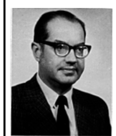
Large Networks
Layering and Hierarchy
(and some war stories)

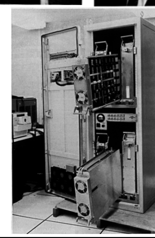
Lecture 23

May 4, 2009

6.02 Spring 2009

A Brief History of the Internet





ARPA: 1957, in response to Sputnik

Paul Baran
- Early 1960s: New approaches for survivable comms systems; "hot potato routing" and decentralized architecture, paper on packet switching over digital comm links

Donald Davies, early 1960s
- Coins the term "packet"

Len Kleinrock (MIT thesis): "Information flow in large communication nets", 1961

J. Licklider & W. Clark (MIT), On-line Man Computer Communication

L. Roberts (MIT then ARPA), first ARPANET plan for time-sharing remote computers

Initial Baby Steps

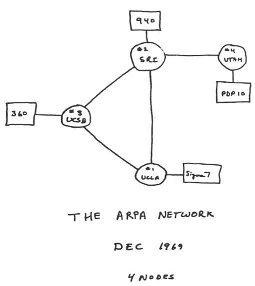
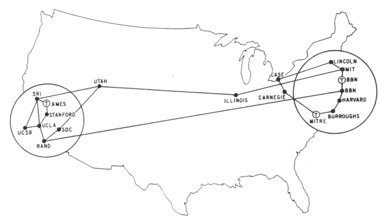


FIGURE 6.2 Drawing of 4 Node Network
(Courtesy of Alex McKenzie)

September 1971

1970, ARPANET hosts start using NCP; first two cross-country lines (BBN-UCLA and MIT-Utah)

"Hostile overlay" atop telephone network



MAP 4 September 1971

1970s: Internetworking Develops

- 1972: modified ARPANET email program, French CYCLADES network (sliding windows)
- 1973: ARPANET becomes international
- 1973-75: Internetworking effort (Cerf, Kahn, et al.)
  - Developed TCP and IP (originally intertwined)
- 1978: TCP and IP split; TCP at end points, IP in the network (layering)

The Internetworking Problem

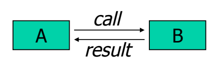
- Many different kinds of networks: want to interconnect them
  - Networks have very different characteristics (packet sizes, error rates, delays, etc.)
- And want to design for growth

Handling Heterogeneity

- Make it very easy to be a node or link on the network (best-effort)
- Universal *network layer*: standardize addressing and forwarding
- Switches maintain no per-connection state on behalf of end points

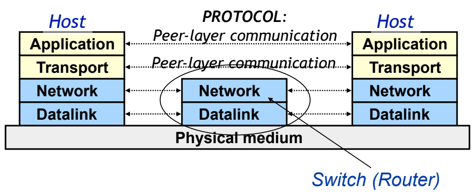
Layering: Modular Network Design

- Layering is a particular form of abstraction



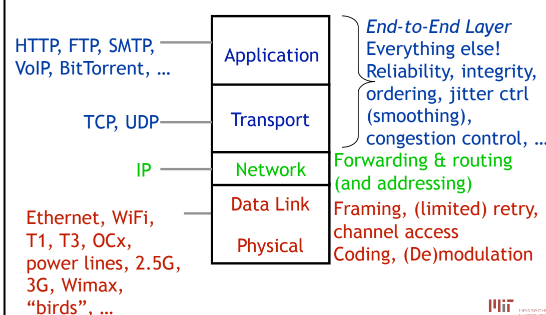
- The system is broken into a **vertical stack of protocols**
- The service provided by one layer is based solely on the service provided by layer below
  - This is the "up/down" interface
- *Peer interfaces* across the network implement *communication protocols*

Layering Interfaces

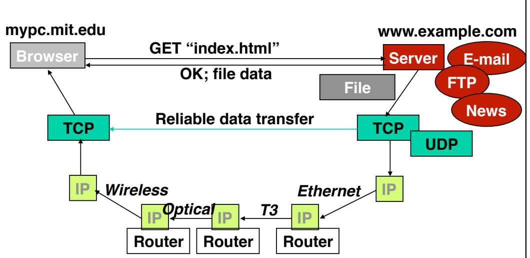


- Link and network layers are implemented everywhere
- The end-to-end layer (i.e., transport and application) is implemented only at hosts (end points)

Layering in the Internet

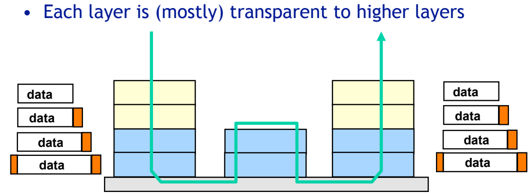


An Example

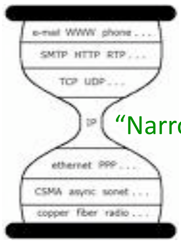


How Layering Works

- Each layer adds/strips off its own header (and possibly a trailer)
- Each layer may split up higher-level data
- Each layer *multiplexes* multiple higher layers
- Each layer is (mostly) transparent to higher layers



The Internet "Hourglass"

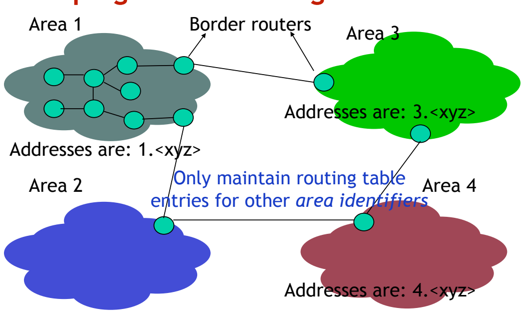


- Many applications, transports, and link protocols
- All over IP: universal network layer

Handling Growth: Topological Addressing

- Per-node routing entries don't scale well
- Solution: Organize network hierarchically
  - Into "areas" or "domains"
  - Similar to how the postal system works
  - Hide detailed information about remote areas
- For this approach to work, node addresses must be *topological*
  - Address should tell network *where* in the network the node is
  - I.e., address is a *location* in the network

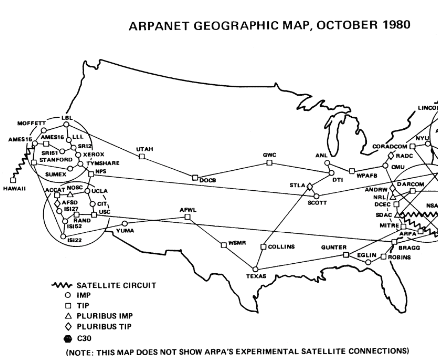
Toplogical Addressing: Ideal Case



And one could have areas within areas, etc.

Reliability

- Routing protocols take care of routing around failures
- End points take care of recovering from lost packets
- Nodes "try their best" to deliver packets over links, but not too hard
- Reliable systems from unreliable components
  - Approach stems from *digital abstraction*
  - Packet switching won't be possible otherwise
- Approach that has usually worked well…



Honeywell 516

Famous ARPANET Routing Bug

- ARPANET used link-state routing much like what we studied
- LSA used circular sequence space of size M
  - Assumed any reordering introduced by network was only a small fraction of M
- To determine if the sequence number wrapped, each node compares the arriving number A to the current number C
- **If |A-C| <= M/2, then accept A, else ignore**

ARPANET Problem

- ARPANET used M = 64
- What happens if three updates with seq numbers 8, 40, 44 exist simultaneously?
  - On seeing 8, will accept 40 next
  - On seeing 40, will accept 44 next
  - On seeing 44, will accept 8 next
- If, somehow, these three are in the system at once, we're in trouble because they will keep getting re-flooded

This Problem Took Down ARPANET on October 27, 1980

- Each node detected integrity of routing software
- Each link had error detection
- Problem triggered by memory error on Honeywell 516
  - 44: 00101100
  - 40: 00101000
  - 08: 00001000
- **No network-level error detection could have prevented problem**
  - Need an "end-to-end" check
- Solution?

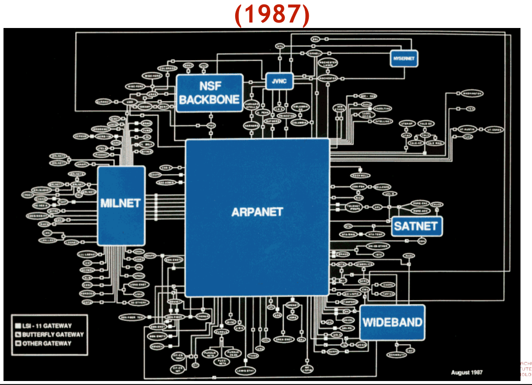
1980s: Rapid Growth

- 1981: many networks (BITNET, CSNET, Minitel, …)
- 1982, US DoD standardizes on TCP/IP
- 1984, Domain Name System (DNS) introduced
- 1986, NSFNet started, NNTP, MX records, big outage in New England
- 1986: Congestion collapse episodes
  - Problems with bad timeout settings
  - Adaptive timers, TCP congestion control solution
  - Athena network file system congestion problems (bad timeout settings)

Severe Congestion Episodes

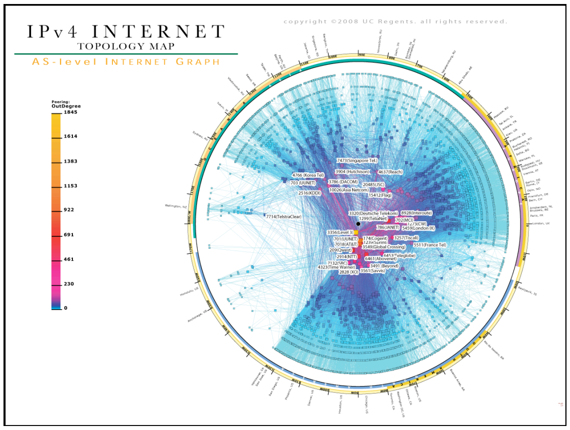
- NFS: Fixed timeouts
- TCP: poor RTT and timeout estimation
- Solution: what we studied, plus more
- RTT estimation as a filtering problem
- Timeout estimation using deviation and mean to reduce spurious retransmissions

Some Decentralized Administration (1987)



1990s

- 1990: no more ARPANET
- 1991: WWW released (Berners-Lee)
- Mid-1990s: NSFNet gets out of backbone
  - ISPs take off
- BGP4: **Path vector protocol** between competing ISPs, who must yet cooperate
- 1996: telcos ask for IP phones to be banned
- 1996-2001: .com bubble starts and bursts
- 2000s: now truly international; more non-PC devices than PCs on the Internet
  - Wireless and mobility take off
- And many exciting challenges remain



BGP Routing Problems in Today's Internet Example: Pakistan "Hijacks" YouTube

- Pakistani government wants Pakistan Telecom (PT) to block YouTube
  - PT advertises its own host as the destination for YouTube's IP address range
- Misconfiguration causes this advert to propagate to PT's ISP (PCCW, Hong Kong)
- PCCW sees that this advert is "more specific" than what it has, so accepts
  - Propagates to other ISPs, who also accept
- Soon, much of the Internet can't reach YouTube!