

Introduction to Web Technology

Diana Inkpen

University of Ottawa
School of Information Technology and Engineering

2005, updated 2021

Content of the course

- **The Internet and the WWW.**
- Internet Connectivity.
- Basic Internet Services.
- Navigating the WWW.
- Building Web pages (HTML).
- Search engines. Information Retrieval.
- Business, Issues, and Controversies.

What is the Internet?

What is the Internet?

- The Internet is a global **network of networks** that enables computers of all kinds to directly and transparently communicate and share services throughout the world.

The Internet and the WWW

- Evolution of the Internet.
- Features of the Internet.
- Features of the WWW.
- Web servers.
- Web clients.
- Search Engines

What year was the Internet born?

The evolution of the Internet

- In 1969 the precursor of Internet is born: ARPAnet.
- ARPA = Advanced Research Projects Agency sponsored by the American Department of Defense (DOD).
- Designed to connect military research centers.
Distributed computer system able to survive a nuclear attack.
- Four nodes: at UCLA (Los Angeles), Stanford Research Institute, UCSB (Santa Barbara), and University of Utah.
- There were in fact two networks: MILnet and ARPAnet.

The evolution of the Internet (2)

- Problem: ARPAnet could connect only networks of the same type.
- In 1970, ARPA starts developing the Transmission Control Protocol / Internet Protocol (TCP/IP), a technology for **connecting networks of different types** (produced by different companies).
- Other networks appear, such as CSNET and BITNET.

The evolution of the Internet (3)

- The concept of **internetworking** was developed. A network connects through a **gateway** that facilitates communication with networks of different types.
⇒ The name **Internet** starts to be used.
- In 1980 NSFnet is created by National Science Foundation to connect supercomputer centers for academic research. By 1990, NSFnet replaces ARPAnet, which ceases to exist.

How big is the Internet?

Year	Computers	Newsgroups	WWW sites
1969	4		
1977	100		
1986	5,000	200	
1990	300,000	1,000	
1993	2 mil	6,000	600
1994	3 mil	10,000	10,000
1997	19 mil	70,000	1 mil
2002	350 mil	100,000	1 billion

2020

?

Features of the Internet

- The Internet = a network of networks.
- People around the world share ideas, information, comments, and stories.

Popular services:

- **Email** (electronic mail) is the most popular service. You can send and receive mail (electronic text), faster and cheaper than surface mail. Example email addresses:
username@computer-address diana@site.uottawa.ca
- **Web browsing** to find information. Example: www.cnn.com

Features of the Internet (2)

- **Mailing lists:** you can subscribe to lists on various subjects to receive emails, you can post (send) messages, you can unsubscribe. Some lists are moderated.
- **Newsgroups** are collections of messages on various subjects. Example comp.internet
- **FTP** (File Transfer Protocol). You can copy files from one computer to another over the Internet.

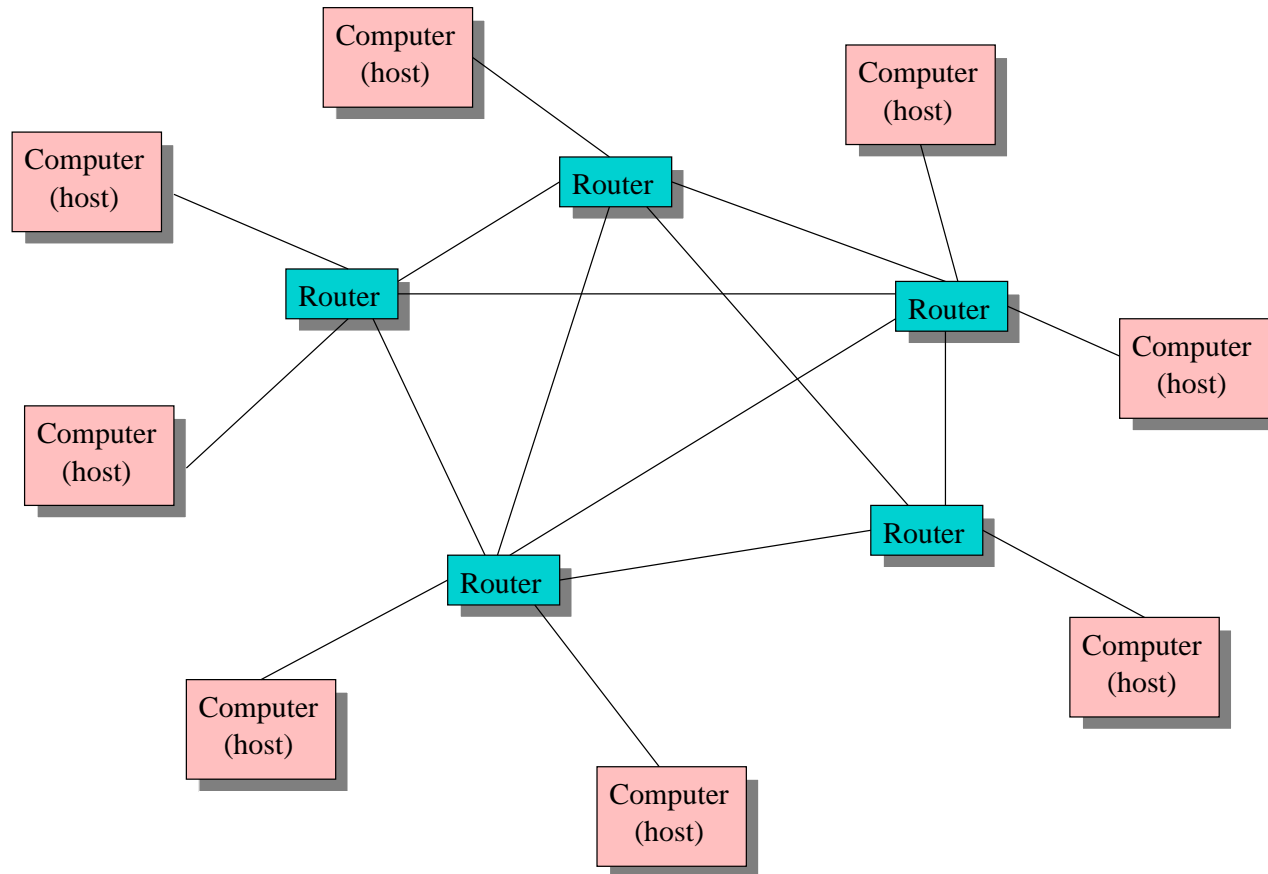
Features of the Internet (3)

- **Telnet** or remote login. Permits your computer to log onto another computer and use it as if you were there. You need to provides your username and password, for security reasons.
- **Chat rooms.** You can exchange messages with other people, anonymously (using a nickname).
- Internet services for companies: **e-commerce**, etc.

How does data travel on Internet?

- Data is divided into **packets**, that are transmitted over the Internet, and re-assembled at destination. No dedicated connection is needed (Packet switching).
- **Routing** = the packets can follow different routes on the Internet. The routers read the addresses and direct the packets (using their routing tables).

Internet routing



TCP/IP family of protocols

- A **protocol** = set of rules that allow different computers or devices to communicate with each other. Implemented in software.
- IP (Internet Protocol) moves packets between nodes.
 - ▶ IP forwards each packet based on a destination address (the IP number).
- TCP (Transmission Control Protocol) is responsible for verifying the correct delivery of data from client to server.
 - ▶ TCP adds support to detect errors or lost data and to trigger retransmission if needed.

Internet addresses

- The Internet authorities assign ranges of **IP addresses** to different organizations. The organizations assign groups of their numbers to departments.
- IP address = four numbers: **n.n.n.n**, where $0 \leq n \leq 255$.
Example: 130.63.236.137 **IP 6**
- **Domain names** were invented in order to avoid the need to memorize numbers.
Example: ugate.site.uottawa.ca

Internet addresses (2)

- **Domain Name Servers** (DNS) translate domain names into IP addresses.
- Hierarchical structure, read from right to left.
- Top-level domains:
 - ▶ .com .org .net .edu .int .mil .gov
 - ▶ country codes: .ca .de .fr
- Example: ugate.site.uottawa.ca
Example: www.microsoft.com

Internet addresses (3)

- Who assigns IP numbers?
Internet Assigned Numbers Authority (IANA).
- Who assigns domain names?
 - ▶ Top level: The Internet Corporation for Assigned Names and Numbers (ICANN).
 - ▶ Second level: Network Solutions Inc., and other registrars accredited by ICANN.
See <http://www.internic.org/> for a list.

Features of WWW

- World Wide Web (WWW, the Web) = a linked collection of documents residing on computers connected to the Internet, named Web servers.
- **Hypertext** = a text document that include **links** to other documents.
- **Hypermedia** = also includes images, video, and sound.
- WWW hypertext system began at CERN (European Centre for Nuclear Research) in Geneva, under Tim Berners-Lee, in 1989.

Features of WWW (2)

- Problem: if the documents have various formats, they cannot be viewed by everyone.
- Text documents (ASCII) can be viewed by everyone, but they cannot encode links, format, and multimedia.
- The documents on the Web (webpages) are text files that contain format information coded in **HyperText Markup Language (HTML)**.

Semantic Web

Web pages (HTML documents)

- Text appearance: bold, italic, font size.
- Links to other pages (hyperlinks).
- Images, sound, video can be embedded.
- Fill-in forms allow interaction with the user.
- Interfaces to databases. Dynamic HTML pages.
- Web pages can contain executables, such as Java applets.
- Web programming languages: Java, Javascript, Perl.

Example of HTML page

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<HTML>
<HEAD>
<TITLE>The ITec Programs at Atkinson</TITLE>
<META name="description" content="Science and Technology Studies">
</HEAD>
<BODY BACKGROUND="../images/pat.gif" BGCOLOR="#FFFFFF">
<TABLE WIDTH=300 BORDER=0 CELLPADDING=0 CELLSPACING=9>
  <TR><TD WIDTH=300><A HREF="itec_people.html"></A></TD></TR>
  <TR><TD WIDTH=300><A HREF="itec_faq.html"></A></TD></TR>
  <TR><TD WIDTH=300><A HREF="itec_core.html"></A></TD></TR>
  <TR><TD WIDTH=300><A HREF="#gen"></A></TD></TR>
  <TR><TD WIDTH=300><A HREF="#health"></A></TD></TR>
  <TR><TD WIDTH=300><A HREF="#write"></A></TD></TR>
</TABLE>
```

Web servers

- The webpages are on Web servers.
- Web servers run software that implements HTTP: httpd, Apache.
- Offer access to a collection of Web pages (a WWW site).
- URL = Uniform Resource Locator. Uniquely identify a webpage: `http://server-name/directory/filename`
- Example: `http://www.site.uottawa.ca/index.htm`
Example: `http://www.site.uottawa.ca/`

Web clients

- Browsers: Mosaic, Netscape Navigator, Internet Explorer.
- HTML pages are text files that contain information about their format.
- The Web browser uses this information to recreate the page on your computer.

HyperText Transfer Protocol

HTTP acts as an interface between a Web client software (i.e. a browser such as Netscape Navigator or Internet Explorer) and a Web server software. A **server** can serve many **clients**.

- Client's HTTP requests connection to the server.
- Server's HTTP sends back server status.
- Client's HTTP requests document.
- Server's HTTP sends document (if available).
- Connection is broken after document is downloaded on your computer. It is not maintained while you view it.

Search engines

- Find and retrieve information on the WWW.
- Document indexing = each document on the Web is represented in the index by its URL and some **keywords**.
- A search engine runs on a Web server. You access it through an HTML interface. You provide your query in the fill-in form. The search engine will give you URLs of pages that match your query.
- Search engines: Google, AltaVista, Yahoo, etc.
- Meta-search engines: Metacrawler, etc.

How big is the Web?

- Google announced (December 2001) that the its total collection of indexed document have passed 3 billions.

This figure is broken down into:

- ▶ 2 billions web pages (HTML documents, but also other files such as Adobe Acrobat or MS Word files).
- ▶ 330 millions images.
- ▶ 700 millions newsgroup messages.

2020 ?

The is also the invisible Web

Summary

You learned:

- What is the Internet?
- What is the WWW?
- Data transfer on the Internet.
- Hypertext, HTML pages.
- Web servers and Web clients.
- Search engines.

References

<http://www.isoc.org/internet/>

<http://www.internet-guide.co.uk/>

<http://library.albany.edu/internet/>

<http://www.northernwebs.com/bc/>

Books:

Introduction to the Internet, by Scott D. James, 2001

How networks work, by Frank Derfler and Les Freed, 2000