Computer Networks & Internet

CS211
Microcomputer Applications

Telecommunications

- Telecommunications, or data communications, is the transmitting of digitized data over a transmission media
- Data transmitted serially as bits
- Explosion of networking has led to digital convergence

Basic Components

- Transmission media
  - Electrical signals
    - Coaxial cable
    - Twisted-pair wire
  - Light pulses
    - Fiber-optic cable
    - Wireless microwave or radio signals

Basic Components (contd)

- Protocols
  - Rules that govern how data is transmitted over the network
  - There are many protocols
    - TCP/IP, the protocol of the Internet, is the most common
    - Computers can have multiple protocols
  - To communicate between computers, all must be using the same protocol

Basic Components (contd)

- Modem
  - For temporary connection
  - Can be internal or external device
  - Modulator/Demodulator
    - Modulator converts analog to digital
    - Demodulator converts digital to analog
  - Typical modem used with existing phone wire
    - Cable modem works with cable TV system
    - ISDN modem works with digital phone lines

What is a network? (Revisited)

- Two or more computers connected together makes a network
- Connection can be temporary or permanent
- Reasons to network
  - Share data
  - Share resources (storage, I/O, multimedia)
  - Share software
**Basic Components (contd)**
- Network Interface card (NIC)
  - For permanent connection
  - Internal card
  - Includes a transceiver
    - Transmitter sends bits onto media
    - Receiver reads bits from media
  - Encodes/decodes bits into correct form for media
    - electrical pulses, light pulses or radio signals

**Other Components (contd)**
- Hub
  - Connects PCs within a star network
- Router
  - Connects multiple LANs normally through common carrier

**Other Components (contd)**
- Server
  - Dedicated high-end PC
  - Used in resource sharing
    - Print server
    - Data server
    - Communications server (or Mail server)

**Network Types (contd)**
- Permanent networks
- Home network
  - 2 - 4 PCs in a home setting
- LAN = Local Area Network
  - Office suite or building
  - Most common type of network
- MAN = Metropolitan Area Network
  - City Network

**Network Types**
- WAN = Wide Area Network
  - Wide geographic area
  - Connection of multiple LANs
  - Normally depends on common carrier to provide connection
  - Examples:
    - Internet
    - Phone system

**Operating System**
- Permanent networks require an operating system
  - Peer-to-peer
    - Small network of 10 or less users
    - All machines are equal (“peers”)
    - Uses regular OS like Windows 9x, NT Workstation
Operating System (contd)

- Server-based
  - Dedicated servers for authentication
  - Centralized control
  - Needs Network operating system (NOS)
    - Microsoft NT or 2000 server, Novell NetWare

Network Topology

- Physical layout of the permanent network
  - Bus
    - Network devices are daisy-chained in a line
  - Ring
    - Similar to a bus except that both ends connect to form a ring
  - Star
    - All devices connect to a single central device

Access Methods

- Only one computer can transmit at a time
  - Simultaneous transmission = collision, both transmissions are lost and need to be retransmitted
- Token Access
  - Used with ring topology
  - One token travels the ring
    - Must have free token to transmit
    - No collisions

Access Methods (contd)

- Ethernet
  - PC listens on media before transmitting
  - Collisions may occur
    - more PCs = more collisions

Internet

- Global WAN, collection of interconnected networks, or LANs

Going Online

The Internet
What is the Internet?

- Largest computer network in the world
  - 1997 estimated 16.1 million “host” computers
  - 1995 estimated 4.8 million “host” computers
  - 1999 150 million users with over 800 million web pages accessible.
  - Network of networks worldwide
- Internet “host”
  - Computer that is connected at all times

What is the Internet?

- Networked computer is a “server”
  - Provides service or information
- Internet servers classified by types of information they offer:
  - Mail server: provides electronic mail
  - WWW server: provides hypertext documents
  - FTP server: makes files accessible to users
- Single Computer can act as number of different servers
  - Host can be both mail and WWW server

What is the Internet?

- Internet is globally-linked network of computers
- Provides people, business & corporations, educational institutions, governmental agencies & countries with ability to electronically communicate over long-distances.

History of the Internet

- Came into existence during Cold War Era
  - U.S. Department of Defense became concerned with how different governmental services would be able to communicate in the event of nuclear war (1958)
  - US Air Force sponsored RAND Corp Project RAND
    - President Dwight D. Eisenhower saw need for the Advanced Research Projects Agency (ARPA) to keep the U.S. at the forefront of technology
    - Conceived of network of computers designed to function despite loss of substantial part of system
- RAND Report On Distributed Communications describes this system
- Each computer or node capable of sending, routing, & receiving information
  - By taking messages & breaking them into different parts—packet switching—
    - Leonard Kleinrock invents packet-switching technology (1962)
    - Then sending them along separate routes to their eventual destination
- Each node treated equally as no single computer is hub: if large parts of system destroyed, information still gets through

J.C.R. Licklider, head of computer research at ARPA, articulates vision of worldwide network
- Larry Roberts publishes a paper proposing the ARPAnet network
- The Day the Infant Internet Uttered its First Words by Leonard Kleinrock
**Internet’s Ancestor**

- ARPANET in 1969
  - Advanced Research Projects Agency Network
  - Department of Defense experiment:
    - reliable networking
      - dynamic rerouting: link is disrupted, traffic automatically rerouted to other links
    - linking DOD & military research contractors
      - universities performing military-funded research
      - ARPAnet unveiled at UCLA

- E-mail introduced by Ray Tomlinson, 1972
- ARPANET was widely successful
  - too much traffic to handle
- ARPANET divided into:
  - MILNET
    - Military Network for military sites
  - ARPANET
    - new, smaller with nonmilitary sites

**The Internet’s Ancestor**

- UCLA, Stanford Research Institute, University of California, Santa Barbara, & University of Utah connected in 1969
- Stanford, MIT’s Lincoln Labs, Carnegie-Mellon & Case-Western Reserve University connected in 1971
- NASA/Ames, Mitre, Burroughs, Rand & University of Illinois added shortly

**Internet’s Ancestor**

- ARPANET & MILNET connected by IP
  - Internet Protocol, technology to route traffic between 2 networks as needed
  - both networks spoke “IP” and could exchange messages
  - IP designed for 1000s of networks to communicate
  - each computer on network able to communicate with any other computer on network
  - current technology: central computer with remote terminals

**The Internet in the Classroom**

- By 1980 university computing moving
  - from small number time-sharing machines to large number smaller desktop workstations
- Advantages of timeshare systems:
  - shared directories of files & e-mail
- Desktop workstations ran UNIX
  - OS developed at UC-Berkeley
  - UNIX included network software
- Desktop workstations sold with necessary network hardware

**The Internet & NSF**

- National Science Foundation set up 5 supercomputer centers for research use
  - researchers send programs using ARPANET to supercomputer centers for processing
- ARPANET did not work, so NSF creates NSFNET to connect supercomputer centers
  - also, setup regional networks to connect users in each region, and NSFNET connecting all regional networks
The Internet & NSF

- NSFNET was very successful
  - By 1990 most of ARPANET traffic moved to NSFNET
  - ARPANET shutdown
  - Supercomputer center plan failed
    - Too expensive, technical problems
- NSFNET permits only research & educational traffic
- Independent commercial IP network services

The Internet TCP/IP

- TCP/IP, a standard of communication
  - ARPA's "Network Control Protocol" or NCP
  - "Transmission Control Protocol" or TCP
    - Breaks information into different packets at its source
      - Then puts them together at its final destination
- "Internet Protocol" or IP
  - Does addressing of information
    - Ensuring packets are sent across multiple nodes
    - And networks using different standards

Did Al Gore invent the Internet?

- In 1988 Albert Gore, then a Tennessee senator, proposes the National Research and Education Network, which would provide top computing facilities to research communities and schools.

Basic components

- Hardware
  - Modem for temporary connection
  - NIC for permanent connection
- Communications software or browser
  - Netscape Communicator, MS Internet Explorer
  - Browser plug-ins
    - Shockwave
    - RealPlayer
    - Acrobat Reader

Basic components (contd)

- Username and password
- TCP/IP
  - Internet protocol

World Wide Web (WWW)

- In 1992 the World Wide Web is born, introduced by Tim Berners-Lee.
- First audio/video multicasts are broadcast over Internet
- In 1993 Internet browser MOSAIC is introduced at University of Illinois by Marc Andreeson
### Creation of the WWW

- In 1994 Real Audio introduced to Internet which allows one to hear audio in near real time. Radio HK, first 24-hour Internet-only radio station, starts broadcasting.

---

### The Internet & WWW

- WWW is different from Internet.
- WWW added:
  - graphics, sounds, pictures, colors, videos to text-only Internet
  - Browser is tool to view
  - Hyperlinks weave information together
  - Hyperlinks are text or graphics on a Web page
    - clicked on by mouse, causes browser to load pages
    - easy to use
    - 27 to 50 million people browsing today

---

### Internet Connection

- Many different ways to access the Internet:
  - ISP (Internet Service Providers)
    - AOL, CompuServe, Prodigy
    - connect to Internet using PC and modem
    - plug and play solution:
      - pop in CD-ROM, follow instructions, on the Net!
  - ISP that offer connection time
    - Browser programs
      - Netscape or Internet Explorer
      - list of ISPs in US, and elsewhere

---

### What can I do on the Internet?

- exchange email
- participate in discussion groups (news groups)
- socialize using chat groups
- retrieve data files
- shop & e-commerce
- virtual tours
- read newspapers

---

### Where does Internet info come from?

- everywhere, anyone can publish on Internet
- no central authority or organization
- vast, growing online library in which anyone can publish anything they want
- consider source of information
- First Amendment protects freedom of electronic expression

- http://www.yahoo.com/ & select Links:
  - Society and Culture, Issues and Causes,
    - Civil Rights, Censorship, Censorship and the Net

---

### Getting around in the Internet

- Uniform Resource Locator (URL)
  - Internet address
  - http://www.vc.cc.tx.us/~swagner
  - http:// is the access protocol of Hypertext Transport Protocol
  - other access methods are ftp (file transfer protocol) and news (newsgroup), etc.
URL (contd)
- www.vc.cc.tx.us is the domain name, or unique name that identifies an Internet host site
  - domain names are read from right to left
- /~swagner is a file or web page found at the host site

Hyperlinks
- Click text or graphics to link to different web page
- Has associated URL “behind” text/graphic

Ports and Search engines
- Yahoo, Infoseek, Excite, Lycos, Google
- Portal
  - choose category, subcategory, etc
- Search engine
  - type in keyword(s)
- Results vary between portal sites and search engines

World Wide Web (www)
- user-friendly
- web pages or multimedia documents
- hyperlink between documents

File Transfer Protocol (ftp)
- upload and download files

Telnet
- remote login

Communication applications
- E-mail or electronic mail
- Newsgroups or electronic bulletin boards
- Internet Relay Chat (IRC) or “chat rooms”
- Instant messaging

E-mail
- is a method to communicate with others online

E-mail addresses:
- username@hostname
  - sopami2@uwm.edu
  - john@aol.com
  - michelle.p@irs.gov
  - tblavre@packers.com
  - joeblow@yahoo.com
  - tgreen@usa.net

E-mail addresses composed of 2 parts separated by @ sign:
- Part before @ sign: mailbox
  - usually your personal name
- Part after @ sign: domain
  - usually computer you are using
  - domain can be group of computers
The Internet & e-mail

- sompami2@alpha2.uwm.edu
  - before the @ sign: sompami2 is mailbox
  - after the @ sign: (domain group) alpha2 is alpha machine
  - uwm identifies institution
  - edu identifies type of institution
    - org, net, com, gov, mil

The Internet & e-mail

- Signature files: short text file appended to email messages with personal info:
  - email address, phone numbers, fax numbers
  - quotations
- “Netiquette” dictates “sig” file no more than 5 lines

The Internet & e-mail

- Windows system:
  - sig file can be named anything
  - sig file can be stored anywhere
  - tell mailer what and where sig file is
- Netscape and IE, options/preference dialog
- UNIX system:
  - .signature
  - stored in user’s home directory

The Internet & e-mail

- To: Brett Favre <bfavre@packers.com>
- Cc: gary@uwm.edu
- Bcc: john@usa.net
- Attachment:
- Subject: Go Packs Go!
- ----- Message Text ----- 
  Yo Brett! Be aggressive! Super Bowl is not out of reach, dude.
- ...
- Wutnipong Sompamitr
  - Ph.D. candidate, CS department
  - University Of Wisconsin-Milwaukee

The Internet & e-mail

- Receive e-mail message
- Save messages in folders or mailboxes
  - inbox
- Forward e-mail message
  - Forward option in header
- Message Reply
  - Reply to sender only
  - Reply to all
Mail filtering is a method to dispose of incoming mail that match certain criteria. Save messages from your mom in a certain folder.

Spam is Junk e-mail. It arrives on any day that ends in "y" and is an electronic advertisement.

How do you get on a Spam-List?
- Actively participate in newsgroups
- Actively participate in mailing lists
- Register with an online service that makes your e-mail address public
- White page directories of e-mail addresses
- Never really sure

Anti-Spam Measures
- Do NOT ask to be REMOVED
  - Indicates to spammers e-mail address is active
  - Return address may not be valid (spammers do not want to be identified)
- Stop spam FAQ:
- Just say "Delete"!!!
- Use Filters
- Get another e-mail address

When attaching a binary file to send
- MIME encodes binary file
- When recipient receives e-mail with attached binary file
  - MIME decodes binary file
- UNIX environment uses own encoding & decoding utilities
  - uuencode & uudecode

How private is e-mail?
- Not totally private
- Any recipient of your e-mail can forward it to others

Rule of Thumb:
Do **NOT** e-mail anything you would not want posted next to the water cooler.

32-bit host address defined by IP in STD 5, RFC 791
- Usually represented in dotted decimal notation
  - 128.121.4.5
    - Address can be split into network number (or network address)
    - And host number unique to each host on network
    - Sometimes also subnet address
Internet Address

- Dot Notation consists of 1 to 4 numbers in hex or octal or decimal
- Each leading number represents 8 bits of address (high byte first) and last number represents rest
  - E.g. address 0x25.32.0xab represents 0x252000ab.
  - Most common form is 4 decimal numbers
    - E.g. 146.169.22.42.
- Many commands will accept address in dot notation in place of hostname

Internet Address

- Manner address is split depends on its "class", A, B or C as determined by high address bits:
  - Class A - high bit 0, 7-bit network number, 24-bit host number. n1.a.a.a 0 <= n1 <= 127
  - Class B - high 2 bits 10, 14-bit network number, 16-bit host number. n1.n2.a.a 128 <= n1 <= 191
  - Class C - high 3 bits 110, 21-bit network number, 8-bit host number. n1.n2.n3.a 192 <= n1 <= 223

Search Engine - Northern Light

- natural language search engine
- enter search in same way you would ask a question or speak a word or phrase
  - what is a good recipe for chicken risotto?
  - bicycle tours in the Canadian Maritimes
  - What is the capital of Sweden?
- searches all sites in its database which contain all of search terms
- http://northernlight.com/

Natural Language Search

- Number of results in search not affected by common words
  - is, a, for, in, and the,
- Do affect the relevance of those results
- Relevancy can increase or decrease based on
  - number of occurrences of query terms.
  - and presence of query terms in document titles
- Result lists are sorted with most relevant appearing closest to the top.

Search on simple words

- More words you enter, more on-target your results will be.
- Examples:
  - ski resorts Vermont
    - (instead of skiing)
  - ergonomic workstation mouse keyboard
    - (instead of ergonomics)
**Northern Light Search Engine: Boolean Search**

- specialized search technique which relies on a system of logic developed by mathematician George Boole (1815-64)
- Boolean searching uses three expressions to determine results: **AND**, **OR**, and **NOT**

**Northern Light Search Engine: Boolean Search (cont.)**

- Using Boolean expressions, you can limit your results to very specific information
- You may also use parentheses in Boolean searches to create complex Boolean expressions
- No limit to the level of nesting which you can use in a query

**Northern Light Search Engine: “AND” Boolean Search**

- cinema AND (China OR Taiwan)
  - search will return results which contain the words “cinema” and “China” or which contain the words “cinema” and “Taiwan” anywhere in the text

**Northern Light Search Engine: “AND” Boolean Search**

- ”War AND Peace”
  - will return documents with the phrase “War and Peace”
    - (such as discussion of the book by Leo Tolstoy), whereas
  - War AND Peace
    - will return documents that contain the word War and the word Peace.

**Northern Light Search Engine: “OR” Boolean Search**

- Mozart OR Beethoven
  - search will return results which contain either the name “Mozart” or the name “Beethoven” anywhere in the text
- Use **OR** to retrieve documents that include any of the search words (rather than most.)

**Northern Light Search Engine: “NOT” Boolean Search**

- ”set dancing” NOT (contra OR country)
  - search will return results which contain the phrase “set dancing” but do not contain the words “contra” or “country” anywhere in the text
- Use **NOT** to indicate a word that must not appear in the documents
If Boolean operators appear in quotes, they are interpreted as a search term, rather than a Boolean operator:

- “War AND Peace” will return documents with the phrase “War and Peace”
- War AND Peace will return documents that contain the word War and the word Peace

“” (double quotes)

- To focus your search on an exact phrase, enter the phrase in double quotes:
- “The Name of the Rose” This search is more likely to return results which contain the entire phrase “the name of the rose”
- Searches in single quotes will return the same results as returns that are not quoted.

If you require that some words or phrases appear in your search results but not necessarily others, use a + sign before that word or phrase that you must have.

- Any words that are not preceded by a symbol will be considered “may have” terms, which means they may appear in the page, but they are not required to appear:

- Scotland +Orkneys Skye
- This search would return results which contain the terms “Scotland” and “Orkneys” but which may or may not contain the term “Skye”.
- Pages which do contain “Skye” would appear higher in the results list than pages which do not.

If certain terms must not appear in your results, eliminate pages containing those words or phrases by placing a - sign before the word or phrase.

- Any term that is not preceded by a symbol will be considered a “may have” term, as above:

- bond -James stock
- This search would return pages which contain the word “bond” but not the word “James”.
- Pages may also contain the word “stock” but it is not a required term.
- Pages which did contain the word “stock” would appear higher in the results list than pages which do not.
Example of + and - searches

- Must include space between the + and - symbols and terms which immediately precede them, but not between the symbols and the terms that follow them:

Northern Light Search Engine: Syntax Search Examples

- +dolphins -NFL
  SPACE

- +recipes for +"chocolate cake" -nuts
  SPACE  SPACE

Truncation Symbols

- 2 truncation symbols (wildcards) in queries
- Must have at least 4 non-wildcard characters in a word before you introduce a wildcard
  - * (asterisk) can be used to replace multiple characters
  - % (percent) symbol is used to replace only one character.

Note that Northern Light automatically stems most common plural and singular forms of words
- A search on cat will also return results containing the word cats
- A search on cats will return results containing the word cat

Northern Light Search Engine: Asterisk Sign ‘*’ Syntax Search

- To search for variants of a word root, use a * symbol to replace a series of letters:
  - theolog*
    - This search would return all pages which contain variants on the root "theolog," including "theology," "theological," "theologia"

Northern Light Search Engine: Percent Sign ‘%’ Syntax Search

- If you wish to replace a single letter in a word, use the % sign.
  - Useful for words that have different regional spellings, or which are commonly misspelled
**Northern Light Search Engine: Percent Sign ‘%’ Syntax Search**
- gene%logy
  This search would return pages containing the words "genealogy" and "geneology" (a common misspelling)
- critic%e
  This search would return pages containing the words "criticize" (American spelling) and "criticise" (British spelling)

**Truncation Symbols**
- Use multiple truncation symbols within a single word
- Expanded words found by using truncation symbols will not affect the relevancy ranking of those sites:
  
  Oso you may want to use additional related terms in your query to ensure that your results are accurate and meaningful

**Northern Light Search Engine: Field Searching**
- Northern Light has classified its index into searchable fields.
  - You can search these fields by adding the field delimiter followed by a colon [:] before your search term.
  - Field delimiters can be used with any search form.

**Northern Light Search Engine: Field Searching**
- Field searches may contain any of the above syntax including Boolean expressions, and may also be used within larger Boolean expressions

**Northern Light Search Engine: Field Searching**
- URL: searches for a URL or partial URL
- TITLE: searches for a document title or partial title
- PUB: searches for Special Collection documents by journal title

**Northern Light Search Engine: Field Searching**
- COMPANY: searches for information about a particular company
- TICKER: searches for information about a company based upon the company’s stock ticker
- TEXT: searches within the text of the document or website
Northern Light Search Engine: Field Searching

- **RECID**: searches for a specific Special Collection document by Northern Light Document ID number (found on Northern Light Special Collection article summary page)
- **SORT: date** sorts results by date, from newest to oldest

Search Engine Relevancy Factor

- Person types keywords (or phrases) to describe what he/she is looking for
- Search engine lists websites containing those keywords in order of relevancy
- Search engine relevancy is largely determined by following major search engine relevancy factors

Keyword Frequency

- How often keyword appears on page or in an area on page
- In general, more times keyword appears on page, more relevant it will be to that search

Keyword Density

- Keyword density is ratio of keyword or keyphrase to total words (depth) on a page
- Keyword density is critical aspect of search engine optimization
- To rank highly, keyword density must not be too high or too low
  - Density of 1% to 7% is considered good
Keyword Position

- Keyword position is search result
  - Higher keyword position is, more likely site will be looked at

Meta Tags

- Meta tag is used by search engines to allow them to more accurately list your site in their indexes
  - HTML tag, use keywords in it

What is Yahoo!

- The Yahoo! Directory is a manually created, browse-able (and searchable) collection of site listings aggregated by human editors
  - Human and editorially filtered directories of unified sources of online information

What is Yahoo!

- If sites related to your search are found within the Yahoo! Index
  - They will be listed under the Categories and/or Web Sites headings (found on the toolbar just above the top search result).
- If sites related to your search are found within the Google index,
  - They will be listed under the Web Pages heading

What is Google?

- Yahoo! partner that specializes in full-text indexing of the entire Web
- Google is completely automated search engine: a robot or spider (computer program) automatically crawls the Web, capturing every word on every page crawled

What is Ask Jeeves?

- Another search engine using natural language.
- You can search pictures, products, news, and web pages.
- www.ask.com
  - www.ai.com
  - www.askjeeves.com
What's a Spider?

- Many search engines utilize optimized computers and software known as "Spiders" to collect information about web sites.
- This info is stored in search engine's database.
- "Spiders" visit URLs to gather information about sites that it needs in order to index your page to their database.

Search Tips

- Comparison of search engines by Ziff Davis, http://www.zdnet.com/pclmag/stories/reviews/0,6755,2327803,00.html
- Northern Light search tips, http://www.northernlight.com/power.html
- General search tips from MS Internet Explorer, http://home.netscape.com/escapes/search/tips_general.html
- Dogpile search instructions, http://www.dogpile.com/help/topics
- Web Searching Tips: http://www.searchenginewatch.com/