Data Centre Network Latency Control
Dagstuhl Seminar, Jul 2016

Data Centre to the Home
Low Delay for All

Bob Briscoe*
Koen De Schepper†
Olga Bondarenko*
Inton Tsang†

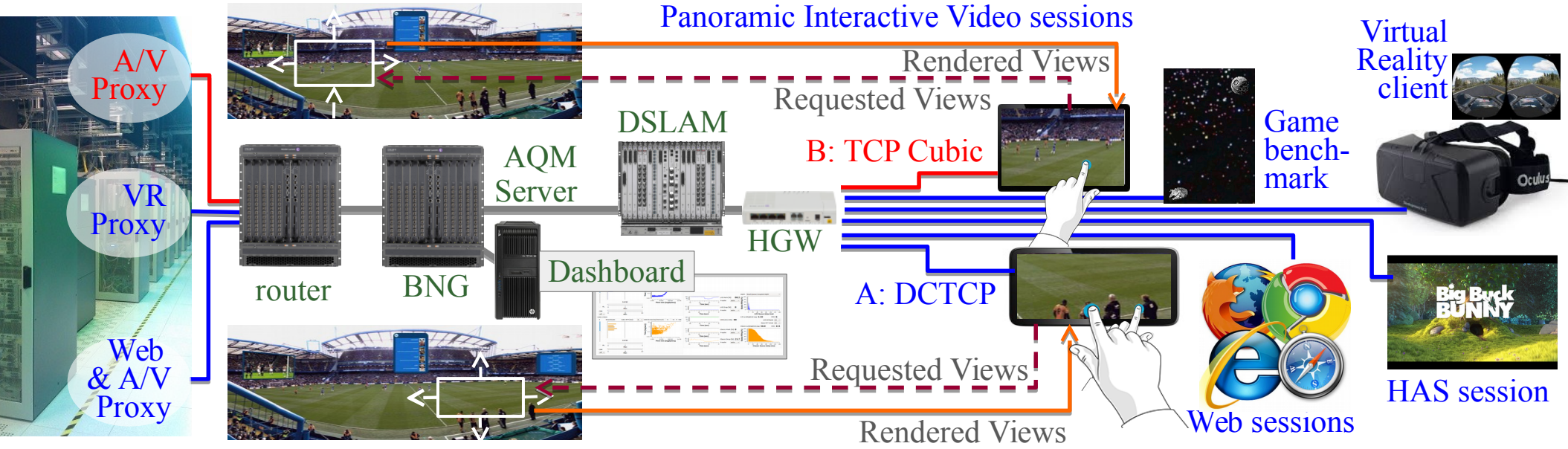
* [simula . research laboratory]

† NOKIA Bell Labs

R/TE
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.

“Data Centre to the Home” DCttH



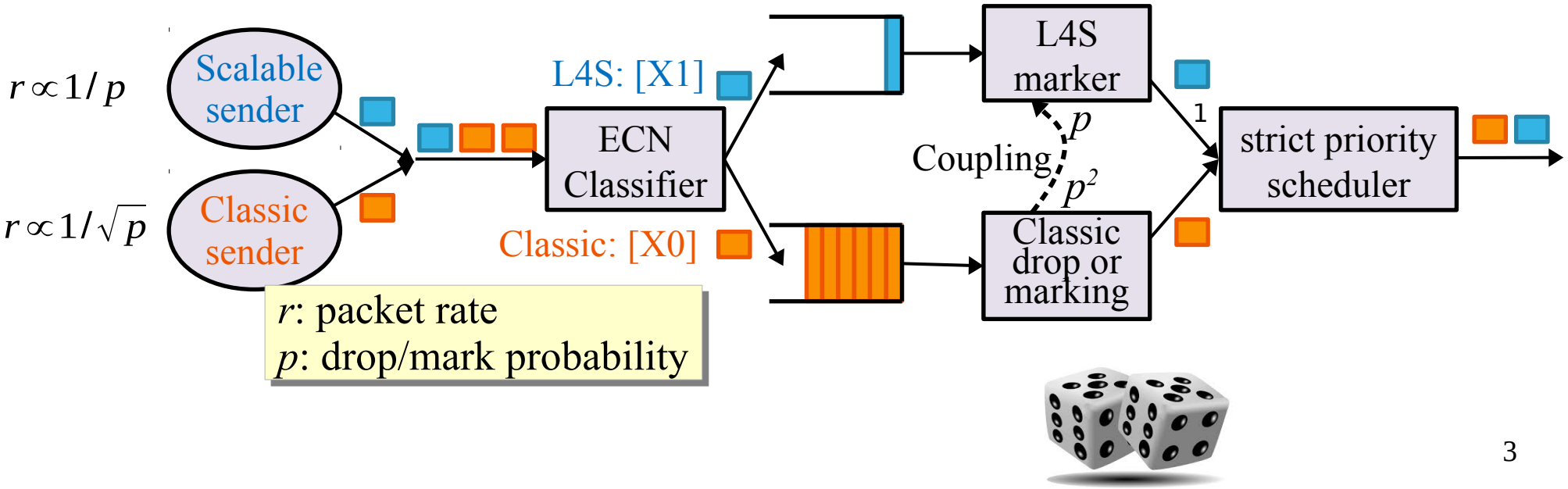
- Ultra-Low Delay for **all** apps
  - *Not* Diffserv; not low delay for some at the expense of others

- heavy load of multiple latency-sensitive apps: **all** packets <1ms queuing delay
  - incl. finger gestures & oculus rift interaction generating video on-the-fly in a DC
  - accessed over real public broadband (7ms base delay)

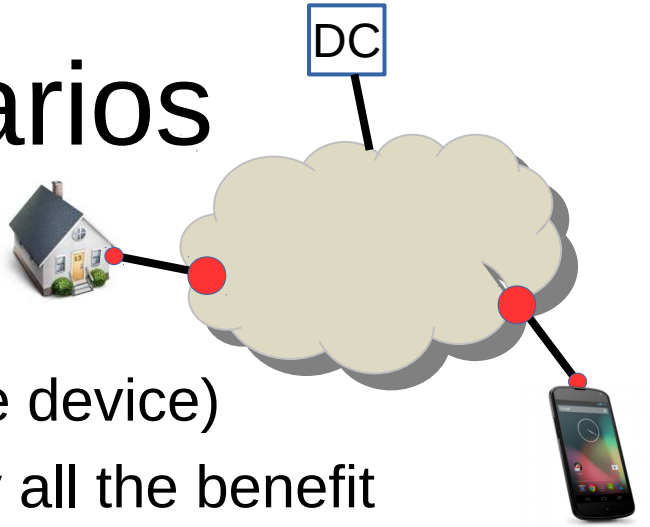
- Aim: to be worth the deployment hassle – much better than today; new app enabler
- L4S: Low Latency, Low Loss, Scalable throughput (L4S)
  - A new service for **all** Internet traffic to transition to

How? DCTCP over the Internet

- DCTCP coexistence with Reno/Cubic
  - Throughput equivalence without flow inspection using...
- DualQ Coupled AQMs: a 'semi-permeable membrane' that:
  - partitions latency (separate queues for L4S & Classic)
  - but pools bandwidth (shared by apps/transport, not by network)

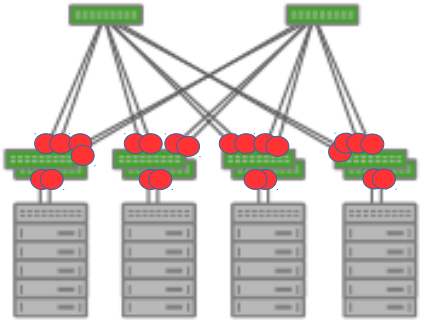


deployment scenarios



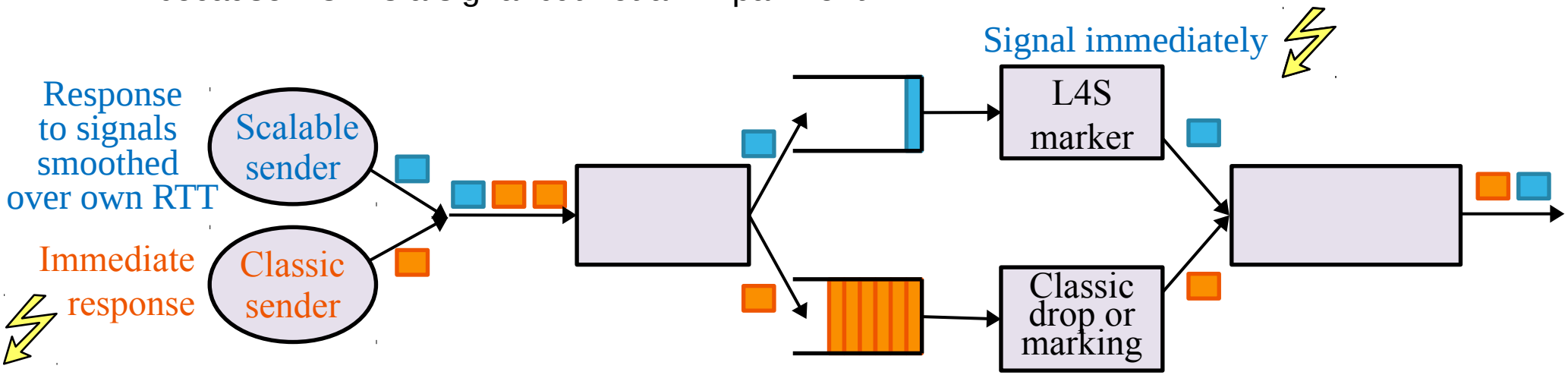
- access bottleneck
  - per 'site' (home, office, campus or mobile device)
  - deploying DualQ here should give nearly all the benefit
  - typically as leaf queues of pre-existing inter-site scheduling hierarchy

- data centre
  - ingress and egress bottleneck would typically give nearly all the benefit
  - e.g. all the outputs of the top-of-rack switch
  - and ingress to inter-DC WAN links



Why is performance so much better?
Immediate signalling

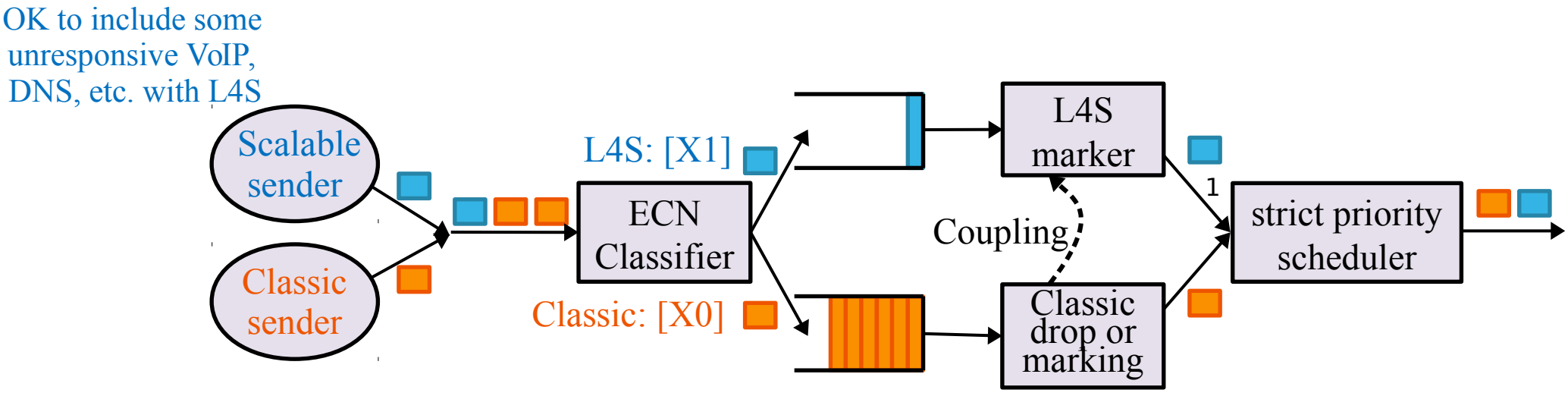
- Today's AQMs defer drop for ~100ms
  1)to allow time for a worst-case RTT response
  because: the network doesn't know each packet's RTT
  2)to avoid drop unless the queue proves persistent
  because: drop is an impairment as well as a signal
- Using ECN for L4S makes it feasible to signal immediately
  - because ECN is a signal but not an impairment



Problem with the Classic approach:
a flow with RTT=5ms
gets no signal for 20 round-trips

Framework for Diverse Solutions

- The DualQ Coupled AQM draft is structured as a framework



many scalable algorithms already: DCTCP, Relentless, S-SCREAM, etc. (not yet with safety features search "TCP Prague Requirements")	Classify at IP layer or lower	2 different classic AQMs already implemented: Curvy RED & PI2	2 different schedulers already implemented: strict priority & MEDF

an incrementally deployable clean-slate

- DCTCP serves us as an existence proof
  - L4S should work with any good 1/p congestion control

- can redesign everything together
  - new AQM
  - new flow-start
  - new steady-state congestion control
- brief time window to solve all those old cc problems properly

- IETF L4S BoF to initiate standardisation next week
- search “dctth” for videos, papers, etc
  or https://riteproject.eu/dctth/