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
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: 

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>hdr</td>
<td>src</td>
<td>dest</td>
<td>type</td>
<td>data</td>
<td>check</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>46-1500 bytes</td>
<td>4</td>
</tr>
</tbody>
</table>

- "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
- protocols 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?
The Internet

- **millions 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 (including wireless)
- **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
- 2020: fragmentation?

- for lots more, http://www.isoc.org/internet/history/
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 128.112.136.10 (IP address)
  – central authority assigns numbers to networks
  – each host computer has unique address (32 bit integer in IPv4, 128 in IPv6), 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

<table>
<thead>
<tr>
<th>net part</th>
<th>host on that net</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>112</td>
</tr>
<tr>
<td>136</td>
<td>10</td>
</tr>
</tbody>
</table>

- written in "dotted decimal" notation: each byte in decimal
  - e.g., 128.112.136.10  = www.princeton.edu

<table>
<thead>
<tr>
<th>128</th>
<th>112</th>
<th>136</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000000</td>
<td>01110000</td>
<td>10001000</td>
<td>00001010</td>
</tr>
</tbody>
</table>
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 myriad newer domains like .biz, .info, .name, .xxx, ...
- 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 computer
  - CS department manages names within, e.g., rinse.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
  – www.princeton.edu == 128.112.136.10
  – carnegiehall.org == 45.60.73.146
  – kernighan.com == 69.46.29.42

• 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
Routing

- networks are connected by gateways or routers
- routing rules direct packets from gateway to gateway trying to get closer to ultimate destination
- routers exchange information frequently about routes
- bottom-up view:
  - gateways move packets from one network to another based on network id
  - if destination on the same network, use physical address
  - otherwise send to a gateway, which passes it to another network
- top-down view:
  - networks connected only through gateways
  - core has a small set of gateways that exchange complete routing info about which nets it knows about and number of hops to reach them
  - autonomous system: group of networks under single authority
  - passes reachability info to core for use by other autonomous systems
  - interior gateway protocols exchange routing info within a single AS
- traceroute: how do you get from here to there?