ICS 351: Today's plan

- TCP congestion control
- Ethernet switching
- Spanning Tree Protocol
Congestion Collapse

- reminder: the network hardware might be working fine, but if the software fails, the network goes down
- e.g. if the routing tables include loops, packets will not get delivered
- imagine a retransmission mechanism where, when a packet is lost, I resend the lost packet and also a new one
- if a packet is lost due to congestion, the first little congestion experienced will likely lead to more congestion
- this happened a few times in the 1970's -- the network hardware was working fine, but almost no data would get through

- the problem was: even when TCP data was lost, TCP would keep sending at the same speed, plus retransmit the missing packets
TCP Congestion Control

- to control congestion, TCP slows down substantially (half the speed) when packets are lost.
- TCP then slowly speeds up its transmission rate when no packets are lost.
- this is controlled by a window that (unlike the flow control window above) is maintained on each sender, and never communicated: the congestion window.
- the effective window is the smaller of the flow control window and the congestion window.
TCP Congestion Control: details

- when packets are lost, the congestion window shrinks to about half its previous size
  - actually it shrinks to one packet (one Maximum Segment Size)
  - then grows exponentially to half the previous window
- every RTT when no packets are lost, the congestion window grows by one packet (one MSS)
- since each TCP can send one window every RTT, shrinking the window slows down sending
- TCP also has other mechanisms to lessen congestion:
  - binary exponential backoff on retransmissions
  - adaptive timers to more reliably detect packet loss
Ethernet Equipment

- much experience so far in lab with Ethernet
- different equipment used to connect Ethernet segments:
  - hubs: broadcast everything
  - switches: broadcast packets addressed to 0xff:ff:ff:ff:ff:ff, and packets for destinations that are not known. Otherwise, transmit selectively if the port of the destination is known.
  - routers: forward packets among different IP networks
- hubs and switches work within a single IP network as a single broadcast medium (but switches don't always broadcast)
- traditional bridges had two interfaces, and forwarded everything from one interface to the other
  hubs and switches both implement this bridging function
Learning Switches

- if a switch gets a packet from A on interface I, it forwards the packet,
- and remembers that A can be reached on interface I
- the next time a packet for A is received on interface I', it is only forwarded on interface I (unless I == I', and then it is not forwarded)
- if there is no record of communication from A (within the last 60 seconds), the packet is broadcast on all interfaces except I'
Broadcast Storms

- given a network with redundant links
- if the network is connected by hubs, every packet will cause collisions with itself
- if the network is connected by switches, any broadcast packet will live forever
- packets may even be multiplied if there is more than one loop
- this is useless traffic that gets in the way of useful traffic -- a "broadcast storm"
Preventing Broadcast Storms

1. have no redundant links in the network, or

2. restrict "broadcast" forwarding by switches:
   - select a root switch, based on priority, using MAC addresses to break ties in case of equal priority
   - find a least-cost path to the root, reached via the root port for this switch
   - for each segment, determine a least-cost switch port to use to reach the root, the designated port for this segment
   - only forward broadcasts along root ports and designated ports
   - root ports and designated ports form a Spanning Tree
Rapid Spanning Tree Protocol (RSTP)

- The regular spanning tree protocol can take tens of seconds to converge after a topology change.
- Instead, a switch can pre-select alternate ports that also lead to the root bridge.
- Broadcast data is only sent on alternate ports when it is determined that the root port is disconnected.
- Similarly for backup paths to individual segments.
- Switches also actively exchange their information, so one switch can quickly hand off forwarding to another switch.