CSE/EE 461 Lecture 5
Network Scaling

Tom Anderson
tom@cs.washington.edu

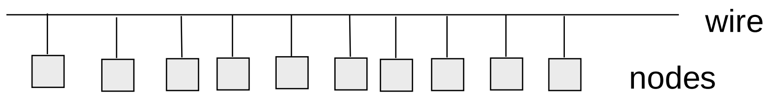
Last Time

How do multiple parties arbitrate access to a communication channel (wire or wireless)?

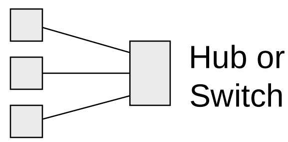
- Ethernet solution
  - Carrier sense – don't send if you can hear someone else
  - Collision detect – **sender** senses if someone else is sending at same time; jams so other sender can detect collision (RTT + epsilon)
  - Binary exponential backoff – choose random delay, starting with small mean, and increasing as detect more collisions

Ethernet

- Ethernet: first practical LAN, Xerox 1980's (802.3)
  - 1-persistent CSMA/CD with binary exponential backoff
  - 10 Mbps over coaxial cable, passive taps
  - Manchester encoding, preamble, 32 bit CRC



- Newer versions
  - Fast (100 Mbps), gigabit (1 Gbps)
  - Switched, point to point wires



Ethernet Frames

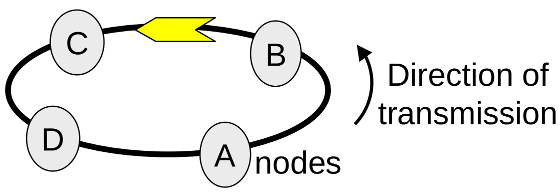
Preamble (8)	Source (6)	Dest (6)	Len (2)	Payload (var)	Pad (var)	CRC (4)

- Min frame 64 bytes, max 1500 bytes
- CSMA/CD jam period is 48 bits
- Max length 2.5km, max between stations 500m (repeaters)
- Addresses unique per adaptor; globally assigned
- Broadcast media:
  - ARP, multicast, promiscuous mode monitoring

Ethernet Evaluation

- Fairness -- backoff favors latest arrival
  - max limit to delay
  - no history -- unfairness averages out
- Stable performance under increasing load
  - Works very well in practice
- What happens as bit rates increase?
  - Need to shorten wires and increase frame size

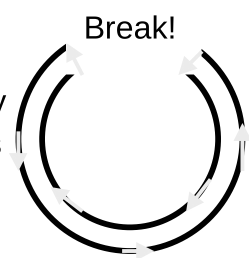
Token Ring (802.5)



- "Permission to send" token rotates around loop
- When token arrives
  - if no data to send, forward the token
  - if data to send, remove token and inject packet
  - remove packet and re-inject token at **sender**
  - Maximum token holding time (THT) bounds access time

FDDI
Fiber Distributed Data Interface

- Roughly a large, fast token ring
  - 100 Mbps and 200km vs 4/16 Mbps LAN
  - Dual counter-rotating rings for better reliability
  - Complex token holding policies for real-time traffic
- Token ring advantages
  - No contention, bounded access delay
  - Support fair, reserved, priority access
- Disadvantages
  - Complexity, reliability, scalability



Why Did Ethernet Win?

- Reliablity
  - Token ring failure mode -- network unusable
  - Ethernet failure mode -- node detached
- Cost
  - Passive tap cheaper to build than active forwarder
  - Volume => lower cost => volume => lower cost …
- Scalability
  - Repeater: copy all packets across two segments
  - Bridge: selectively repeat packets across two segs
  - Switch: bridge k segments; Hub: repeater for k segs

Wireless Communication

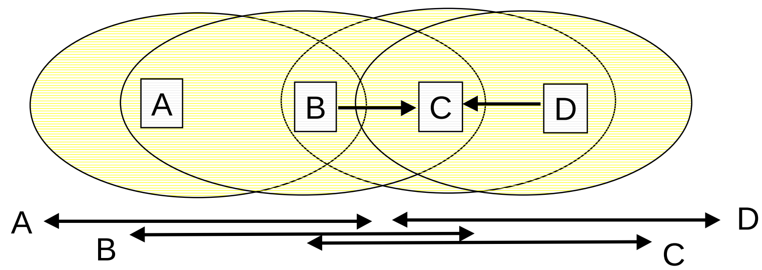
Wireless is more complicated than wired …

- Cannot detect collisions
  - Transmitter swamps co-located receiver
- Different transmitters have different coverage areas
  - Asymmetries lead to hidden/exposed terminal problems

CSMA with Collision Avoidance

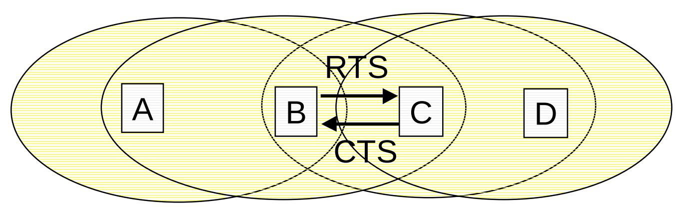
- Since we can't detect collisions, we avoid them
  - CSMA/CA as opposed to CSMA/CD
- If medium busy, choose random backoff interval
  - Wait for that many idle timeslots to pass before sending
- If a collision is inferred, retransmit with binary exponential backoff (like Ethernet)
  - Use CRC and ACK from receiver to infer "no collision"
  - Again, exponential backoff helps us adapt "p" as needed

Hidden and Exposed Terminals



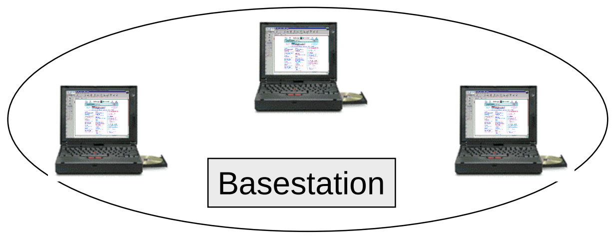
- Hidden terminals: B and D can both send to C but can't hear each other
- Exposed terminals: B, C can hear each other but can safely send to A, D

RTS / CTS Protocols (MACA)



- B asks C: Request To Send (RTS)
- A hears RTS and defers to allow the CTS
- C replies to B with Clear To Send (CTS)
- D hears CTS and defers to allow the data
- B sends to C

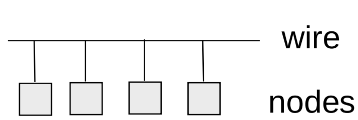
802.11 Wireless LANs

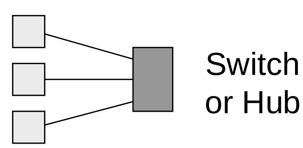


- Emerging standard: wireless plus wired system or pure wireless (ad hoc)
- Avoids collisions (CSMA/CA (p-persistence), RTS/CTS
- Built on new links (spread spectrum, or diffuse infrared)

Limits of a LAN

- One shared wire can limit us in terms of:
  - Distance
  - Number of nodes
  - Performance





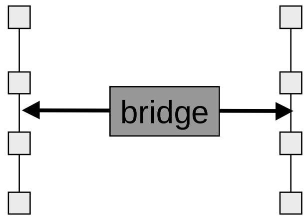
- How do we scale to a larger, faster network?

Network Scaling Options

- Scaling beyond one wire
  - Intra-network: bridges, hubs, switches
    - Myrinet, Ethernet, ATM
  - Inter-network: routers
    - scaling and heterogeneity
    - need cross-network addressing standard (IP)
- Forwarding: how does the switch/router know where to send the packet?
  - source routing (Myrinet)
  - global addresses (Ethernet, IP)
  - virtual circuits (ATM, MPLS)

Bridges and Extended LANs

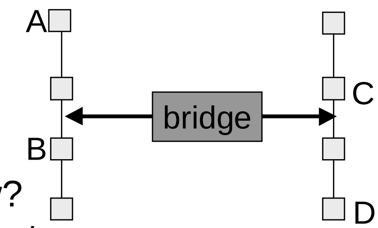
- "Transparently" interconnect LANs with bridge or switch
  - Receive frames from each LAN; selectively forward to the others
  - Each LAN is its own collision domain



a

Backward Learning Algorithm

- To optimize overall performance:
  - Shouldn't forward A→B or C→D, should forward A→C and D→B

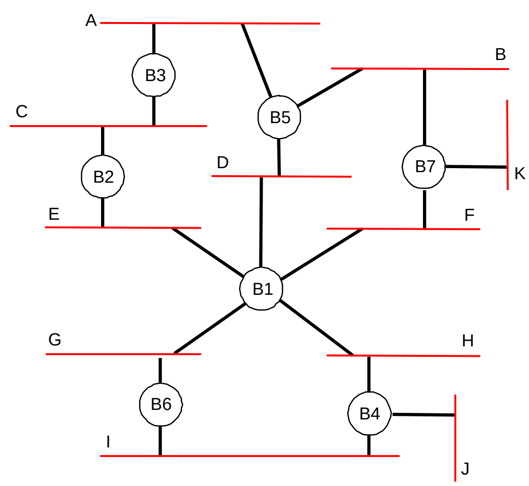


- How does the bridge know?
  - Learn who is where by observing source addresses
  - Forward using destination address; age for robustness

Why stop at one bridge?

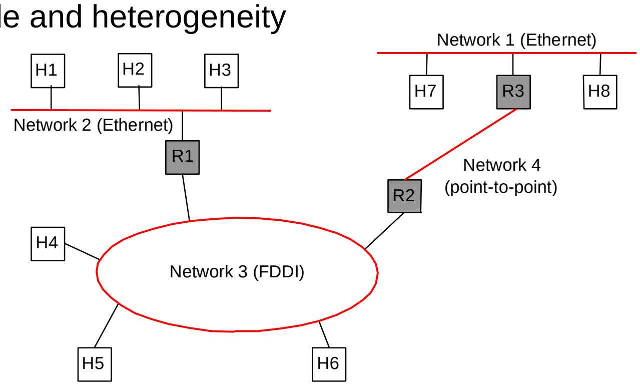
Need to know where to forward!

Analogous to IP routing



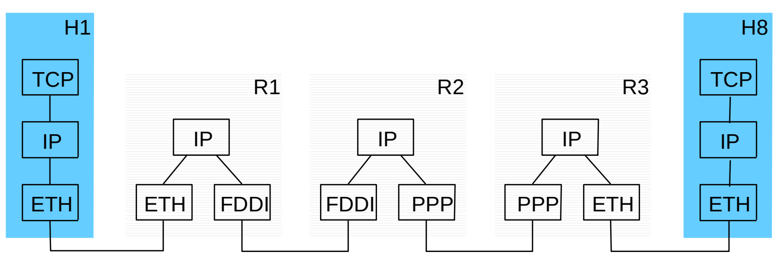
Internetworks

- Set of interconnected networks, e.g., the Internet
  - Scale and heterogeneity



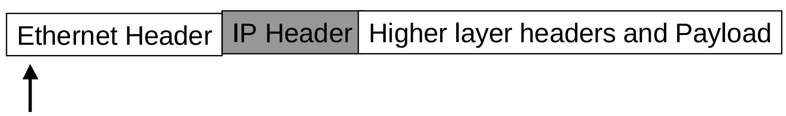
In terms of protocol stacks

- IP provides a global addressing layer across heterogeneous networks
- Routers interconnect networks
- Address Resolution Protocol (ARP) translates from IP addresses to MAC addresses
- Packet is the IP data unit; frame is the LAN data unit



In terms of packet formats

- View of a packet on the wire on network 1 or 2
- Routers work with IP header, not higher
  - Officially, to do otherwise is a “layer violation”
  - But many commercial products combine routing with higher layer functions (ex: layer 7 switch)
- Routers strip and add link layer headers

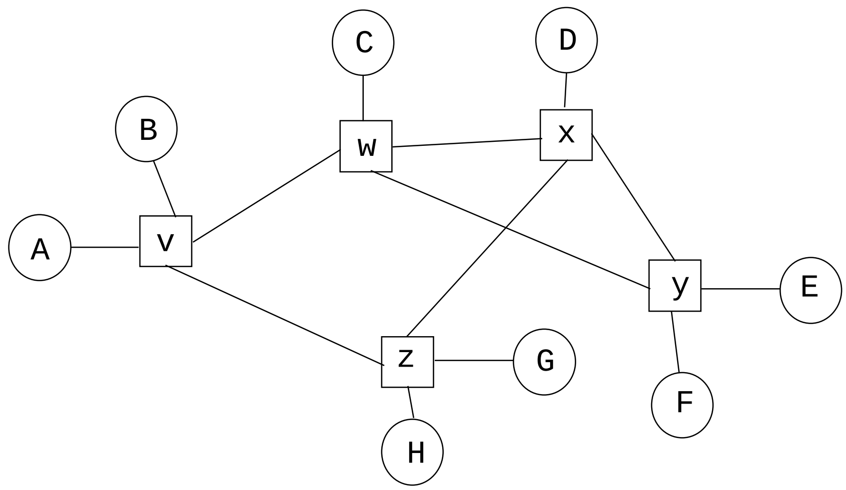


Front of packet to left (and uppermost)

Forwarding and Routing

- Forwarding is the process that each router goes through for every packet to send it on its way
  - Involves local decisions
- Routing is the process that all routers go through to calculate the routing tables
  - Involves global decisions

Forwarding Mechanics Example



Source Routing (Myrinet)

- List entire path in packet
  - Ex: A-> F (east, south, south)
- Advantages
  - Switches can be very simple and fast
- Disadvantages
  - Variable (unbounded) header size
  - Sources must know topology (incl. failures)
- Typical use: machine room networks

Global Addresses (Ethernet, IP)

- Each packet has destination address
- Each switch/router has forwarding table of destination -> next hop
  - At v and x: F -> east
  - At w and y: F-> south
  - At z: F-> north
  - Forwarding decision made independently for each arriving packet
- Distributed algorithm for calculating tables

IP Router Table Size

- One entry for every host on the Internet?
  - 100M entries, doubling every year
- One entry for every LAN?
  - every host on LAN shares prefix
  - still too many, doubling every year
- One entry for every organization?
  - every host in organization shares prefix
  - requires careful allocation

Virtual Circuits (ATM, MPLS)

- Each connection has destination address; each packet has virtual circuit ID (VCI)
- Each switch has forwarding table of connection -> next hop
  - at connection setup, allocate virtual circuit ID (VCI) at each switch in path
  - (input #, input VCI) -> (output #, output VCI)
    - At v: (west=A, 12) -> (east=w, 2)
    - At w: (west=v, 2) -> (south=y, 7)
    - At y: (north=w, 7) -> (south=F, 4)

Virtual Circuits

- Advantages
  - more efficient lookup (smaller tables)
  - more flexible (different path for each circuit)
  - can reserve bandwidth at connection setup
- Disadvantages
  - still need to route connection setup request
  - more complex failure recovery