion) addresses, ran out in 2011.
- IPv6 eight two-byte words (128 bits):
  - ___.____.____.____.____.____.____.____
  - 3.4×10^38 addresses
  - Deployed, but still not widely used

DNS and IP addresses

- DNS (Domain Name Service) maps host names from URLs into numeric IP addresses.
- You type in a URL, a chain of DNS servers figure out what the IP address is and pass it back to your computer, which then knows how to make a proper resource request.
- You could type in the IP address directly!
  - http://128.119.240.37/

Building Web Pages

- Building a simple Web page is easy,
- Building a complicated Web page is hard!
- Many Web design tools exist:
  - Adobe Dreamweaver
  - Microsoft Expression Web
- We will build our Web pages using text editors:
  - Windows Notepad
  - Macintosh TextEdit
  - UNIX emacs

Simple Web Files (.txt files)

This is a simple Web page.
It is just a plain-text file, as created in Windows Notepad, or Mac Text Edit, or UNIX emacs (text editors). Browsers render it in monospace as shown in the editor.

Canonical Web Page (.html files)

```html
<html>
  <head>
    <title>My Web Page</title>
  </head>

  <body>
    Hello!
  </body>
</html>
```