What Are Our Goals When Allocating Resources

- End-to-end bandwidth
- Fairness
- Link latency
- Stability (~predictability; needed for QoS. ~Robustness, able to predict network behavior)
- Minimize packet loss
- Security

Factors that we care about when meeting the goals

- Link bandwidth
- Buffer

Problem with Blasting Link with Large Congestion Window Size

- Router buffer < window size.
- Link bandwidth < access link

How to Find Out About Bottleneck Bandwidth and Buffer Size

- Slow start and ramp up.
- Packet-pair (Send 2 packets over constrained link. Time gap between them is an indication of bottleneck bandwidth. But this is a noisy signal because other packets can come between them.)
- High-level control (a la traffic condition on the news). Problem is information are not shared among Autonomous Systems (and not even among hosts in ASes) to make this possible.

TCP

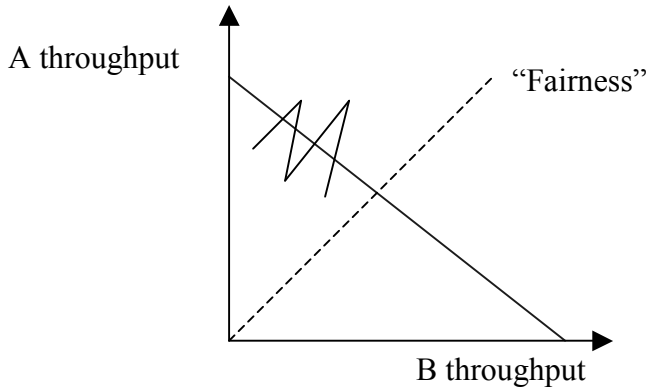
- Slow start – Additive Increase, Multiplicative Decrease
- Need to slow start because need to generate packets (and hence acks) for ack-clocking.
- Problem: Window size get very much larger than link-capacity+buffer size before packets are dropped. (When end-host sends at bottleneck bandwidth, nothing is queued. When it sends at a faster rate, queuing occurs. This causes the window size to become much larger than link bandwidth.)

(aside: ack pacing can use in place of ack-clocking. With RTT and window size, we know how to space packets when sending them.)

Congestion Avoidance

- After packet loss, half window size, exponentially increase window size, then additive increase beyond that.

Intuition of Fairness



- As window size changes, it tends towards "fairness" line.

(aside: the amount of bandwidth that a flow grabs is inversely proportional to its RTT.)

Fast Retransmit

- Use duplicate acks (3 of them) to assume that a packet is lost and retransmit it.

(aside: selective acks – bit mask of all packets received. Acks packets received out of order, thus allowing sender to deduce which packet may have been dropped and retransmit.)

Fast Recovery

- Use duplicate acks to clock out new data (instead of letting the link drain and then slow-start.)