

## Topic 5: Networks and distributed computing

*Motivation:*

- Much of the benefit of computing is in being connected with other users and with resources
- Users need some knowledge of underlying concepts
- The speed and connectivity enabled by networking have major social implications

Reading:  
P-O,  
Chs. 5-7

*Hypothesis:* **The networked character of computing today accelerates the rate of change in computing and in society**

### Networking concepts

asynchronous communication	HTTP	protocol
bandwidth	hyperlink	query processor
broadcast	Internet	routing
browser	interoperability	search engine
client	IP	secure communication
client/server computing	IP address	server
crawler	ISP	spyware
data compression	LAN	synchronous
data hierarchy	latency	TCP/IP
distributed computing	middleware	transaction processing
domain name	multicast	URL
e-business	network	web page
encryption	packet	web service
enterprise computing	pathname	web site
gateway	point-to-point	Wi-Fi
HTML		World Wide Web
		XML

## Eras of computer connectivity

- *Centralized* (mainframes, connected to processorless terminals)
- *Distributed* (networked PCs, 1980s-90s)
- *Networked* (communication via HTML using Internet, 1990s to present)
- *Internet*: a worldwide collection of computers and cables communicating using a standard
- 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

## Distributed computing

- Networking makes possible systems whose nodes are different platforms running different software
- *Middleware* helps coordinate computational and data resources on distributed systems
- Examples of distributed systems: Internet-based business systems, bank ATMs and retailer transaction handlers, enterprise databases
- *Software agents* may visit web sites

## Communication of data

- *LAN*: Local area network, as at FSC
- *Synchronous* (waiting turn) vs. *asynchronous* (not waiting turn) communications
- *Bandwidth*: Data capacity of a communication channel
- *Circuit switching* reserves a channel; *packet switching* enables concurrent use of a channel for many messages
- *Internet*: a worldwide network of networks

## The Ethernet and Wi-Fi standards

- *Ethernet* provides common access to LAN cable
- Packets are broadcast to all devices using a *bus*
- They are accepted only by devices to which the packets are addressed
- A communications *protocol* (rule) detects colliding sends, triggers resend
- *Wi-Fi*: a related standard for wireless networks
  - Reduced security
  - Slow transmission
  - Interference

## Client/server computing

- Client initiates all actions, server responds
- *Examples:* browser is client, web-site host is server serving HTML files; email uses server
- No continuous connection, just requests/responses
- Some web pages are generated specifically for one transaction, e.g., flight reservations
- *Middleware:* software, e.g., DBMS, that handles added layers (other than client and server) in  $n$ -tiered C/S relationship
- Server becomes client of middleware server:  
(C → S → M → S → C)

## The Internet

- *World Wide Web:* An abstract information space within the Internet in which servers host web pages, clients access Web servers via browsers
- *URL* (Uniform Resource Locator)
  - the address of a web page
  - components: **protocol**, **server**, **pathname**  
(<http://www.framingham.edu/faculty/dkeil>)
- *HTML* (Hypertext Markup Language) enables nonlinear, formatted documents
- Users connect to Internet via Internet Service Providers (ISPs)

## Protocols of the Internet

- *Protocol*: A set of rules for communicating, e.g., to encode or decode, to clean up noise
- *HTTP*: Hypertext Transfer Protocol, governs retrieval of web sites by users
- *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)

## Java applets

- 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
- Java is an object-oriented language designed especially for Internet use

## 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

## 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

## 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 research

- 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
- Most legitimate sources have been *peer reviewed* by other experts

## Malware

- *Virus* replicates self from file to file
- *Trojan horse* disguises self as a useful application
- *Worm* replicates from computer to computer
- *Botnets* are sets of computers (zombies) controlled remotely in background
- Email address spoofing: Email return addresses are not necessarily genuine, can deceive users into infecting their computers

## Spyware and popup hosts

- *Spyware*: Communicates via Internet without user permission
- Some undesired apps host popup messages
- This malware is installed either by tricking the user to click “Install,” or via security weaknesses in Windows associated with ActiveX controls
- Even opening an email attachment or visiting a web site can install malware

## File-sharing software

- *Songs and videos* available in MP3 and other compressed formats for MP3 players, iPods, etc.
- Software is generally free, enables users’ PCs to act as clients of other users’ PCs and to turn their computers into servers for other users
- One problem: This swapping occurs without asking permission of songs’ copyright owners
- Government and industry challenged Napster software developer
- Recent example of file-sharing: Limewire
- Some users report software is ill-behaved

## Privacy and security

- *Secure web sites* are password-protected
- *Secure connections* are those that encrypt the data communicated; protocol is *https*
- Encryption transforms a message into a difficult-to-read text; decryption transforms encrypted text to readable form
- Encryption/decryption use algorithms and a private key; some algorithms also permit a public key
- *Strong encryption* is hard to “break”
- Public keys enable client browsers to encrypt messages for sending securely to recipients

## Email privacy

- PATRIOT Act relaxes search-warrant rules
- Sniffing software may scan all messages entering or leaving an ISP to find email associated with a person under investigation
- Email intercepts by law enforcement require court order
- Employers are not legally restricted in looking at employees' emails
- Suggestion: emails are like postcards

## Transaction processing

- 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

## E-commerce

- *Definition*: “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

## 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

## 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

## Distributed systems

- *Tiered (layered) architectures*
  - single data storage tier
  - business logic tier to process transactions
  - many-noded user tier enabling data entry and showing query results
- *Grid computing*: coordination of computers as equals to achieve high processing speed, often in background
- *Cluster computing*: Fault-tolerant use of multiple dedicated servers, with load balancing

## Quality of service

- QoS is level of performance seen from user point of view, e.g., response time
- Mission of a multi-client system includes good QoS for *all* clients
- Metrics
  - *Response time*: delay before request fulfillment
  - *Throughput*: quantity of data processed per unit of time
  - *Accuracy*
  - *Downtime*
  - *Capacity*: in storage, number of users, connections, packets

## Email and email attachments

- Email is web-based (Yahoo, Hotmail, Gmail) or store-and-forward (ISP emails and networks like FSC), i.e., transferred from SMTP server and residing at client machine
- 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

## References

J. Parsons, D. Oja. *Computer Concepts*, 9<sup>th</sup> ed.  
Thomson, 2007.

L. Snyder. *Fluency with Information  
Technology: Skills, Concepts, and  
Capabilities*. Addison Wesley, 2006.