Congestion Control: Beyond TCP

Yuhang Ren (yuhangr2)
Department of Electrical and Computer Engineering

Dec 17th, 2021

Main topics today

- Motivation: improve TCP
- Explicit Congestion Notification (ECN)
    - DEC bit
    - Source Quench
    - Sally Floyd’s modification to ECN
- BBR (Bottleneck Bandwidth and Round-trip propagation time)

Motivation

- Review of TCP:
  - A packet drop indicates network is congested
  - Detect congestion by three DupACKs or timeouts
  - Decrease CW and SST during congestion
- Questions:
  - Can we do better?
  - Is it too slow to react to congestion by waiting for three DupACKs?
  - Is it good to avoid congestion only when packet drops occur?

Explicit Congestion Notification (ECN)

- An extension to the Internet Protocol and to the Transmission Control Protocol
- Explicitly declare congestion in packet headers
- Routers sets ECN bit to 1 in packets if queue is about to overflow
- Receiver echoes ECN by setting ECN bit to 1 in ACKs
- Sender responds to ECN

ECN: DEC bit

- Routers set ECN bit when the average queue size exceeds a certain threshold
- Receiver replies ACKs with ECN bit set to 1
- Sender calculates how many packets it received with the congestion indication bit set to one
- During the last window, calculate how many ACKs have ECN=1:
  - If less than half, linearly increase congestion window
  - Else, exponentially decrease the congestion window

ICMP Source Quench

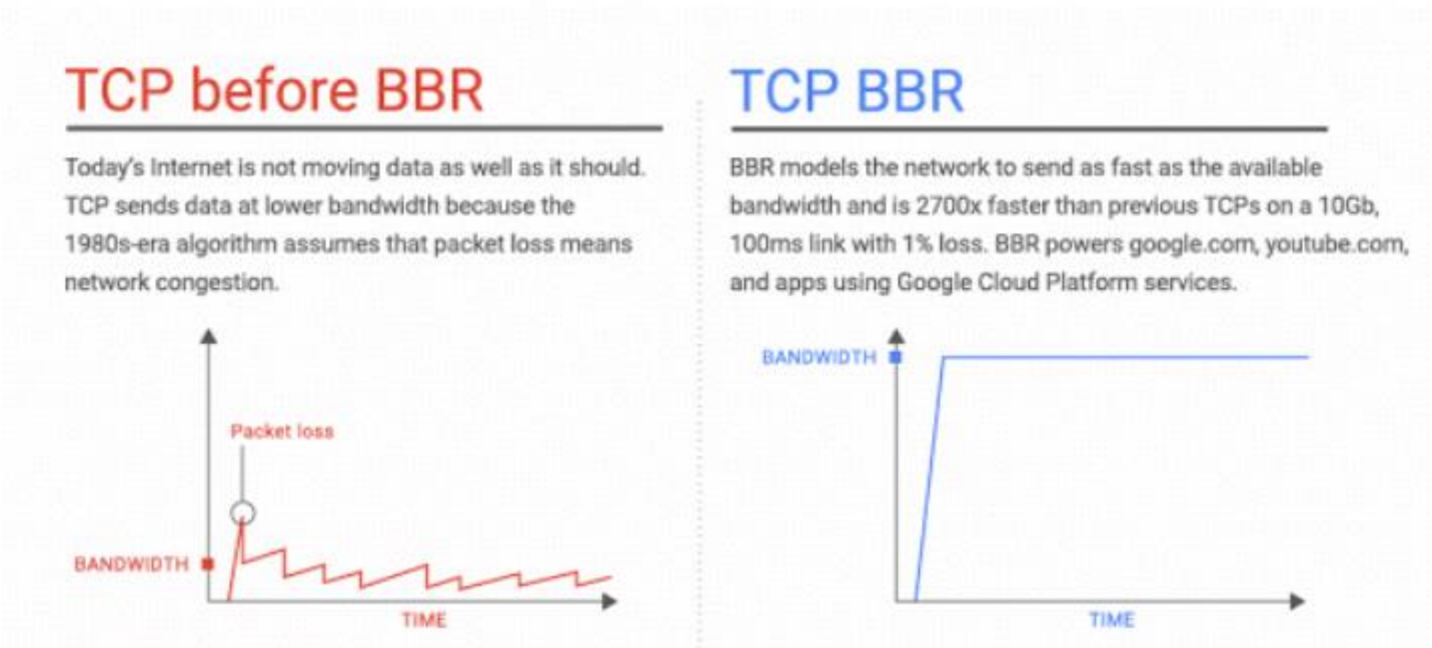
- ICMP: Internet Control Message Protocol
- Source Quench: a message in IP header that requests hosts to cut back the rate at which it is sending traffic to the internet destination
- TCP responds to a Source Quench by reducing CW to 1 and initiate Slow Start
- Problem:
  - Violates end-to-end rule
  - Consumes network bandwidth

ECN, revised by Sally Floyd

- Still uses ECN bit to indicate congestion, but TCP responds differently
- Halves both CW and SST
- A single ECN message needs response, but TCP responds at most once per RTT.
- Succeeding ECN bits are ignored
- On receiving three DupACKs immediately after an ECN is handled, do not reduce CW
- Upon timeouts, do not decrease SST if it has been decreased within the last RTT

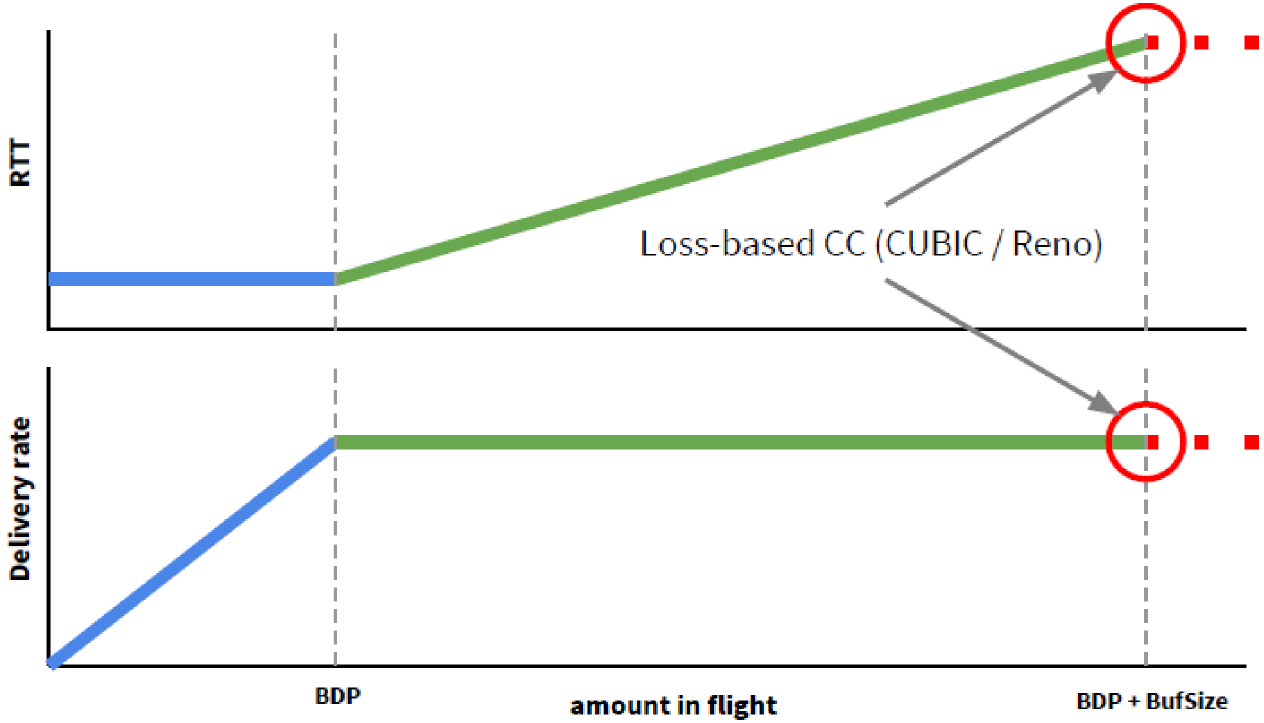
BBR (Bottleneck Bandwidth and Round-trip propagation time)

- A new congestion control algorithm developed at Google
- An upgraded version of TCP



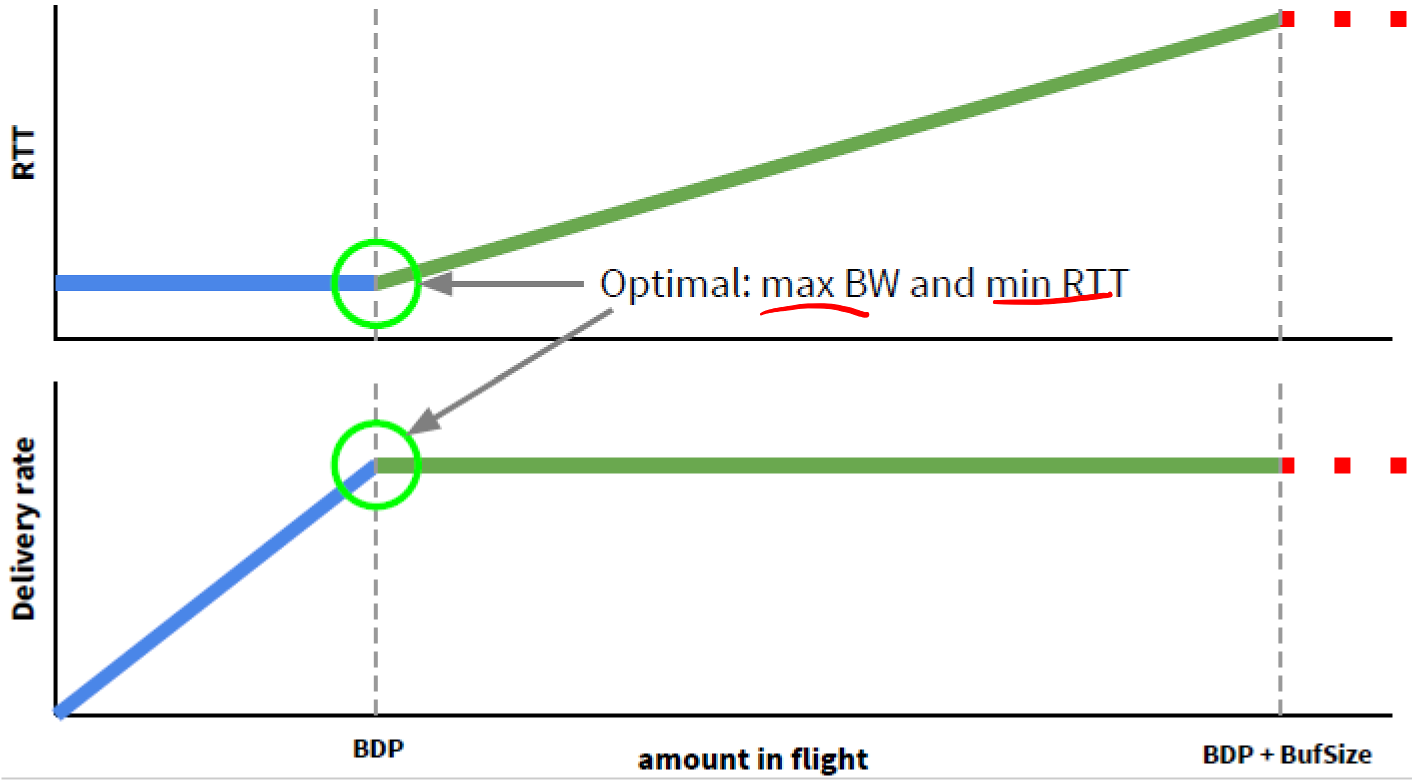
BBR: motivation

- BBR seeks high throughput with a small queue by **probing BW** and **RTT**

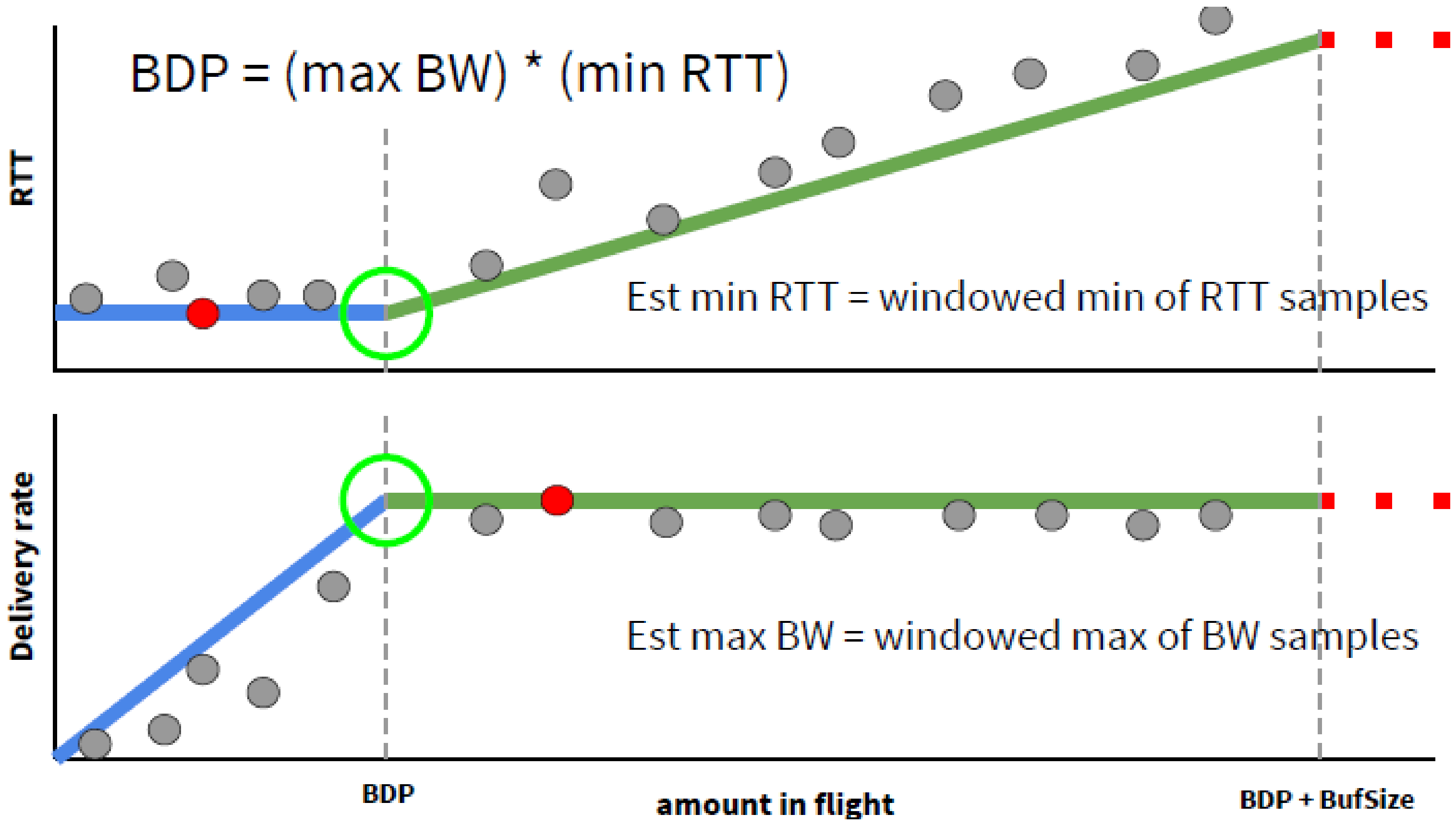


BDP: bandwidth-delay product

BBR: motivation



Probing



BDP = (max BW) * (min RTT)
RTT
Est min RTT = windowed min of RTT samples
Delivery rate
Est max BW = windowed max of BW samples
BDP
amount in flight
BDP + BufSize

Core Design

- **Model** network path
  - Update estimates of max BW and min RTT on each ACK
- **Control** sending based on the model, to…
  - Pace near estimated BW, to reduce queues and loss
  - Vary pacing rate to keep inflight near BDP
- **Result** in…
  - much larger goodput
  - Tolerance to packet losses for up to ~15%
  - Low queueing delay and latency

Summary

- Motivation: improve TCP
- Explicit Congestion Notification (ECN)
  - DEC bit
  - Source Quench
  - Sally Floyd's modification to ECN
- BBR (Bottleneck Bandwidth and Round-trip propagation time)