

Scaling TCP's Congestion Window for Small Round Trip Times

Bob Briscoe, BT*

Koen De Schepper, Bell Labs

RITE – Reducing Internet Transport Latency

IETF-93, Praha, July 2015



* now independent

Summary of 4pp tech report:
<<http://www.bobbriscoe.net/projects/latency/sub-mss-w.pdf>>

R / T E

REDUCING INTERNET TRANSPORT LATENCY

The authors' contributions are part-funded by the European Community under its Seventh Framework Programme through the Reducing Internet Transport Latency (RITE) project (ICT-317700). The views expressed here are solely those of the authors.



Problem

At low RTT, TCP overrides AQM and builds large queue

✧ Capacity seeking (TCP) → need AQM deployment

- queues \Downarrow → RTT \Downarrow
- $\text{cwnd} = \text{packet rate} * \text{RTT}$
- So, AQM → $\text{cwnd} \Downarrow$

✧ During testing of DCTCP, as number of flows, $n \nearrow$

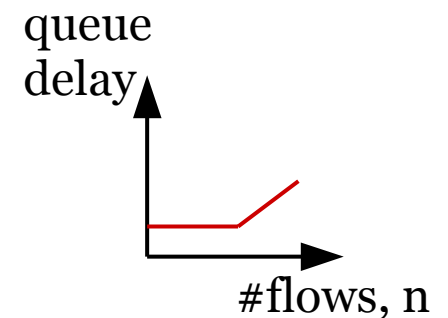
- noticed queue growth after a plateau

✧ due to min cwnd of 2 in TCP

- preserves ACK clock (without delayed ACKs, would still need min cwnd, but 1 not 2)

✧ Problem seen before, with low packet rates (developing world, congestion)

- but low RTT can also cause low cwnd



RTT : Round Trip Time

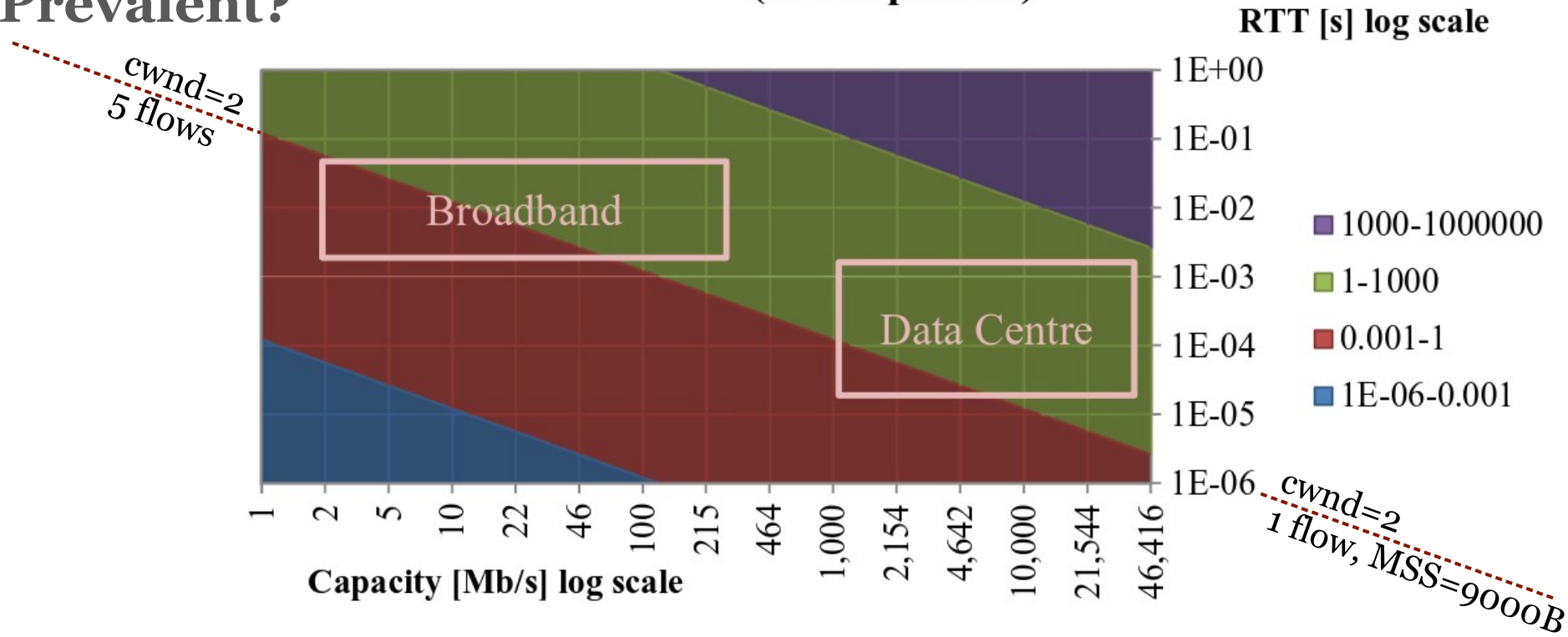
AQM: Active Queue Management

cwnd: TCP congestion window

Problem

How Prevalent?

Window [pkt/RTT]
if capacity is shared by 10 flows
(1500B packets)



- ❖ Need another box for “Developing World”
 - need to dig out data on typical #flows

- ❖ We have not noticed this before, because:
 - not much experience of short queues
 - even AQM typically not ultra-shallow
 - we have not been looking

Towards a solution

What not to do?

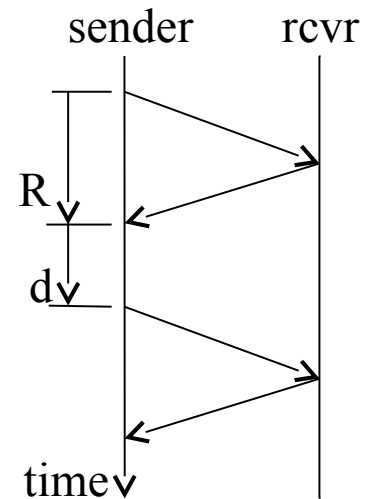
- ✧ “Use smaller packets?” **No**
 - could be harmful if packet congestion was the cause of low cwnd
- ✧ “This is a fundamental limit of TCP's ACK clock?” **No**
 - if the rate the network needs is unachievable with an ACK clock
 - get yourself a different clock
- ✧ “Invent an AQM to fix this?” **No**
 - No matter how much an AQM asks TCP to reduce cwnd
 - TCP rounds back up to 2. We have to fix TCP



Towards a solution

What to do?

- ✧ A new pacing mode will be needed, beneath ACK-clock mode
 - if $\text{cwnd} < \text{SMSS}$, send a segment every $1/\text{cwnd}$ round trips
- ✧ more precisely,
 - send a packet of s bytes every s/cwnd round-trips
 - where $s = \min(\text{SMSS}, \text{send_queue})$
- ✧ TCP Nice is an existence proof



Summary

This is a significant bug in TCP

- ✧ TCP can override AQM's efforts to reduce queue delay
 - in a significant range of scenarios
- ✧ TCP congestion control RFC5681 will need to be changed
 - IMO, SHOULD implement sub-MSS window
- ✧ This is a call for researchers to work on this
 - pacing is never easy

