Topic 6: The Internet

1. The Internet
2. The World Wide Web
3. The Internet and business

Objectives

6a. Use and explain the basic terminology of the Internet
6b. Distinguish the Internet from the World Wide Web
6c. Explain the role of the Internet in commerce

Reading: Evans et al, Chs. 3, 8, 13
1. The Internet

- The Internet is a worldwide network of networks that links almost all computing devices worldwide
- *Agreed standards* enable exchange of data
- The Internet has a physical infrastructure and has governing bodies to make decisions
- Users connect to Internet via Internet Service Providers (ISPs)

Internet history

- Developed by Defense Advanced Research Projects Agency, DARPA, as a decentralized solution to danger of a disruption of military communication in case of nuclear war
- Mosaic (later, Netscape) browser, mid-1990s, enabled millions to use the Internet easily
6. The Internet

Protocols of the Internet

- **Protocol**: A set of rules for communicating, e.g., to encode or decode, or to correct transmission errors
- **Internet Protocol addresses**: four 8-bit numbers separated by dots
- **TCP/IP** (Transmission Control Protocol / Internet Protocol) converts messages into packets that are independently routed using multiple hops, in point-to-point communication
- **Domain names**: DNS servers convert names to IP addresses

IP addresses

- Internet Protocol (IP) handles addressing of packets
- Each device on the Internet has an IP address
- IP address is a 32-bit number, shown as four 8-bit values in {0..255}, and expressed in decimal notation, e.g., 204.109.1.127
- Servers usually have permanent (static) IP addresses, workstation IPs are usually assigned at time of Internet connection (dynamically)
6. The Internet

The Internet’s infrastructure

- **Backbone**: main paths, a set of networks owned by large commercial or governmental organizations, e.g., Verizon, AT&T, Sprint Nextel, Qwest
- **Physical medium**: optical carrier lines, carrying up to 40 Gb/sec on OC-768 lines
- **Major governing organizations**: Internet Society, Internet Engineering Task Force, Internet Architecture Board, Internet Corporation for Assigned Names and Numbers, World Wide Web Consortium
- **Internet exchange points** (IXPs) enable networks to connect directly to each other using switches

2. The World Wide Web

- **World Wide Web**: An abstract information space within the Internet in which servers host web pages, clients access Web servers via browsers
- **HTTP**: Hypertext Transfer Protocol, governs retrieval of web sites by users
- **URL** (Uniform Resource Locator)
  - the address of a web page
  - components: protocol, server, pathname
    (http://www.framingham.edu/~dkeil)
- **HTML** (Hypertext Markup Language) enables nonlinear, formatted documents
Steps of display of a web page

1. User clicks a hyperlink or enters a URL
2. User system (as client) sends URL and request to domain name server for IP address of URL’s server
3. Client sends request for page to URL’s server via IP address
4. Server sends HTML file to client

Interoperability requires standards

- All protocols support interoperability
- HTML enables web-based data to be self-formatting (appearance)
- CGI (Common Gateway Interface), an extension of HTML that supports scripts, e.g., for forms
- Java virtual machine supports applets downloaded from Internet to different platforms
- XML (Extensible Markup Language) enables data to be self describing (meaning)
- W3C (World Wide Web Consortium) sets standards
Support for Web-page interaction

- *Common Gateway Interface* (CGI) provides for running *scripts* in languages such as Perl, C, C++, on web browsers
- Some scripts support user interaction, such as form filling
- Scripts are run on server side
- Dynamic HTML (DHTML) allows modification of HTML after client loads it

Java applets

- Java is an object-oriented language designed especially for Internet use
- Programs in the Java language compiled to run on a browser’s Java Virtual Machine are *applets*
- Browsers for many platforms have JVMs that run the same compiled code
- The JVM cannot access files, make network connections, or allow programs to start other programs
Search engines and research

- A web crawler at Google, Yahoo, etc., reads billions of web pages and creates an index (CiteSeer reads online academic research)
- The search engine responds to queries by performing lookup in the index
- Services such as CiteSeer provide info on what papers have cited the paper in question
- Many papers are available in Acrobat form (.PDF); others in Postscript (.PS)
- With more information available, more critical tools need to be used to select info

Online data and search

- Libraries are still the best source of pre-1985 information; catalogs are online
- Data hierarchies narrow search, step by step: root, branch, leaf
- Evaluate sources for credibility
- Primary vs. secondary/tertiary sources
- Search techniques: Use of quotes, AND, NOT, refined search
- Question: Is Wikipedia a reliable research source?
Online research

- Preferred sources are those that have been peer reviewed by other experts
- Consider using Google Scholar
- To narrow your search, use two or more keywords you want to relate
- Proper bibliographic entries begin with author names and include publication information; bibliographic references may be checked at www.easybib.com
- Researchers must properly acknowledge all work used, including by quoting if text is used verbatim

Boolean operators in Web searches

- AND, OR, NOT
- AND is presupposed; e.g., a search for internet protocol implies a search for web pages that contain both the words “internet” and “protocol”
- To search for a phrase, use quotes, so that searching for pages containing the phrase “Internet Protocol” requires keyword “internet protocol” with the quotes
- * is a string wild card, % a character wild card
Web 2.0 (interactive Web)

- **Social networks** (Facebook, MySpace): Web sites that enable users to post “profile” info about themselves and to connect to “friends”
- **Instant Messaging**: Real-time text conversations
- **Podcasts**: Subscribers receive text, audio, or video files from providers as they are produced
- **Blogs**: public personal journals with support for user response
- **Wikis**: Web documents editable by multiple users as a collaboration tool

3. The Internet and business

- **E commerce**: “Conducting business using electronic data communication” (L. Snyder, Ch. 16)
- **Variations**: web shopping, electronic funds transfer, point-of-sale transactions, business to business commerce, networked meetings
- **Features**: diverse audience, client/server environment, transaction based, need for interoperability standards, unreliability of systems
- **Intranet**: Private network accessible only to employees but using Internet’s infrastructure
Models of E commerce

- B2B: Business to business
- B2C: Business to consumer (e.g., Amazon.com)
- C2C: Consumer to consumer (e.g., Ebay)
- B2G: Business to government
- Use of online transaction processing greatly reduces costs of commerce
- E commerce has disadvantages, e.g., lack of human interaction that may provide useful information to customer

Transaction processing and data integrity

- A transaction is an atomic (indivisible) series of steps
- Simultaneous requests require database lock to preserve atomicity
- Another term for this is serialized behavior
- Example: edits of a document by multiple authors should be serialized so that conflicting versions never coexist
Enterprise computing

- *Enterprise computing systems* serve organizations of hundreds or thousands of users
- May comprise many information systems (transaction processing, management information, decision support)
- *Scaling up* adds processors or storage to a machine; *scaling out* adds computers to a system
- Enterprise systems may share a database, distribute a database, or synchronize databases
- *System integration* connects hardware, software systems for data sharing

Technologies in enterprise computing

- *Blade servers*: slim rack-mounted server computers that occupy little space
- *RAID*: Redundant array of independent disks for fast fault-tolerant storage
- *Storage area networks*: Sets of devices acting as a node in a larger network
- *Mirroring*: Real-time maintenance of a duplicate of a data set
- *Network attached storage*: Serverless storage used in a network
- Input by magnetic ink, optical character recognition, and bar-code reading
- Multiprocessor architectures
Email and email attachments

• Most email today is *web based*, accessible from any Internet computer
• It is considered preferable to send PDF files rather than MS Word files
  – Some recipients may not be able to open an MS Word file
  – When any attached file is edited and saved, note that it is saved to a temporary folder; files to edit should be saved before editing

Information systems jobs

• *Network administrators* (engineers) plan, install, and test networks
• *Database administrators* assure DB integrity and the performance of DB servers
• *Web server administrators* install and maintain and assure access
• *Telcom technicians* manage telecommunications infrastructure and assure access
• *Software and web development*
Training for IT careers

- Two and four year degrees in
  - Computer science
  - Management information systems
  - Computer engineering
- Certifications, e.g.,
  - Cisco Systems
  - Microsoft’s MSCE

Internet concepts

atomic transaction  Internet  server
blog  Internet security  social network
browser  interoperability  spyware
client  intranet  TCP/IP
client/server computing  IP  URL
crawler  IP address  web page
data integrity  ISP  web service
domain name  packet  web site
e-business  pathname  Web 2.0
enterprise computing  protocol  Wi-Fi
HTTP  query processor  wiki
HTML  script  World Wide Web
HTTP  search engine  XML
hyperlink
6. The Internet

References

