Communications and networking

- **history and background**
  - telephone system
  - local area networks
- **Internet**
  - architecture: what the pieces are and how they fit together
  - names and addresses: what's your name and number?
    Domain Name System, IP addresses
  - routing: how to get from here to there
    traceroute, ping
  - fundamental protocols and layers
    IP, TCP
  - higher level protocols and services:
    HTTP, SSH, SMTP, IMAP, ...; web, email, instant messaging, peer to peer, ...
- **Web**
  - what makes it work: URL, HTTP, HTML, browser

Early technology:

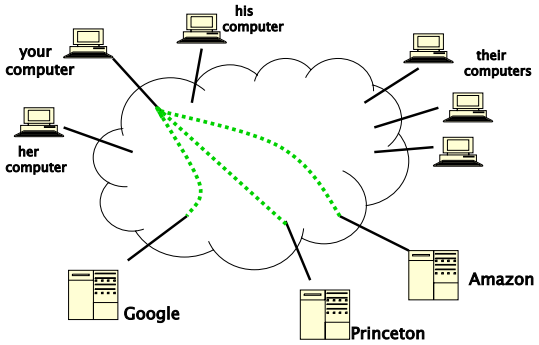
- **runners ("Rejoice, we conquer")**
- **riders ("Neither snow nor rain nor heat nor gloom of night...")**
- **signal lights ("1 if by land and 2 if by sea")**
- **fires, mirrors, flags, drums, voices, pigeons, …**

- **Optical telegraph (Chappe, 1792)**
  - faced many of the problems of modern networks
  - protocols: rules by which independent parties exchange info
  - bandwidth limitations: how to send information faster
  - error detection and recovery
  - security and privacy: protecting info from eavesdroppers & imposters
    Gerard Holzmann, *The Early History of Data Networks*
- **Telegraph (Morse, 1850's)**
  - similar issues (and a new technology that killed the old one)
    Tom Standage, *The Victorian Internet*

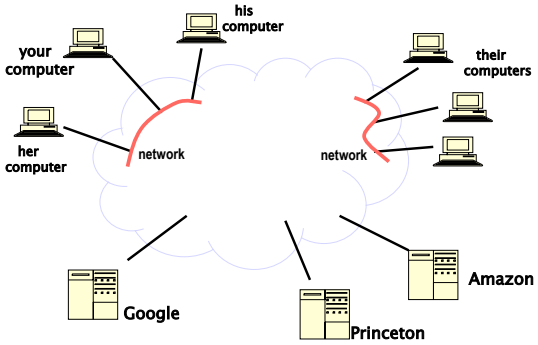
Telephone system (Alexander Graham Bell, 1876)

- **organizing principles, all based on voice traffic:**
  - voice calls need only a narrow bandwidth channel
  - a call uses a dedicated circuit, with long setup and hold times
  - telephone number is a unique identifier
  - fixed routing for a specific call
  - parallel signaling network; data separated from control
  - simple user interface: all intelligence inside network
  - guarantees on quality of service; high reliability
- **running out of some resources (area codes, 800/888/877/866, ...)**
- **traffic model changing rapidly (cell phones, data, ...)**
- **technology changing rapidly (wireless, Internet, ...)**
- **worldwide evolution from highly regulated and/or government-operated to deregulated / private**
  - highly competitive
  - incumbent carriers threatened by Internet

The Internet



Local Area Networks



Local Area Networks; Ethernet

- **a LAN connects computers ("hosts") in a small geographical area**
- **Ethernet is the most widely used LAN technology**
  - developed by Bob Metcalfe & David Boggs at Xerox PARC, 1973
  - each host has a unique 48-bit identification number
  - data sent from one host to another in "packets" of 100-1500 bytes
    including source and destination address and error checking bits
    typical data rate 10-1000 Mbits/sec; limits on cable length

packet:

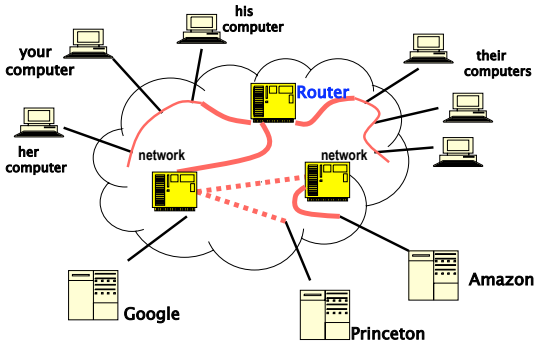
hdr	src	dest	TYPE	data	check
8	6	6	2	46-1500 bytes	4

- **"broadcast" technology: data sent to all connected hosts**
  - sender broadcasts, but if it detects someone else sending, stops, waits a random interval, tries again
- **wireless Ethernet uses radio to carry signals**
  - logical behavior is exactly like a wired Ethernet

Connecting networks (wide area networks / WAN)

- **how do we connect LANs to each other?**
  - LANs may have different properties
  - may be far away
- **names & addresses now needed to find other networks and hosts**
- **routing needed to find a path if multiple networks are involved**
  - can't have each network connected directly to all others
- **need to agree on format of information and how it is exchanged**
  - especially if networks are different kinds that use
    - different format for packets
    - different physical and electrical properties
    - different names and addresses themselves
- **how do we handle errors, delays, overload, etc.?**
- **how does it scale as the number of networks gets really big?**

Gateways and Routers



The Internet

- **a huge number of independent networks that are connected**
  - NOT a giant computer or a single network
  - each network may serve many host computers
- **nearby computers are connected by a local area network**
  - most often Ethernet
- **information travels through networks in small "packets"**
  - each packet independent of all others
    - like individual envelopes through the mail
  - all packets have the same format
  - standard protocols for format of info and behavior
- **networks connected by specialized gateway computers (routers)**
  - route packets of information from one network to the next
  - gateways continuously exchange routing information
- **each packet passes through multiple gateways**
  - gateway passes packet to gateway that is closer to ultimate destination
  - gateways usually operated by different companies

Internet History

- **1961: packet switching concept (Leonard Kleinrock, MIT, UCLA)**
- **1960's: ARPANET, funding from DARPA (Dept of Defense)**
- **1969: first Internet communication**
- **1972: first network email**
- **1973: basic protocols: TCP/IP (Bob Kahn *64, Vint Cerf)**
- **1980's: National Science Foundation funding, NSFNet (Al Gore)**
- **1980's: Internet Engineering Task Force for technical decisions**
- **1990's: commercialization, Web, dot-com boom**
- **2000: dot-com bust**
- **2010: universal availability**

- **for lots more, http://www.isoc.org/internet/history/**

The Internet

Essential Features

- Interconnection of large number of **heterogeneous, independent networks**
  - competitors
- These networks **connected** by **gateway computers**
- Information is packaged in **packets** with **standard form**
- **Protocals** dictate how **packets must be handled** as travel from gateway to gateway through Internet
  - Protocals define Internet

Basic mechanisms

- **names** for computers
  - princeton.edu, finance.yahoo.com, www.whitehouse.gov, kernighan.net, ...
- **addresses** for identifying networks and computers
  - each has a unique number like 128.112.128.81 (IP address)
  - central authority assigns numbers to networks
  - each host computer has unique address (32 bit integer in IPv4), assigned locally according to what network it's on
- **Domain Name System** to convert names to addresses
- **routing** for finding paths from network to network
- **protocols** (rules) for packaging and transporting information
  - IP, or "Internet Protocol": a uniform transport mechanism at IP level, all information is in a common format
  - below IP, different hardware uses different protocols
  - above IP, higher-level protocols for handling web pages, mail, login ...

Internet (IP) addresses

- **each network and each connected computer has an IP address**
- **IP address: a unique 32-bit number in IPv4 (IPv6 is 128 bits)**
  - 1st part is network id, assigned centrally in blocks
    (Internet Assigned Numbers Authority -> Internet Service Provider -> you)
  - 2nd part is host id within that network
    assigned locally, often dynamically

net part	host on that net

- **written in "dotted decimal" notation: each byte in decimal**
  - e.g., 128.112.128.81 = princeton.edu

128	112	128	81

10000000	01110000	10000000	01010001

Domain names

- **a hierarchical naming scheme**
  - central authority (ICANN) manages top level of names
- **top level domains include .com, .edu, .gov, .xx for country XX**
  - and newer domains like .biz, .info, .name , …
- **each domain delegates responsibilities to levels below**
  - for administration and translation into addresses
- **each level is responsible for names within it**
  - princeton.edu handles all of princeton
  - delegates cs.princeton.edu to a CS machine
  - CS department manages names within, e.g., bolle.cs.princeton.edu
- **names impose logical structure, not physical or geographical**

ICANN

- **Internet Corporation for Assigned Names and Numbers**
  - non-profit corporation, established 1998 by Dept of Commerce
  - technical coordination of the Internet
  - www.icann.org
- **"coordinates the assignment of the following identifiers that must be globally unique for the Internet to function:**
  - Internet domain names
  - IP address numbers
  - protocol parameter and port numbers
- **"coordinates the stable operation of the Internet's root server system"**

Domain name system (DNS)

- **DNS converts names to IP addresses and vice versa**
  - princeton.edu == 128.112.128.81
  - carnegiehall.org == 63.131.135.199
  - kernighan.org == 67.18.147.42
- **hierarchical searching for addresses**
  - central authority controls top level domain names (.com, etc.)
  - delegates responsibilities for searching to levels below
  - each level responsible for names and addresses within it
    princeton.edu handles address lookup for all of princeton
    delegates cs.princeton.edu to a CS machine
- **top level domains handled by 13 root servers**
- **lookup for a name asks a local name server first**
  - if not known locally, asks a server higher up, …, to root server
  - recently-used names are cached to speed up access
- **names impose logical structure, not physical or geographical**

Inter-network Routing

- **networks are connected by gateways or routers**
- **routing rules direct packets from gateway to gateway**
  - trying to get closer to ultimate destination
- **at gateway**
  - if destination on connecting network, use physical address and local routing
  - Otherwise use network id to determine next gateway and network to reach it
- **routers exchange information frequently about routes**
  - which nets each knows about and number of hops to reach them
- **autonomous system: group of networks under single authority**
  - interior gateway protocols exchange routing info within a single AS
- **traceroute: how do you get from here to there?**

Protocols

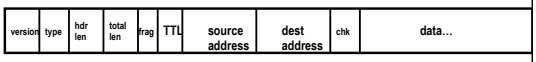
- **precise rules that govern communication between two parties**
- **TCP/IP: the basic Internet protocols**
- **IP: Internet protocol (bottom level)**
  - all packets shipped from network to network as IP packets
  - no guarantees on quality of service or reliability: "best effort"
  - each physical network has its own format for carrying IP packets
- **TCP: transmission control protocol**
  - creates a reliable 2-way data stream using IP
    errors are detected and corrected
  - most things we think of as "Internet" use TCP
- **"application-level" protocols, mostly built from TCP**
  - HTTP (web), SMTP (mail), SSH (secure login), FTP (file transfer), …
- **UDP: user datagram protocol**
  - simple unreliable datagram protocol (errors not detected)
  - used in DNS, remote file systems, voice over IP …

Packets

- **packet: a sequence of bytes carrying information**
  - usually over a network connection
- **bytes have a specific sequence, format, organization**
  - usually as specified in a protocol
- **typical network packet includes**
  - source (where it comes from)
  - destination (where it goes to)
  - size or length information (how big is the data part)
  - miscellaneous information (type, version, info to detect errors, ...)
  - the data itself ("payload")
- **typical sizes range from**
  - a few bytes
  - 150-1500 (Ethernet packets)
  - 100-65000 (IP packets)

What's in an IP packet

- **a "header" that contains**
  - protocol version, type of packet, length of header, length of data
  - fragmentation info in case it was broken into pieces
  - time to live: maximum number of hops before packet is discarded
    each gateway decreases this by 1
  - source & destination addresses (32 bits for IPv4, 128 bits for IPv6)
  - checksum of header information
    redundant info to detect errors in header information only, not data itself
  - etc.; about 20-40 bytes in header
- **actual data**
  - up to 64 KB of payload
  - IPv4:



IP: Internet Protocol

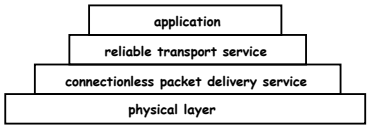
- **IP provides an unreliable connectionless packet delivery service**
  - every packet has full source & destination addresses
  - every packet is independent of all others
- **IP packets are *datagrams***
  - individually addressed packages, like postcards in the postal system
    "connectionless"
  - stateless: no memory from one packet to next
    each packet is independent of others, even if in sequence and going same place
  - unreliable: packets can be lost or duplicated ("best effort" delivery)
  - packets can be delivered out of order
  - contents can be wrong (though error rates are usually very low)
  - no speed control: packets can arrive too fast to be processed
  - limited size: long messages have to be split up and then reassembled
- **higher level protocols use IP packets to carry information**
- **IP packets are carried on a wide variety of physical media**

TCP: Transmission Control Protocol

- **a reliable 2-way byte stream built with IP**
- **a TCP connection is established to a specific host**
  - and a specific "port" at that host
- **each port provides a specific service**
  - SSH = 22, SMTP = 25, HTTP = 80, ...
- **a message is broken into 1 or more packets**
- **each TCP packet has a header (src, dest, etc) + data**
  - header includes checksum for error detection, and sequence number to preserve order and detect missing or duplicated packets
- **each TCP packet is wrapped in an IP packet and sent**
  - has to be positively acknowledged to ensure that it arrived safely
    otherwise, re-send it after a time interval
- **TCP is the basis of most higher-level protocols**

Higher level protocols

- **SSH: secure login**
- **SMTP: mail transfer**
- **HTTP: hypertext transfer -> Web**
- **protocol layering:**
  - a single protocol can't do everything
  - higher-level protocols build elaborate operations out of simpler ones
  - each layer uses only the services of the one directly below
  - and provides the services expected by the layer above
  - all communication is between peer levels: layer N destination receives exactly the object sent by layer N source



Encapsulation

- **each piece of data at one level is wrapped up with a header and sent as a packet at the next lower level**
- **lowest level is what moves across specific network**

How are things connected?

- **local nets connected to local Internet Service Provider (ISP)**
- **these in turn connect to regional ISPs**
- **and then to larger ones like UUNet, AT&T, Sprint, …**
- **traffic exchanged at Internet exchanges**
  - large and small, formal and informal, profit and non-profit
- **bandwidth (bit-carrying capacity) of connections is usually higher the larger the ISP**
  - cable modem, DSL 500 Kbps – 4Mbps (you to your ISP)
  - dedicated telephone lines 1.5-45 Mbps (local ISP, big company to ISP)
  - optical fiber 155 Mbps and up (large carriers)

Coping with bandwidth limits

- **data flows no faster than the slowest link**
- **limits to how much data can pass per unit time**
  - no guarantees about packet delivery
  - no bandwidth, delay or quality of service guarantees
    IP telephony is hard because voice traffic requires limited delay, jitter
    video is somewhat easier but needs a lot more bandwidth
- **caching**
  - save previous data so it doesn't have to be retrieved again
- **compression, encoding**
  - to improve use of available bandwidth
  - don't send redundant or unnecessary information
    text, code, etc., can be compressed and recreated exactly
    music, pictures, movies are compressed with some information discarded
- **home connectivity**
  - telephone modem 56 Kbps
  - ADSL, cable modem 1-4 Mbps
  - wireless 1-50 Mbps
  - fiber 50 Mbps?

Internet Ideas

- **packets versus circuits**
  - different models (mail vs phone)
- **names and addresses**
  - what is it called, how to find it
- **routing**
  - how to get from here to there
- **protocols and standards**
  - Internet works because of IP as common mechanism
    higher level protocols all use IP
    specific hardware technologies carry IP packets
- **layering**
  - divide system into layers
    each of which provides services to next higher level
    while calling on service of next lower level
  - a way to organize and control complexity, hide details

Internet technical issues:

- **privacy & security are hard**
  - data passes through shared unregulated dispersed media and sites
    scattered over the whole world
  - it's hard to control access & protect information along the way
  - many network technologies (e.g., Ethernet, wireless) use broadcast
    encryption necessary to maintain privacy
  - many mechanisms are not robust against intentional misuse
  - it's easy to lie about who you are
- **service guarantees are hard**
  - no assurance of reliable delivery, let alone of bandwidth, delay or jitter
- **some resources are running low**
  - especially IPv4 addresses
  - IPv6 (the next generation) uses 128-bit addresses
    acceptance has been slow but is growing
- **but it has handled exponential growth amazingly well**