George Verghese
verghese@mit.edu

6.02
11/23/10

Recs. 18 & 19

Round-trip time (RTT) is a key parameter in transport protocols

RTT = $\underbrace{\text{Transmission time}}$

to work through the length of the data packet and the ACK packet = packet length (in bits), divided by the transmission rate (in bits/sec)

\+ $\underbrace{\text{Propagation time}}$

characteristic of the channel, and can be comparable with other components of RTT
= length of link (metres), divided by speed of propagation (metres/sec)

\+ $\underbrace{\text{Processing time}}$

computation at switches to run CRCs on headers, read headers, etc.

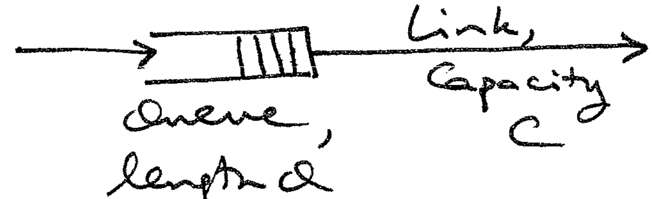
\+ $\underbrace{\text{Queueing delay}}$

associated with packets waiting in queue to be processed and sent on
= { Queue length (bits), divided by processing rate (bits/sec) } Little's law!

averages

e.g. Path with one bottleneck link

modeled as



(average)

By Little's law, delay in queue when link is transmitting at capacity is given by $\frac{Q}{C}$. So the delay across the link varies with queue length; when $Q = 0$, the queue contributes no delay because packets are processed right away. Also, transmission rate $< C \Rightarrow Q=0$.

$\therefore$ RTT on path (assuming just this one bottleneck link, i.e., no other links with capacity $\leq C$ & associated queues) is

$$RTT = \underbrace{RTT_{min}}_{\substack{\text{i.e., RTT} \\ \text{when } Q=0, \\ \text{which happens} \\ \text{when rate of} \\ \text{transmission is } < C}} + \underbrace{\frac{Q}{C}}_{\substack{\text{Queuing delay} \\ \text{associated with} \\ \text{transmission at} \\ \text{rate } C}}$$

Sending data to the input of the queue at rate $> C$ will cause queue to fill up and bits/packets to be lost, so input rate needs to be $\leq C$.

How does all this balance out when the sender uses a sliding-window protocol? $\Longrightarrow$

Assuming no packets are lost, the sender's data rate is $\frac{W}{RTT}$ ← max. # of unacknowledged packets allowed in protocol.

So we need

$$\frac{W}{RTT_{min} + \frac{Q}{C}} \leq C$$

or $W \leq C \cdot RTT_{min} + Q$.

When $Q = 0$, the round-trip time is RTT_{min}, and the rate is $\frac{W}{RTT_{min}} \leq C$.

When W crosses above $C \cdot RTT_{min}$, the queue starts to be occupied. If Q_{max} is the maximum tolerable (average) queue length, then the largest tolerable window size is

(because we are using Little's law)

$$W_{max} \leq C \cdot RTT_{min} + Q_{max}$$

(to avoid bits being dropped from the queue)

Above this, awkward/bad things start to happen!

Try Problems 9 & 10 in Chapter 20 of lecture notes.

Filtering measured RTT to get a smoothed estimate sRTT, using an

EWMA → <u>exponentially-weighted</u> <u>moving-average</u> filter:

$$s[n+1] = (1-\alpha)\, s[n] + \alpha\, r[n] \qquad (*)$$

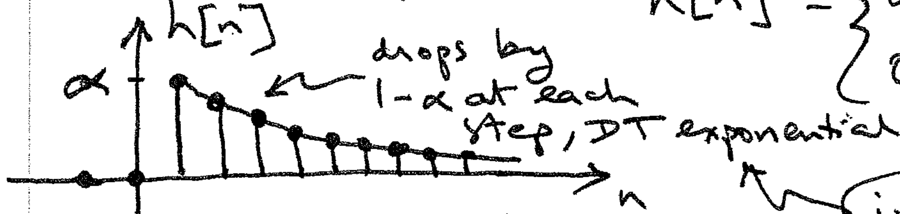
$0 < \alpha < 1$

$s[n]$: smoothed estimate sRTT at time n

$r[n]$: measured RTT at time n

This is a causal LTI system, with unit sample response

$$h[n] = \begin{cases} \alpha(1-\alpha)^{n-1} & \text{for } n \geq 1 \\ 0 & \text{for } n < 1 \end{cases}$$



i.e., geometric series.

So $$s[n] = \alpha \sum_{-\infty}^{n-1} r[m]\,(1-\alpha)^{n-m-1}$$

$$= \alpha \left\{ r[n-1] + (1-\alpha)\, r[n-2] + (1-\alpha)^2 r[n-3] + \ldots \right\}$$

i.e., EWMA.

System is stable. So if $r[n]$ goes to a value R and stays there, $s[n]$ asymptotically (exponentially) approaches a steady-state value S. What is S? Determine by substitution in Eq. (*):

As $(1-\alpha)^n$

$$S = (1-\alpha)S + \alpha R \implies S = R,$$

as desired!

Getting a feel for $(1-\alpha)^n$:
$(1-\alpha)$ for $\alpha = \frac{1}{8}$ is $\frac{7}{8}$.
$\left(\frac{7}{8}\right)^n < 0.5$ for $n \geq 6$, not before
$\left(\frac{7}{8}\right)^n < 0.05$ for $n \geq 23$, not before

This explains the choice of the coefficient α in front of $r[n]$ in Eq. (*)