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

\[ RTT = \text{Transmission time} \]

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).

+ Propagation time

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

+ Processing time

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

+ Queuing delay

Associated with packets waiting in queue to be processed and sent on = \( \frac{\text{Queue length (bits)}}{\text{by processing rate (bits/sec)}} \) Little's law!
e.g. link with one bottleneck link modeled as

\( \text{Capacity} \quad \text{Queue} \quad \text{Delay} \quad \text{Length} \quad (\text{average}) \)

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 < \( C \) \( \Rightarrow Q = 0 \).

\[ \text{RTT on path (assuming just this one bottleneck link, i.e., no other links with capacity} \leq C \text{ & associated queues)} \] is

\[ \text{RTT} = \text{RTT}_{\text{min}} + \frac{Q}{C}, \]

i.e., RTT when \( Q = 0 \), which happens when rate \( g \) transmission is < \( C \)

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?
Assuming no packets are lost, the sender's data rate is \[
\frac{W}{\text{RTT}} = \text{max. # of unacknowledged packets allowed in protocol.}
\]

So we need
\[
\frac{W}{\text{RTT}_\text{min} + \frac{Q}{C}} \leq C
\]

or \[
W \leq C \cdot \text{RTT}_\text{min} + Q.
\]

When \( Q = 0 \), the round-trip time is \( \text{RTT}_\text{min} \), and the rate is \[
\frac{W}{\text{RTT}_\text{min}} \leq C.
\]

When \( W \) crosses above \( C \cdot \text{RTT}_\text{min} \), the queue starts to be occupied. If \( Q_{\text{max}} \) is the maximum tolerable (average) queue length, then the largest tolerable window size is
\[
W_{\text{max}} \leq C \cdot \text{RTT}_\text{min} + Q_{\text{max}}
\]

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 exponentially-weighted moving-average filter:

\[ s[n+1] = (1-\alpha) s[n] + \alpha r[n] \tag{\ast} \]

\[ 0 < \alpha < 1 \]

Smoothed estimate at time \( n \)

This is a causal LTI system, with unit sample response

\[ h[n] = \sum_{m=0}^{n-1} \alpha (1-\alpha)^m \text{ for } n \geq 1 \]

\[ 0 \text{ for } n < 1 \]

So

\[ s[n] = \alpha \sum_{m=0}^{n-1} r[m] (1-\alpha)^m \]

\[ = \alpha \sum_{m=0}^{n-1} r[n-m-1] + (1-\alpha) r[n-2] + (1-\alpha)^2 r[n-3] + \ldots \]

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. (\ast):

\[ S = (1-\alpha) S + \alpha R \Rightarrow S = R, \]

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 \( n \geq 6 \), before.

\( \left(\frac{7}{8}\right)^n < 0.05 \) for \( n \geq 23 \), not before.

This explains the choice of the coefficient \( \alpha \) in front of \( r[n] \) in Eq. (\ast).