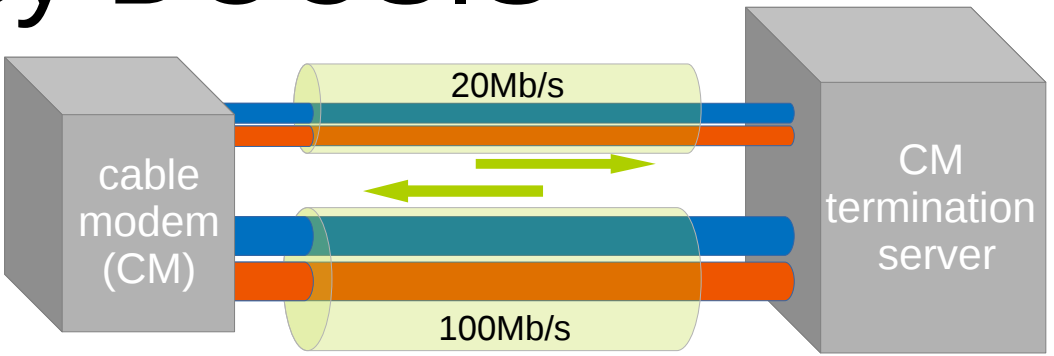
Buffer size?
We took so long to answer that the question has changed

Bob Briscoe
Independent & CableLabs
research@bobbriscoe.net
b.briscoe-contractor@cablelabs.com

It depends...

1) ...on the congestion controllers in use
   - not Reno, not even Cubic in a few years
   - BBR-like & DCTCP-like (yes, for the public Internet)
   - in particular flow-start, re-start & large adjustments
2) ...on the bottleneck buffer behaviour
   - AQM
   - ECN, in particular L4S-ECN

Low Latency DOCSIS*

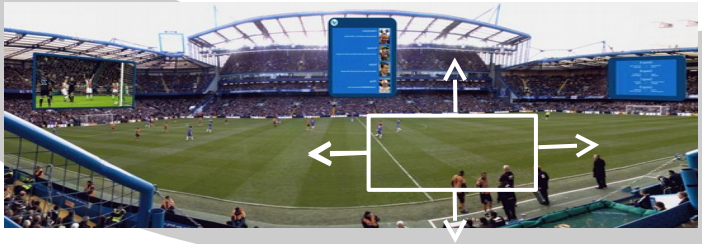


- Specs published Jan 2019
  - up: s/w upgrade to DOCSIS 3.1 CMs
  - down: for any version CM
- DualQ architecture, but...
  - ✗ not old-school QoS; not low latency through bandwidth priority
    - no bandwidth allocated to either queue – only to aggregate
    - low latency queue can fully utilize pooled capacity
  - ✔ enabler to cut end-systems loose from Reno/Cubic/BBR constraints
- Classifies by sender's behaviour
  - Non-Queue-Building: 'scalable' congestion controls + light traffic
  - QB Queue-Building: 'classic' congestion controls (Reno, Cubic, BBR)

* Data Over Cable Service Interface Specification

Ultra-low latency for every application

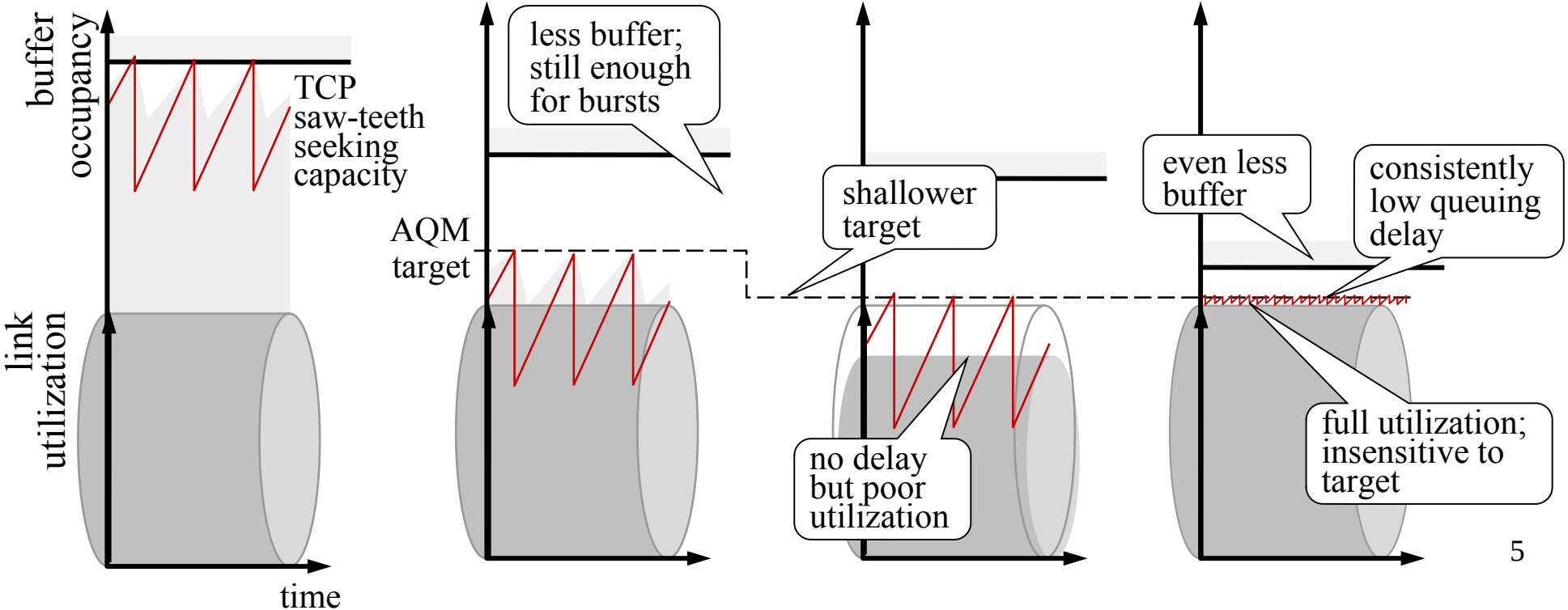
- Not only non-queue-building traffic
  - DNS, gaming, voice, SSH, ACKs, HTTP requests, etc
- Capacity-seeking traffic as well
  - TCP, QUIC, RMCAT for WebRTC
  - web, HD video conferencing, interactive video, cloud-rendered virtual reality, augmented reality, remote presence, remote control, interactive light-field experiences,...



[L4S-MMSYS]

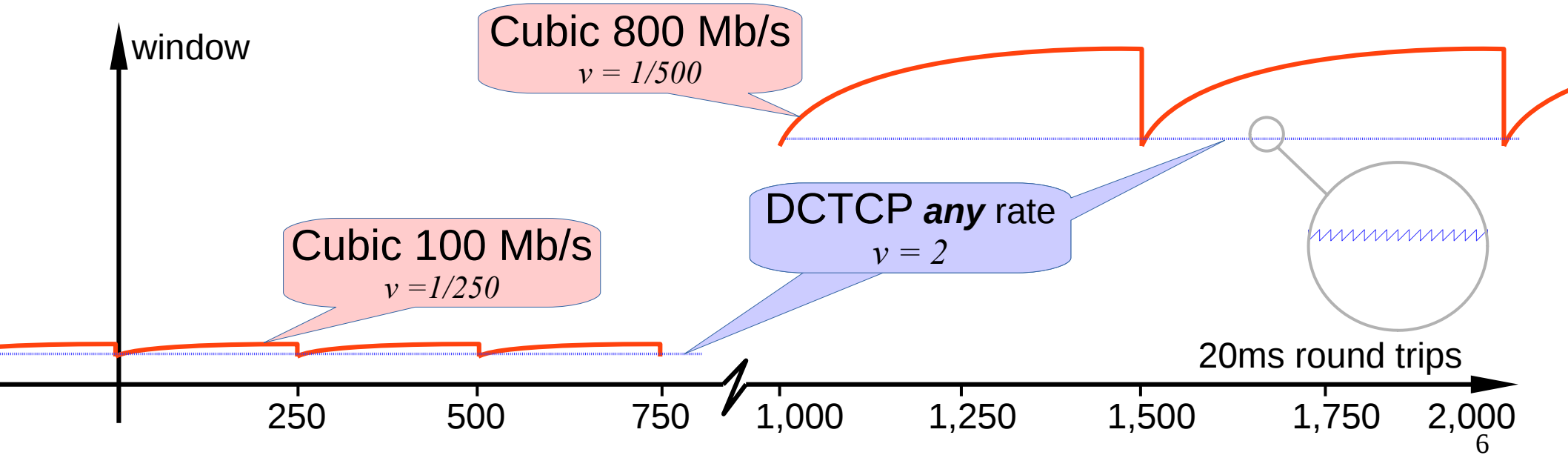
The trick: scalable congestion control

	① Today (typical)	② Today (at best)	③ Unacceptable	④ L4S
Bottleneck	Bloated drop-tail buffer	AQM	Shallower AQM	Immediate AQM
Sender CC	Classic	Classic	Classic	Scalable (tiny saw-teeth)



'Scalable' congestion control?

- example: Data Center TCP
- invariant average congestion signals per round trip (v)
- at any rate: queuing delay remains low with full utilization

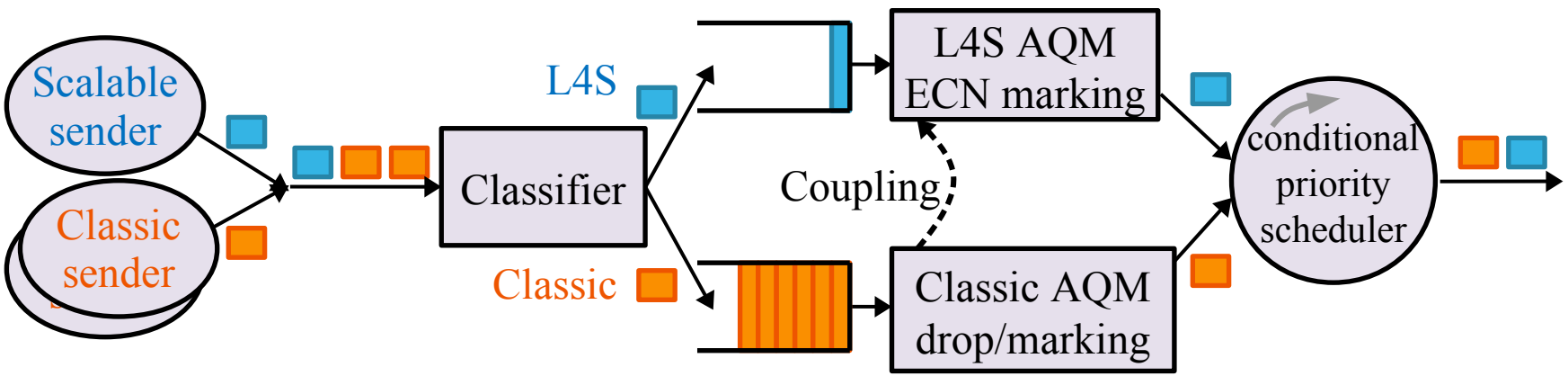


DualQ Coupled AQM

latency isolation, but bandwidth pooling

- L4S-ECN: senders set ECT(1) → classifies into L4S queue

Codepoint	IP-ECN bits	Meaning
Not-ECT	00	Not ECN-Capable Transport
ECT(0)	10	Classic ECN-Capable Transport
ECT(1)	01	L4S ECN-Capable Transport
CE	11	Congestion Experienced

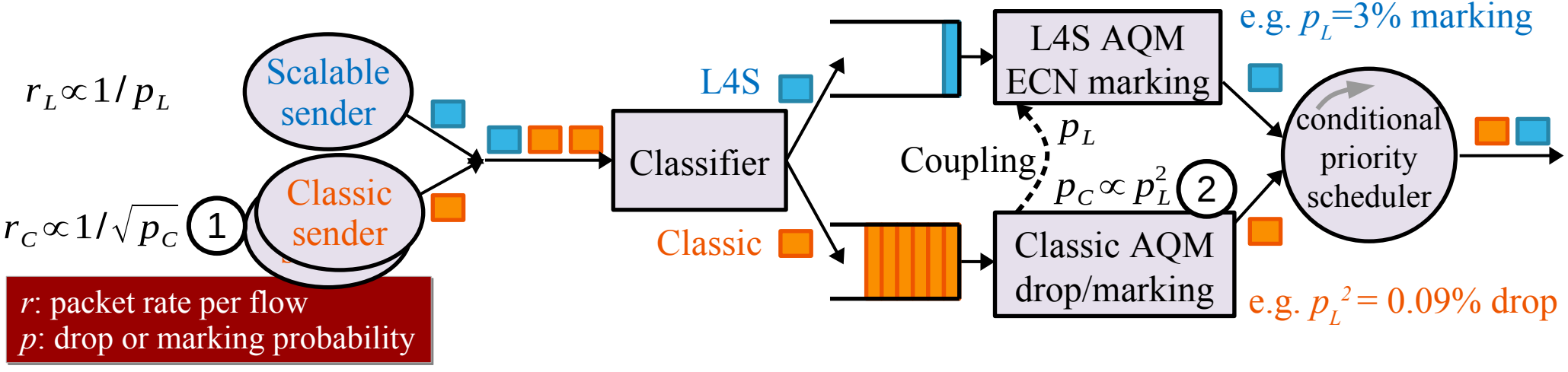


[RFC8311], [ietf-ecn-l4s-id], [dualpi2-netdev], [ietf-dualq-aqm]

DualQ Coupled AQM

latency isolation, but bandwidth pooling

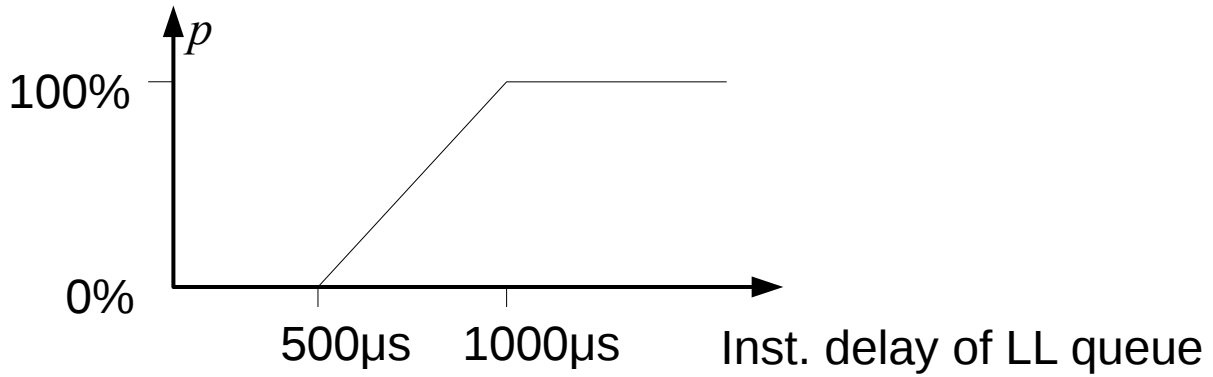
- how do n+m flows get 1/(n+m) of the combined capacity?
  - (1) classic congestion control (TCP & QUIC):
    rate depends on the square root of the drop level
  - (2) counterbalanced by the squaring



- no flow ID inspection, no bandwidth priority

[RFC8311], [ietf-ecn-l4s-id], [dualpi2-netdev], [ietf-dualq-aqm]

AQM for Low Latency Queue



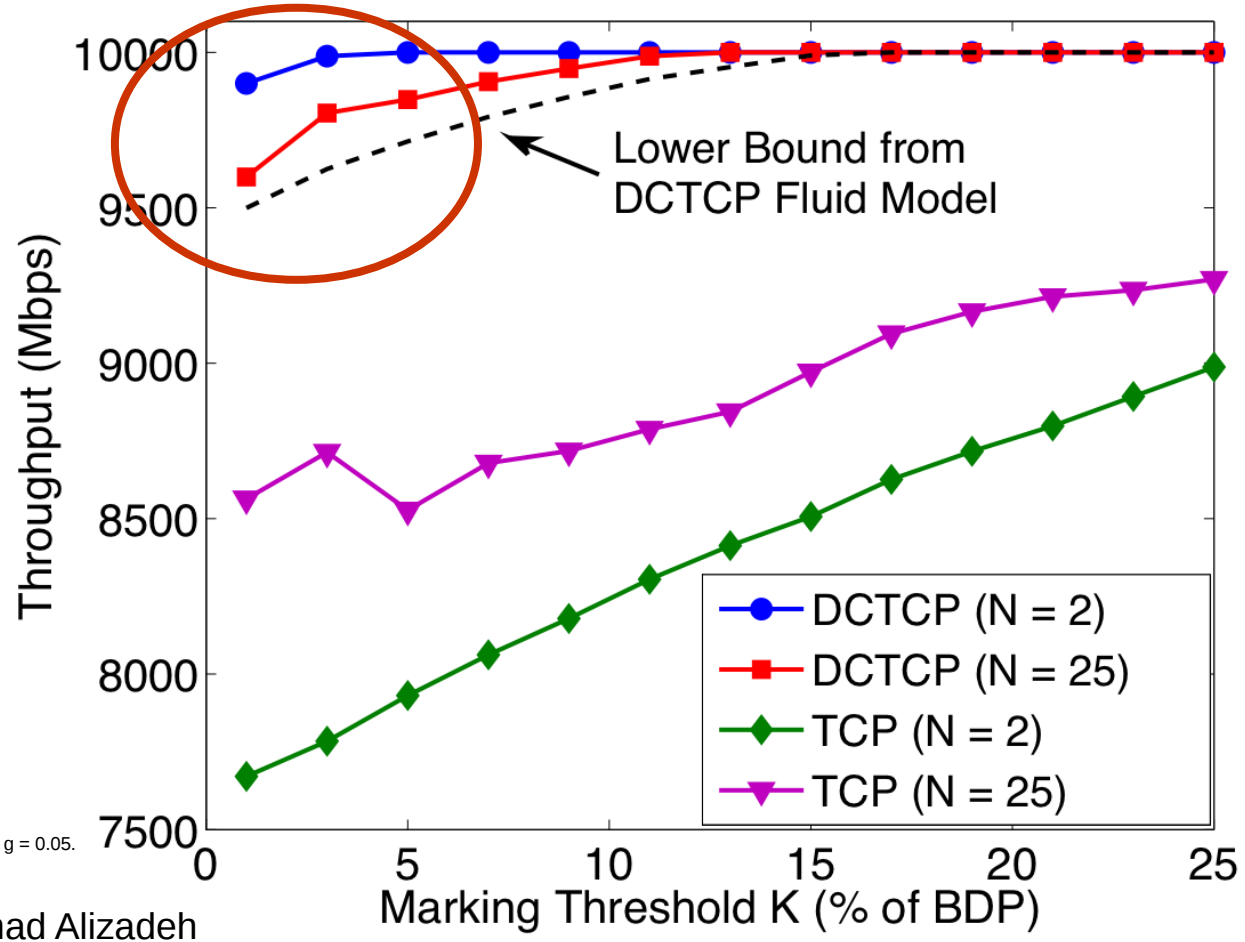
min threshold: 500µs (or 2 MTU @ max sustained rate, if greater)

buffer size: 10ms

DCTCP: insensitivity to low ECN threshold

For DCTCP:
Throughput > 94%
as *K* → 0

For TCP:
Throughput → 75%

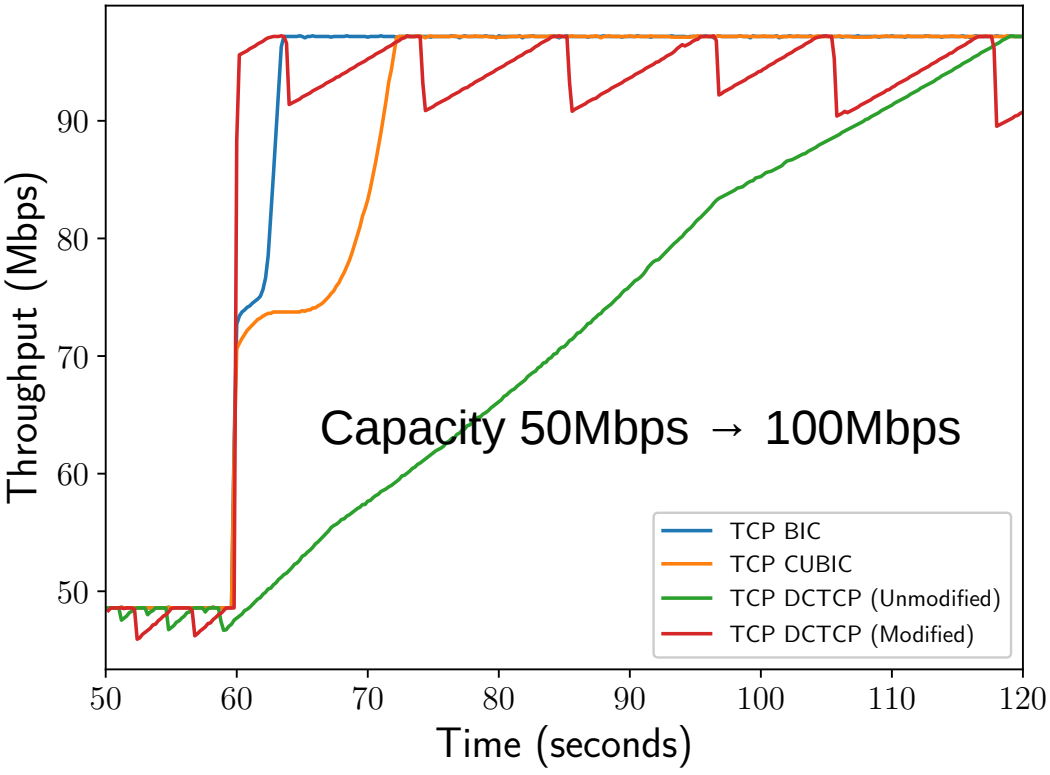


Parameters:
link capacity = 10Gbps
RTT = 480μs
smoothing constant (at source), g = 0.05.

Slide courtesy of Murari Sridharan/Mohammad Alizadeh

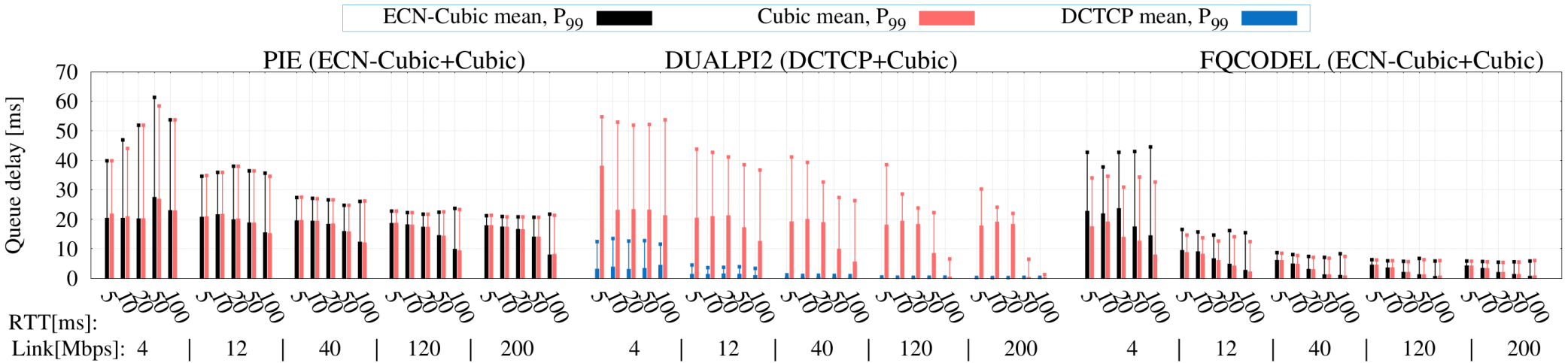
flow-start

- Modify DCTCP with paced chirping
- DCTCP's high ECN-marking freq
  → rapidly detect when it stops



- ~7 RTTs to regain capacity (RTT: 100ms)
- Qdelay overshoot
  - DCTCP+paced chirping ~1ms
  - (Cu)bic ~50ms (0.5 RTT)

Comparison with 'Classic' AQMs



traffic: heavy web workload + single longer-running flows

Take away messages

- buffer sizing in core depends on
  - bottleneck behaviour (prob. access network)
  - sender congestion control
- only get benefits of new CCs if isolate from old CCs
- flow start is the critical path for buffer sizing

- Low Latency DOCSIS modems
  - instrumented for Qdelay histogram logging
  - virtual queue-ready (use ECN marking to fly just below capacity)

Buffer size?

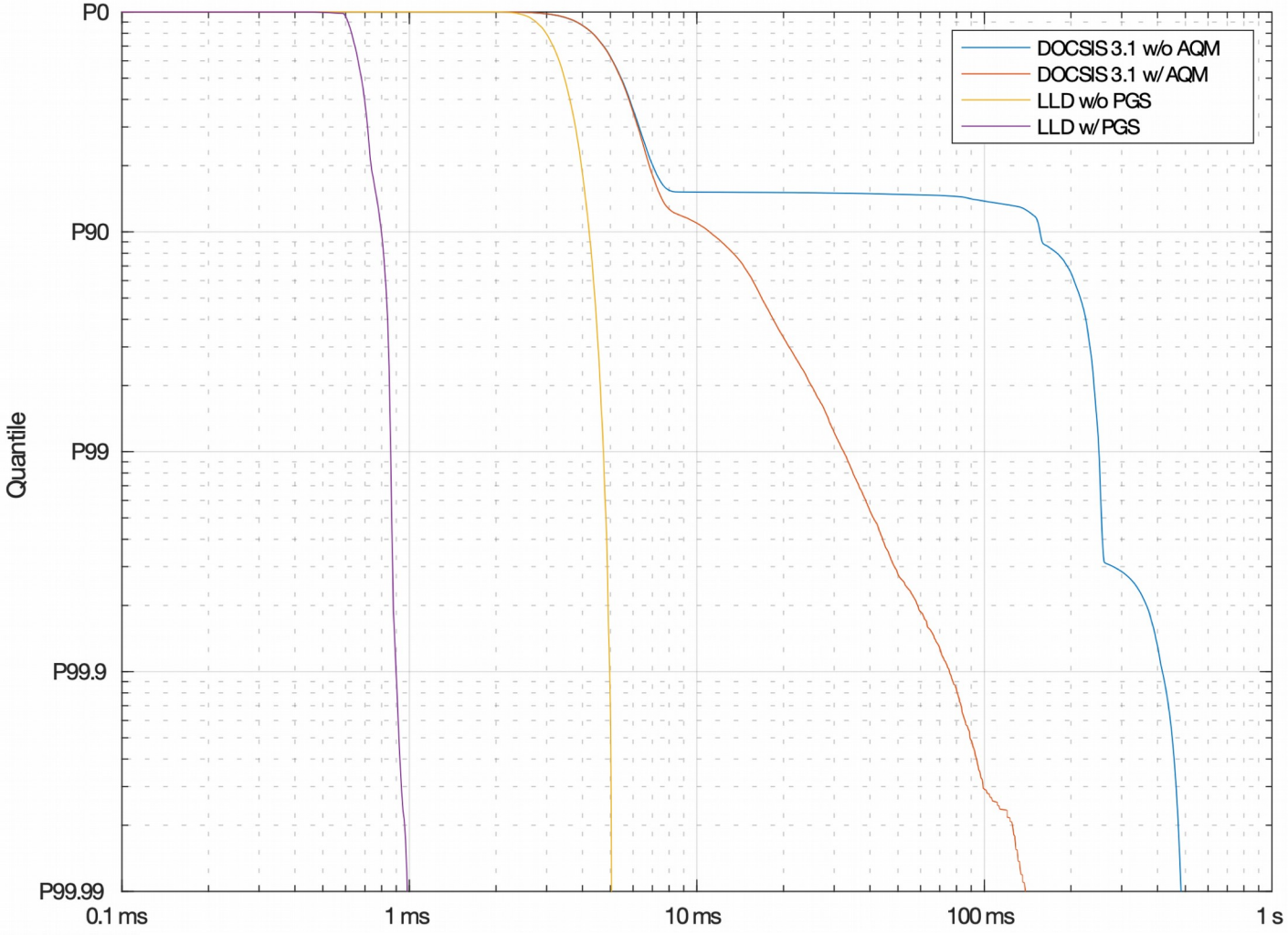
Q&A

spare slides

Scripture prophesized this

“We are concerned that the congestion control noise sensitivity is quadratic in *w* but it will take at least another generation of network evolution to reach window sizes where this will be significant.”

In footnote 6 of:
Jacobson, V. & Karels, M.J., "Congestion Avoidance and Control," Laurence Berkeley Labs Technical Report (November 1988) (a slightly modified version of the original published at SIGCOMM in Aug'88) URL: <http://ee.lbl.gov/papers/congavoid.pdf>



LL DOCSIS

Traffic:

- Low Latency Service Flow
  - game traffic
- Classic Service Flow
  - heavy load

Low Latency DOCSIS access link latency; Low Latency Service Flow

more info

All via L4S Landing page: https://riteproject.eu/dctth/

Linux Netdev

- [tcp-prague-netdev] Briscoe, B., De Schepper, K., Albisser, O., Misund, J., Tilmans, O., Kühlewind, M. & Ahmed, A. S., “Implementing the 'TCP Prague' Requirements for L4S” in Proc. Netdev 0x13 (Mar 2019)
- [dualpi2-netdev] “DUALPI2 - Low Latency, Low Loss and Scalable (L4S) AQM” in Proc. Netdev 0x13 (Mar 2019)
- [paced-chirping-netdev] “Paced Chirping - Rethinking TCP start-up” in Proc. Netdev 0x13 (Mar 2019)
- [AccECN-netdev] Mirja Kühlewind, “State of ECN and improving congestion feedback with AccECN in Linux” in Proc. Netdev 2.2 (Dec 2017)
- [RACK-netdev] Cheng, Y. & Cardwell, N.,”Making Linux TCP Fast” in Proc. Netdev 1.2 (Oct 2016)

IETF

- [ietf-l4s-arch] Briscoe (Ed.), B., De Schepper, K. & Bagnulo, M., "Low Latency, Low Loss, Scalable Throughput (L4S) Internet Service: Architecture," IETF Internet Draft draft-ietf-tsvwg-l4s-arch-03 (Oct 2018) (Work in Progress)
- [RFC8311] Black, D. “Explicit Congestion Notification (ECN) Experimentation” IETF RFC8311 (Jan 2018)
- [ietf-l4s-id] De Schepper, K., Briscoe (Ed.), B. & Tsang, I.-J., "Identifying Modified Explicit Congestion Notification (ECN) Semantics for Ultra-Low Queuing Delay (L4S)," IETF Internet Draft draft-ietf-tsvwg-ecn-l4s-id-05 (Nov 2018) (Work in Progress)
- [RFC8257] Bensley, S., Thaler, D., Balasubramanian, P., Eggert, L. & Judd, G., "Data Center TCP (DCTCP): TCP Congestion Control for Data Centers," RFC Editor RFC8257 (October 2017)
- [ietf-dualq-aqm] De Schepper, K., Briscoe (Ed.), B., Albisser, O. & Tsang, I.-J., "DualQ Coupled AQM for Low Latency, Low Loss and Scalable Throughput," IETF Internet Draft draft-ietf-tsvwg-aqm-dualq-coupled-08 (Nov 2018) (Work in Progress)
- [RFC7560] Kühlewind, M., Scheffenegger, R. & Briscoe, B. “Problem Statement and Requirements for Increased Accuracy in Explicit Congestion Notification (ECN) Feedback” IETF RFC7560 (2015)
- [ietf-AccECN] Briscoe, B., Scheffenegger, R. & Kühlewind, M., "More Accurate ECN Feedback in TCP," IETF Internet Draft draft-ietf-tcpm-accurate-ecn-07 (Jul 2018) (Work in Progress)
- [ietf-RACK] Cheng, Y., Cardwell, N., Dukkipati, N. & Jha, P., “RACK: a time-based fast loss detection algorithm for TCP” IETF Internet Draft draft-ietf-tcpm-rack-04 (Jul 2018) (Work in Progress)
- [ietf-ECN++] Bagnulo, M. & Briscoe, B., “ECN++: Adding Explicit Congestion Notification (ECN) to TCP Control “Packets IETF Internet Draft draft-ietf-tcpm-generalized-ecn-03 (Oct 2018) (Work in Progress)

DOCSIS

- [LLDOCSIS-spec] DOCSIS® 3.1 MAC and Upper Layer Protocols Interface (MULPI) Specification (i17+)
- [LLDOCSIS-overview] White, G., Sundaresan, K., and Briscoe, B. “Low Latency DOCSIS: Technology Overview” CableLabs White Paper (Feb 2019)

Research papers

- [DCttH] De Schepper, K., Bondarenko, O., Tsang, I.-J. & Briscoe, B., "`Data Centre to the Home': Deployable Ultra-Low Queuing Delay for All," RITE Project Technical report (June 2015)
- [PI2] De Schepper, K., Bondarenko, O., Tsang, I.-J. & Briscoe, B., "PI^2 : A Linearized AQM for both Classic and Scalable TCP," In: Proc. ACM CoNEXT 2016 pp.105-119 ACM (December 2016)
- [L4S-MMSYS] Bondarenko, O., De Schepper, K., Tsang, I.-J., Briscoe, B., Petlund, A. & Griwodz, C., "Ultra-Low Delay for All: Live Experience, Live Analysis," In: Proc. ACM Multimedia Systems; Demo Session pp.33:1-33:4 ACM (May 2016)
- [DCTCP] Alizadeh, M., Greenberg, A., Maltz, D.A., Padhye, J., Patel, P., Prabhakar, B., Sengupta, S. & Sridharan, M., "Data Center TCP (DCTCP)," Proc. ACM SIGCOMM'10, Computer Communication Review 40(4):63--74 (October 2010)
- [DCTCP-analysis] Alizadeh, M., Javanmard, A. & Prabhakar, B., "Analysis of DCTCP: Stability, Convergence, and Fairness," In: Proc. ACM SIGMETRICS'11 (2011)
- [CC-scaling-tensions] Briscoe, B. & De Schepper, K., "Resolving Tensions between Congestion Control Scaling Requirements," Simula Technical Report TR-CS-2016-001 (July 2017)
- [low-RTT-scaling] Briscoe, B. & De Schepper, K., "Scaling TCP's Congestion Window for Small Round Trip Times," BT Technical report TR-TUB8-2015-002 (May 2015)
- [paced-chirping] Misund, Joakim and Briscoe, Bob, “Paced Chirping: Rapid flow start with very low queuing delay” In Proc IEEE Global Internet Symposium 2019 (Apr/May 2019)