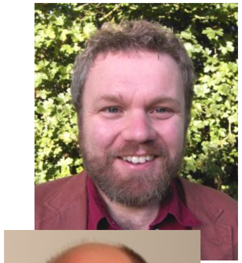
TCP Prague:

Resolving Tensions between Congestion Control Scaling Requirements

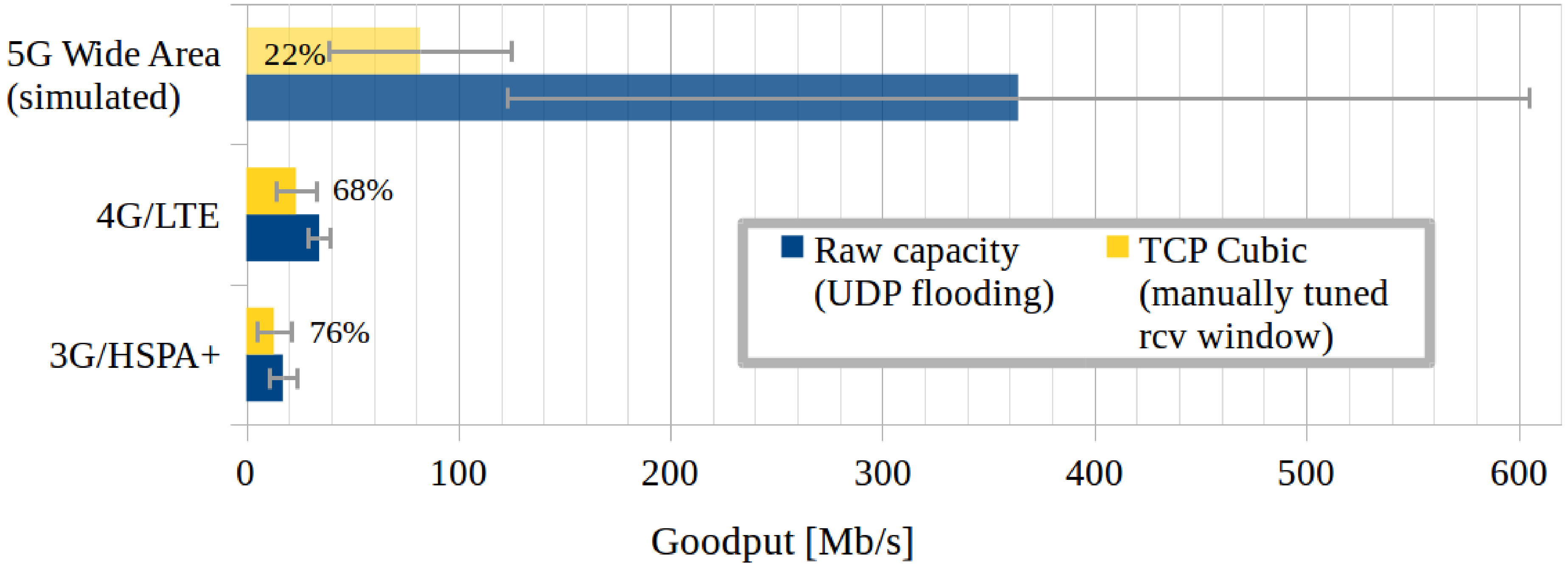
Supporting discussion paper:

https://riteproject.files.wordpress.com/2015/10/ccdi_tr.pdf



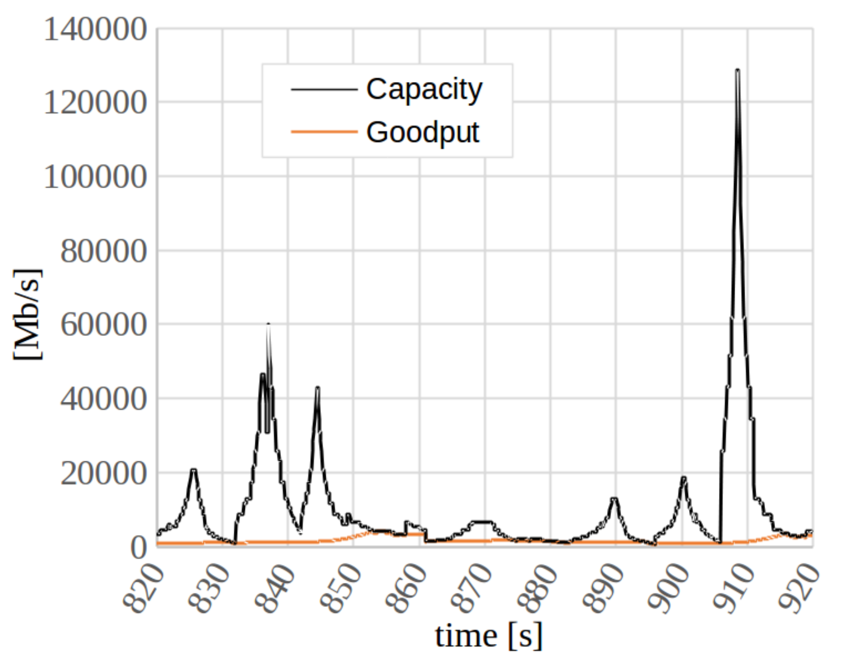
Bob Briscoe bobbriscoe.net
Koen De Schepper NOKIA Bell Labs

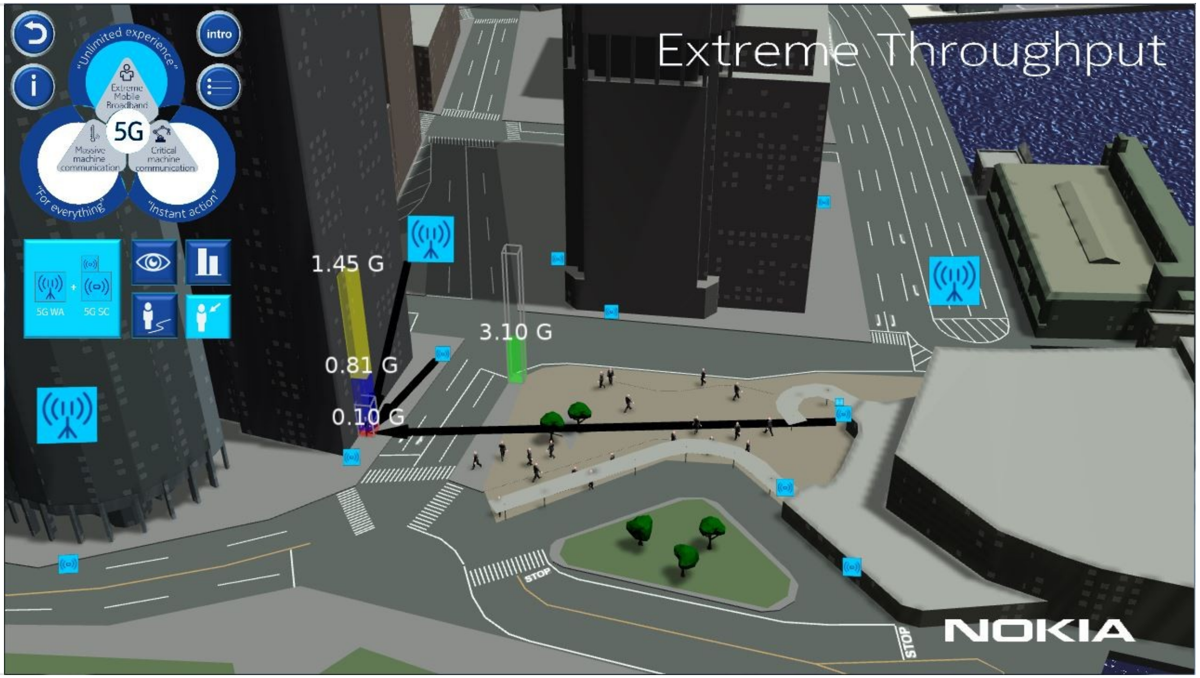
Scaling congestion control dynamics



- whiskers: 1 standard deviation.
- %-age figures: resulting utilization of raw capacity
- 3G and 4G results captured over a production network [1]
- 5G Wide Area macro-cell results were simulated (next slide)
- Poor handling of dynamics by TCP's receive window was manually overridden in all cases (to focus on congestion control dynamics)
  - with autotuned rwnd and Cubic CC, 4G/LTE mean utilization was 43%

Why Such Poor Utilization?





- Right: video still of simulation to establish capacity dynamics
  - pedestrian mobility model
  - 3.5GHz macrocell (red) combined with 28GHz (blue) and 73 GHz (yellow) mmWave microcells
- Left: Goodput is for TCP Cubic over 20ms base RTT
  - each 10s grid line = 500 RTT
  - mean goodput: 1.2Gb/s; mean utilization: 19%

WiFi is no better

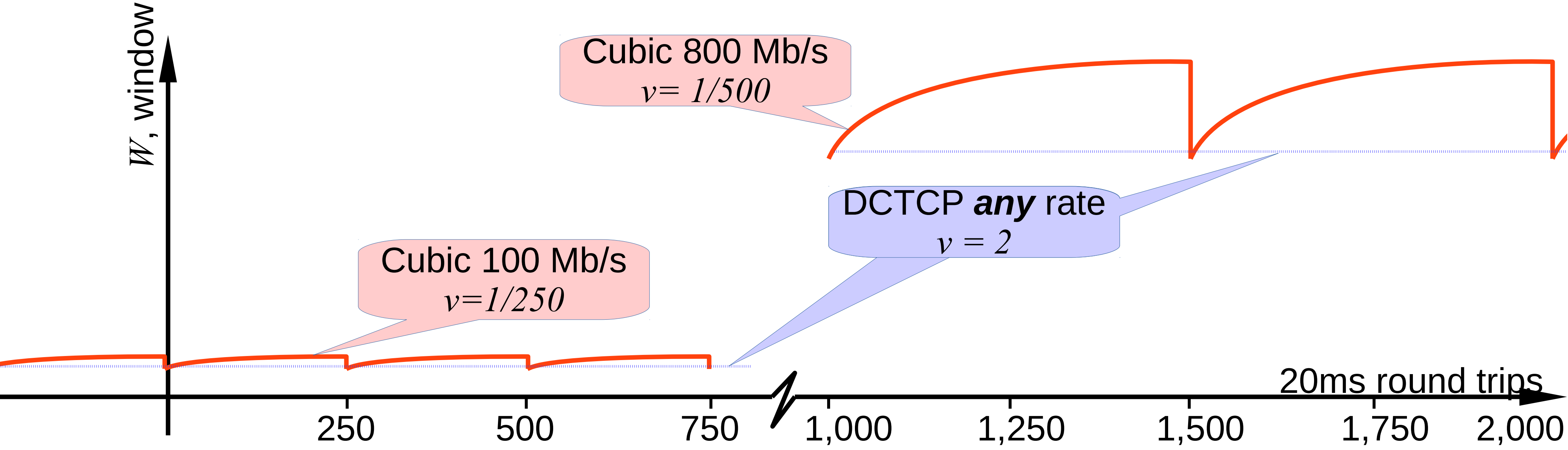
- 802.11ad
  - 60GHz 3Gb/s channel, static, office environment, occasional human blocking, Compound TCP
  - median goodput 280Mb/s, utilization 9.3%
  - near-perfect beam-forming improved utilization to 16% [2]

BDP = Bandwidth-Delay Product

Lesson from design of high-BDP TCP protocols

- keep the number of round trips between drops small
- equivalently, keep the number of drops per round trip large

v : number of congestion signals per round trip



Cubic 800 Mb/s
$v = 1/500$

DCTCP ***any*** rate
$v = 2$

Cubic 100 Mb/s
$v = 1/250$

v : number of congestion signals per round trip
W: congestion window
p: dropping or marking probability

Req#1. Scalable Congestion Signalling

- congestion signalling is scalable if $v \geq v_0$ (1)
  where v_0 is a reasonable min
- v = (segments per RTT, W) * (probability each will be marked, p)

  $v = Wp$

  substitute in scalability constraint (1)

  $W \geq v_0 / p$ (2)
- can easily derive constraint on steady-state TCP equations from this...

  General congestion control formula: $W \propto \frac{1}{p^B}$,
- To satisfy (2), $B \geq 1$

	B
Reno	½
Cubic	¾
DCTCP (prob. AQM)	1
DCTCP (step AQM)	2

Approach to scaling the dynamics

- capacity decrease or other flows arrive: not the problem
- If capacity increases, or other flows depart, no information, except
    - ACK rate increases briefly while queue empties(1)
    - the next mark never comes...

- If normally 500 RTTs between marks, it takes ~1000 RTTs to notice their absence
- If normally 2 marks per RTT, it takes 1 RTT to notice
- Then sender can start probing for capacity

(1) Goal: ultra-low standing queue;
The closer to our goal, the less we will detect this

Req#2: Limited RTT-dependence

- We have lived with this. Why change?
- Bufferbloat has cushioned us from the impact of RTT-dependent CC
- Low queuing delay leads large RTT flows to starve

E.g: base RTT ratio $R_1/R_2 = 200/2 = 100$

	Qdelay q	Total RTT imbalance $(R_1+q)/(R_2+q)$	
Drop tail	200 ms	$\frac{(200+200)}{(2+200)}$	≈ 2
PIE AQM	15 ms	$\frac{(200+15)}{(2+15)}$	≈ 13
L4S AQM	500 µs	$\frac{(200+0.5)}{(2+0.5)}$	≈ 80

Note: this is not an anti-starvation requirement not a strong 'fairness' requirement

Tension between Reqs 1 & 2

- Scalable congestion signalling $pW \geq v_0$
- Limited RTT-dependence (pW/R const) $pW \propto R$

v : number of congestion signals per round trip
W: congestion window
p: dopping or marking probability
R: Total Round trip time

“Compromise 5” betw Reqs 1 & 2

- signals per RTT

$$pW = \frac{v_0}{lg(R_0/R+1)}$$

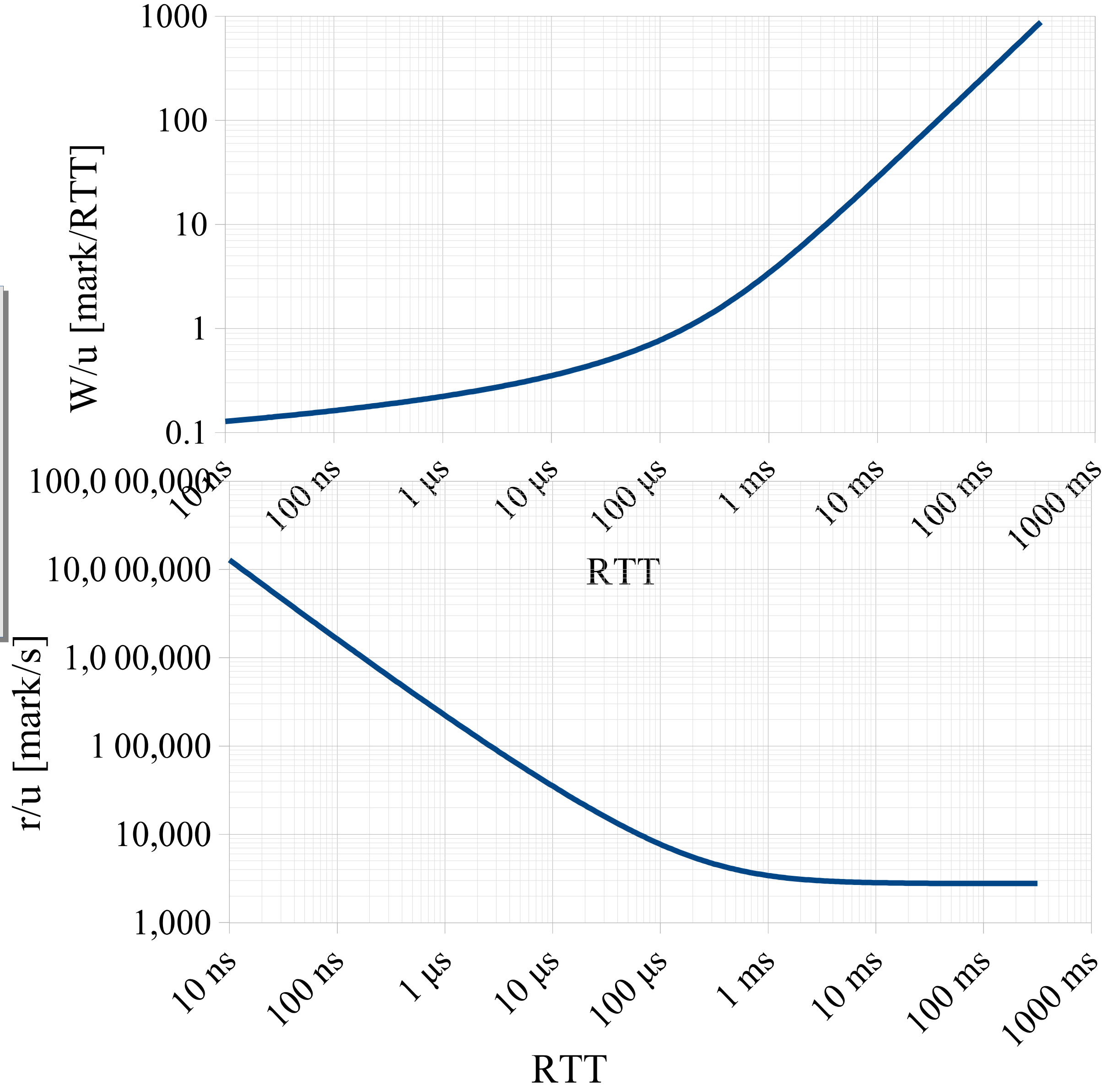
scalable signalling

AND

$>>R_0$ RTT-independent
$<<R_0$ not RTT-dependent

- flow rate

$$\frac{pW}{R} = \frac{v_0}{R\,lg(R_0/R+1)}$$



sorry for confusing you all: $p \approx 1/u$

Resolving Tensions between Congestion Control Scaling Requirements

6 scalability requirements (RTT, rate):

1. Scalable congestion signaling
2. Limited RTT-dependence
3. Unlimited responsiveness
4. Low relative queuing delay
5. Unsaturated signaling
6. Coexistence with Classic TCP

Link to paper:

[Resolving Tensions Between Congestion Control Scaling Requirements](https://riteproject.files.wordpress.com/2015/10/ccdi_tr.pdf)
https://riteproject.files.wordpress.com/2015/10/ccdi_tr.pdf

Link to experiment videos:

BBR with AQM: https://youtu.be/4eYfyKYe9nM

BBR with Cubic: https://youtu.be/akO1HN2ey48

more info

[1] K. Liu and J. Y. B. Lee, "On Improving TCP Performance over Mobile Data Networks," IEEE Transactions on Mobile Computing, 2016.

[2] Sur, S., Zhang, X., Ramanathan, P. & Chandra, R., "BeamSpy: Enabling Robust 60 GHz Links Under Blockage," In: 13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16) pp.193-206 (2016)