CSE/EE 461

TCP and network congestion

Last Time …

- More on the Transport Layer
- Focus
  - How do we manage connections?
- Topics
  - Three-Way Handshake
  - Close and TIME_WAIT

Application
Presentation
Session
Transport
Network
Data Link
Physical

2

This Lecture

- Focus
  - How should senders pace themselves to avoid stressing the network?
- Topics
  - congestion collapse
  - congestion control

Application
Presentation
Session
Transport
Network
Data Link
Physical

3

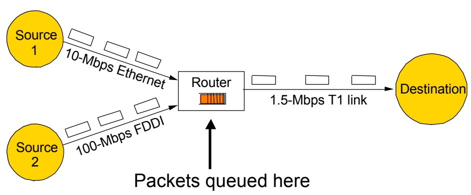
Deciding When to Retransmit

- How do you know when a packet has been lost?

```
again:
    Send(p);
    Wait(t);
    if (!p.acked)
        goto again;
```

- How long should the timer **t** be?
  - Too big: inefficient (large delays, poor use of bandwidth)
  - Too small: may retransmit unnecessarily (causing extra traffic)
  - A good retransmission timer is important for good performance
- Right timer is based on the round trip time (RTT)
  - Which varies greatly in the wide area (path length and queuing)

Congestion from in the network

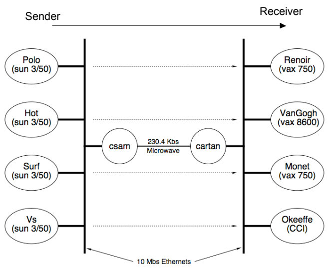


- Buffers at routers used to absorb bursts when input rate > output
- Loss (drops) occur when sending rate is persistently > drain rate

Congestion Collapse

- In the limit, early retransmissions lead to congestion collapse
  - e.g., 1000x drop in effective bandwidth of network
  - sending more packets into the network when it is overloaded exacerbates the problem of congestion (overflow router queues)
  - network stays busy but very little useful work is being done
- This happened in real life ~1987
  - Led to Van Jacobson's TCP algorithms
    - these form the basis of congestion control in the Internet today
    - [See "Congestion Avoidance and Control", SIGCOMM'88]
  - Researchers asked two questions:
    - Was TCP misbehaving?
    - Could TCP be "trained" to work better under 'absymal network conditions?'

A Scenario



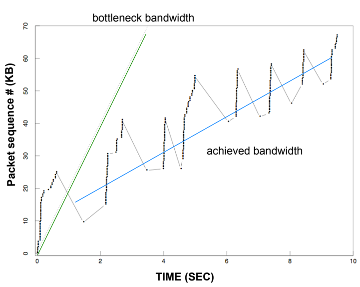
Receiver window size is 16KB.

Bottleneck router buffer size is 15 KB.

Data bandwidth is about 20KB/s

7

Effects of early retransmission



Slope is bandwidth.

Steep smooth upward slope == means good bandwidth.

Downward slope means retransmissions (*bad)*.

8

If only…

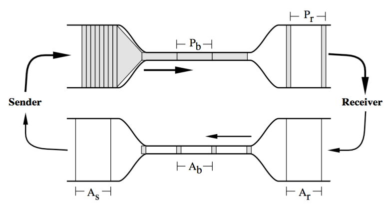
- We knew RTT and Current Router Queue Size,
  - then we would send:

    MIN(Router Queue Size, Effective Window Size)
  - and not retransmit a packet until it had been sent RTT ago.
- But we don't know these things
  - so we have to estimate them
- They change over time because of other data sources
  - so we have to continually adapt them

9

Ideal packet flow: stable equilibrium

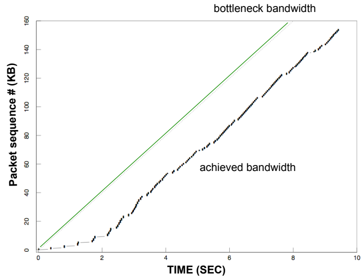
Pr = Interpacket spacing --> mirrors that of slowest link



As = Inter-ACK spacing --> mirrors that of slowest downstream link

10

Modern TCP in previous scenario



Notice:

• no retransmissions, (and thus no packet loss)

• achieved BW = bottleneck BW

11

1988 Observations on Congestion Collapse

- Implementation, not the protocol, leads to collapse
  - choices about when to retransmit, when to "back off" because of losses
- "Obvious" ways of doing things lead to non-obvious and undesirable results
  - "send effective-window-size # packets, wait rtt, try again"
- Remedial algorithms achieve network stability by forcing the transport connection to obey a 'packet conservation' principle.
  - for connection in equilibrium (stable with full window in transit), packet flow is conservative
    - a new packet not put in network until an old packet leaves

12

Resulting TCP/IP Improvements

- *Slow-start*
- *Round-trip time variance estimation*
- *Exponential retransmit timer backoff*
- *More aggressive receiver ack policy*
- *Dynamic window sizing on congestion*
- Clamped retransmit backoff (Karn)
- Fast Retransmit

Packet Conservation Principle

Congestion control means: "Finding places that violate the conservation of packets principle and then fixing them."

13

Key ideas

- Routers queue packets
  - if queue overflows, packet loss occurs
  - happens when network is "congested"
- Retransmissions deal with loss
  - need to retransmit sensibly
    - too early: needless retransmission
    - too late: lost bandwidth
- Senders must control their transmission pace
  - flow control: send no more than receiver can handle
  - congestion control: send no more than network can handle

14