6.033 Spring 2019
Lecture #11

• Reliable Transport
• Window-based Congestion Control
How do we **route** (and address) scalably, while dealing with issues of policy and economy?  

How do we **transport** data scalably, while dealing with varying application demands?  

How do we **adapt** new applications and technologies to an inflexible architecture?
Reliable Transport

- Sending Application
- Reliable Sender

Unreliable network

- Receiving Application
- Reliable Receiver

Each byte of data is delivered exactly once and in-order.
W = 5

timeout

sender

receiver
notice that (in this example) the timeout expired before the sender got an ACK indicating that 7 had been received
question: what is the correct value for $W$?

- too small $\rightarrow$ underutilized network
- too large $\rightarrow$ congestion
question: how can a single reliable sender, using a sliding-window protocol, set its window size to maximize utilization — but prevent congestion and unfairness — given that there are many other end points using the network, all with different, changing demands?
The network is fully utilized when the bottleneck link is “full.”
$R_1$ (S₁’s sending rate) + $R_2$ (S₂’s sending rate) = B

- **efficiency** (utilization)
- the network is fully utilized when the bottleneck link is “full”
efficiency (utilization)
the network is fully utilized when the bottleneck link is “full”

fairness
the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 = R_2$

$R_1 + R_2 = B$
R_1 \quad (S_1's \ sending \ rate) \\

S_1 \ is \ sending \ more \ than \ S_2 \\

S_2 \ is \ sending \ more \ than \ S_1 \\

R_2 \quad (S_2's \ sending \ rate) \\

efficiency \\
\text{(utilization)} \\
\text{the \ network \ is \ fully \ utilized \ when \ the \ bottleneck \ link \ is \ “full”}

fairness \\
\text{the \ network \ is \ fair \ when \ } S_1 \text{ and } S_2 \text{ are \ sending \ at \ the \ same \ rate}
efficiency (utilization)

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 + R_2 = B$

$R_1 = R_2$
The network is fully utilized when the bottleneck link is “full.”

The network is fair when $S_1$ and $S_2$ are sending at the same rate.

Efficiency (utilization)

$R_1 + R_2 = B$

Fairness

$R_1 = R_2$
The network is fully utilized when the bottleneck link is “full”.

The network is fair when $S_1$ and $S_2$ are sending at the same rate.

**Efficiency** (utilization)

$$R_1 + R_2 = B$$

**Fairness**

$$R_1 = R_2$$
efficiency
(utilization)
the network is fully utilized when the bottleneck link is “full”

fairness
the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 + R_2 = B$

$R_1 = R_2$

$B$

$R_1$ (S₁’s sending rate)

$R_2$ (S₂’s sending rate)
efficiency (utilization)

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when $S_1$ and $S_2$ are sending at the same rate

\[ R_1 + R_2 = B \]

\[ R_1 = R_2 \]
efficiency (utilization)

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 + R_2 = B$

$R_1 = R_2$

$B$

$R_1$ (S₁’s sending rate)

$R_2$ (S₂’s sending rate)
efficiency (utilization)

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when $S_1$ and $S_2$ are sending at the same rate

$$R_1 = R_2$$

$$(S_1's$ sending rate)$$

$$(S_2's$ sending rate)$$
efficiency

( utilization )

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 = R_2$

$R_1 + R_2 = B$
efficiency
(utility)
the network is fully utilized when the bottleneck link is “full”

fairness
the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 + R_2 = B$

$R_1 = R_2$

$R_1$ (S\(_1\)'s sending rate)

$R_2$ (S\(_2\)'s sending rate)

B

B
efficiency
(utility)
the network is fully utilized when the bottleneck link is “full”

fairness
the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 + R_2 = B$

$R_1 = R_2$

$R_1$ (S_1’s sending rate)

$R_2$ (S_2’s sending rate)

B

B

lacurts@mit.edu
efficiency
(utility)
the network is fully utilized when the bottleneck link is “full”

fairness
the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 + R_2 = B$

$(S_1$’s sending rate) $R_1$

$(S_2$’s sending rate) $R_2$

$B$
The network is fully utilized when the bottleneck link is “full”.

The network is fair when $S_1$ and $S_2$ are sending at the same rate, i.e., $R_1 = R_2$.
efficiency (utilization)

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 = R_2$

$R_1 + R_2 = B$
efficiency (utilization)

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 + R_2 = B$

$R_1 = R_2$
efficiency (utilization)

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 + R_2 = B$

$R_1 = R_2$
efficiency (utilization)

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 = R_2$

$R_1 + R_2 = B$
efficiency (utilization)

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 + R_2 = B$
efficiency
(utilization)
the network is fully utilized when the bottleneck link is “full”

fairness
the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 + R_2 = B$
efficiency (utilization)

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when \( S_1 \) and \( S_2 \) are sending at the same rate

\[ R_1 = R_2 \]

\[ R_1 + R_2 = B \]
efficiency (utilization)

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when $S_1$ and $S_2$ are sending at the same rate

$R_1 + R_2 = B$
efficiency (utilization)

the network is fully utilized when the bottleneck link is “full”

fairness

the network is fair when $S_1$ and $S_2$ are sending at the same rate $R_1 = R_2$

eventually, $R_1$ and $R_2$ will come to oscillate around the fixed point
something has happened to packet 7
in practice, if a single packet is lost, the three “dup” ACKs will be received before the RTO for that packet expires
AIMD + Slow Start

Window Size vs. Time (RTTs)

retransmission due to timeout

0 5 10 15 20 25
Time (RTTs)

0 2 4 6 8 10 12 14 16
Window Size
• **TCP** provides **reliable transport** along with **congestion control**: senders increase their window additively until they experience loss, and then back off multiplicatively. Senders also use slow-start and fast-retransmit/fast-recovery to quickly increase the window and recover from loss.

• TCP has been a massive success, but **senders don’t react to congestion until queues are already full**. Is there a better way?