Ultra-Low Queuing Delay for All
Bar BoF

Bob Briscoe

[simula . research laboratory]

Koen De Schepper

NOKIA Bell Labs

IETF-95 Apr 2016

Note Well

Any submission to the IETF intended by the Contributor for publication as all or part of an IETF Internet-Draft or RFC and any statement made within the context of an IETF activity is considered an "IETF Contribution". Such statements include oral statements in IETF sessions, as well as written and electronic communications made at any time or place, which are addressed to:

- The IETF plenary session
- The IESG, or any member thereof on behalf of the IESG
- Any IETF mailing list, including the IETF list itself, any working group or design team list, or any other list functioning under IETF auspices
- Any IETF working group or portion thereof
- Any Birds of a Feather (BOF) session
- The IAB or any member thereof on behalf of the IAB
- The RFC Editor or the Internet-Drafts function

All IETF Contributions are subject to the rules of RFC 5378 and RFC 3979 (updated by RFC 4879).

Statements made outside of an IETF session, mailing list or other function, that are clearly not intended to be input to an IETF activity, group or function, are not IETF Contributions in the context of this notice. Please consult RFC 5378 and RFC 3979 for details.

A participant in any IETF activity is deemed to accept all IETF rules of process, as documented in Best Current Practices RFCs and IESG Statements.

A participant in any IETF activity acknowledges that written, audio and video records of meetings may be made and may be available to the public.

Proposed Agenda

09:00

- Note Well, Agenda Bashing [Phil]
- Introduction/Background (very brief) [Bob]
- Clarification Questions

09:15

- Status updates [Koen to lead], e.g.:
    - what people are doing on parts of the problem:
    - planned work
    - evaluation work
    - interest in implementing
    - willingness to review

09:40

- Build a standardisation roadmap [Bob to lead]
- Build a BoF for the Berlin time-frame
    - should it be non-WG forming?
    - volunteers to help with organisation / writing problem statement, etc.
    - which mailing list?
    - what name?
- Discussion / Q&A

09:55 end

Low Latency Low Loss Scalable throughput (L4S) – background

Latency so low it sticks to your finger

- Recall: demo at Prague IETF (aqm wg & bits-n-bites)
  - see https://riteproject.eu/dctth/ for videos, papers, etc
- L4S could incrementally replace "best efforts"
  - ultra-low queuing delay
  - zero congestion loss
  - scalable throughput (beyond Reno, Compound, Cubic)
- Eventually for *all* Internet traffic
- Aim: to be worth the deployment hassle – so much better than today

very high level

- problem: TCP is the elephant in the room
- solution: build another room without the elephant

TCP

Scalable
TCP
+ everything
else
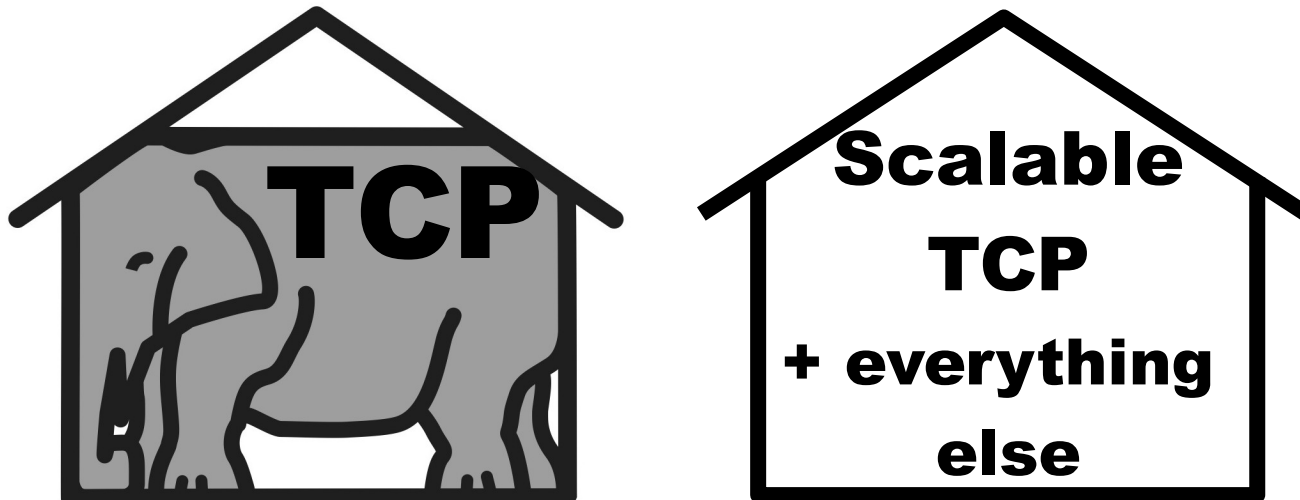
Current status

DualQ:

- Added and evaluated overload mechanisms

- Evaluation on a bigger range of RTTs and link capacities

- Evaluation of mixed RTTs

- Alternative Classic AQMs: PI2

TCP-Prague:

- Simulation of some alternative scalable CC mechanisms
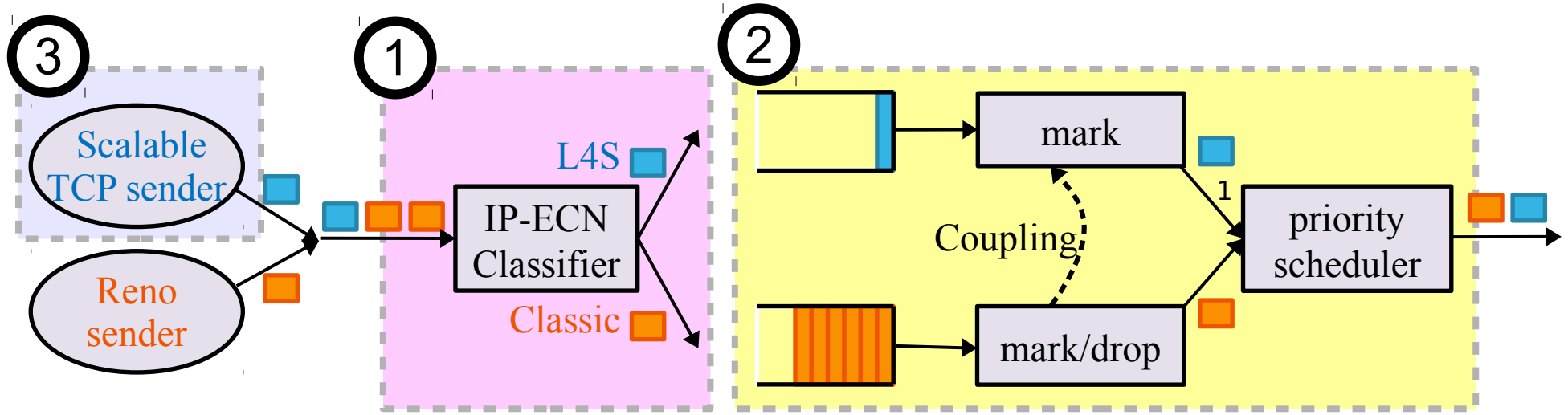
Planned

DualQ:

- Source cleanup and Open Source release of DualPI2

TCP-Prague:

- RTT independence

- Faster adapting to available line rate (both in slow start and congestion avoidance)

- Linux implementation of simulated ideas

- Evaluation

# 3 parts to standardise

1)	The identifier	draft-briscoe-tsvwg-ecn-l4s-id
2)	The DualQ AQM	draft-briscoe-aqm-dualq-coupled
3)	Scalable transport	draft-ietf-tcpm-accurate-ecn draft-ietf-tcpm-dctcp (bis)

- #1, #2 are as general as possible

- #3a) fixes TCP feedback, other transports are already OK

- #3b) one concrete example transport behaviour: DCTCP

An incrementally deployable clean slate
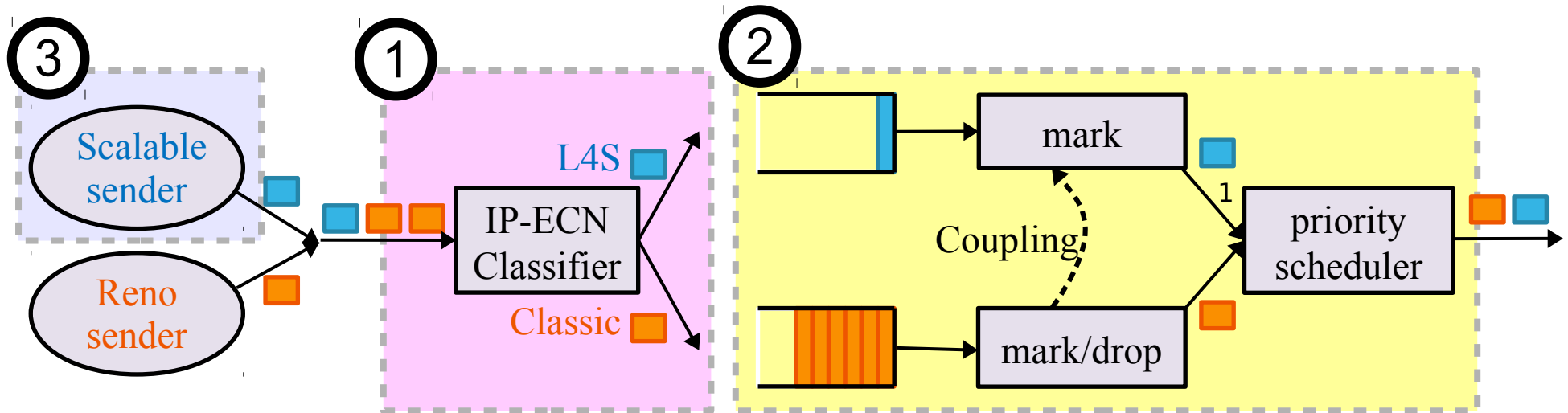
1)	The identifier	draft-briscoe-tsvwg-ecn-l4s-id	tsvwg?
2)	The DualQ AQM	draft-briscoe-aqm-dualq-coupled	aqm?
3)	Scalable transport	draft-ietf-tcpm-accurate-ecn draft-ietf-tcpm-dctcp (bis) "TCP Prague" "SCTP Prague" "RMCAT Prague"	tcpm? tcpm? tcpm? tcpm? rmcat?

"TCP Prague" Requirements

	Requirement	Title of I-D or potential I-D	Ref	WG	TCP	DCTCP	DCTCP bis	TCP Prague	"SCTP Prague"?	"rmcat Prague"?
1	Fall back to Reno/Cubic on loss	DCTCP: TCP Congestion Control for Datacenters	draft-ietf-tcpm-dctcp	TCPM		Y	Y	Y	Y	Y
2	Negotiate altered feedback semantics	More Accurate ECN Feedback in TCP	draft-ietf-tcpm-accurate-ecn	TCPM	Y	Y	Y	Y	n/a	n/a
3	Use of a standardised packet identifier	Identifying Modified ECN Semantics for Ultra-Low Queuing Delay	draft-briscoe-tsvwg-ecn-l4s-id	TSVWG?		Y	Y	Y	Y	Y
4	If base-RTT is low, handle a window of less than 2, don't grow queue	Scaling TCP's Congestion Window for Small Round Trip Times	slides-93-iccrg-5.p	TCPM?	Y	Y	Y	Y	Y	?
5	Reduce RTT-dependence	Contribution-driven		TCPM?			Y	Y	Y	?
6	Smooth ECN feedback over flow's RTT, not RTT hard-coded for DCs	Contribution-driven		TCPM?		Y?	Y	Y	Y	?
7	Fall back to Reno/Cubic if classic ECN bottleneck detected	Contribution-driven		TCPM?				Y	Y	?
8	Faster-than-additive increase	Contribution-driven	e.g. Adaptive Acceleration (A^{2}DTCP)	TCPM?			Y	Y	Y	?
9	Less drastic exit from slow-start	Contribution-driven		TCPM?			Y	Y	Y	?

BoF

- Berlin timeframe

- Non-WG forming?

- Volunteers?

    - organisation

    - problem statement

- Which mailing list?

- What name?

Q&A

large saw teeth can ruin the quality of your experience

Q) why is queuing delay of DCTCP so low?
A1) Finer saw-teeth of a Scalable TCP

Today (at best)
TCP on end-systems
AQM at bottlenecks

if change bottlenecks
alone

DCTCP
change bottlenecks
and TCP

buffer occupancy

buffer size

buffer kept for bursts

TCP saw-teeth
seeking the
operating point

shallower
operating
point

lower queuing
delay and more
predictably low

smooth TCP:
more smaller
saw-teeth

AQM
operating
point

line utilisation

cuts delay but
poorer line
utilisation

good line
utilisation

time

also highly insensitive
to threshold configuration