On Designing a Congestion Control Algorithm for Low Flow Durations and Zero Loss

Nandita Dukkipati

High Performance Networking Group
Stanford University

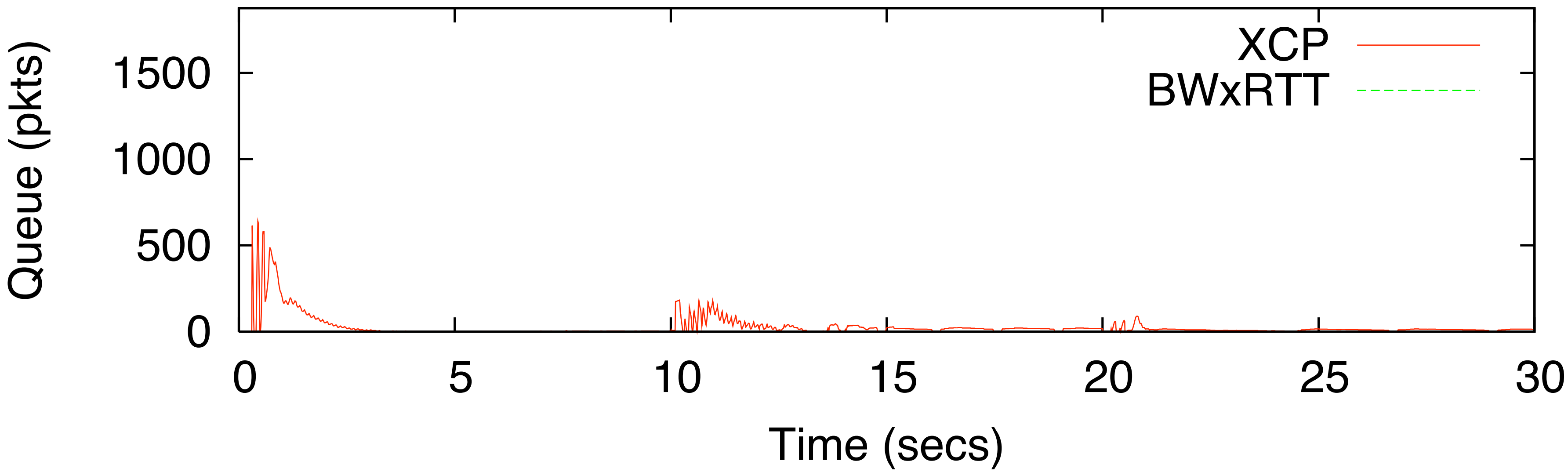
Introduction

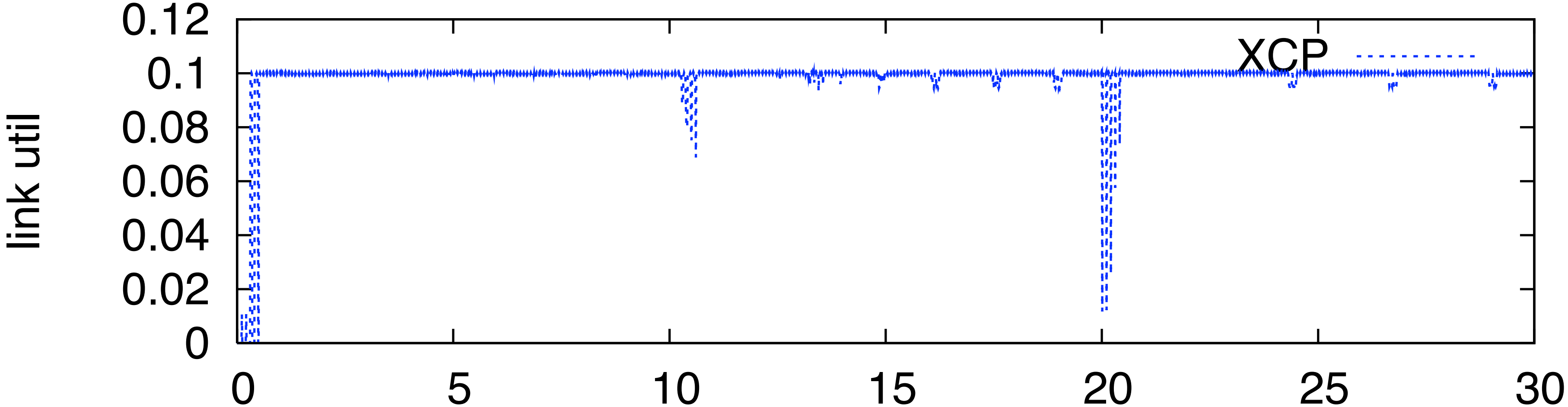
- Designing congestion control, if we get to start again ?
- Choose TCP --- No, no and no!
- Two goals:
  - Finish flows quickly
  - Don't lose packets
- Know how to achieve goals individually
  - RCP: fast, sometimes lossy
  - XCP: sometimes slow, zero loss
- This talk: Can I have both?

XCP's Strength: Bounded Worst Case

Small buffer occupancy and zero loss for <u>any</u> traffic pattern

Long flows: high link utilization

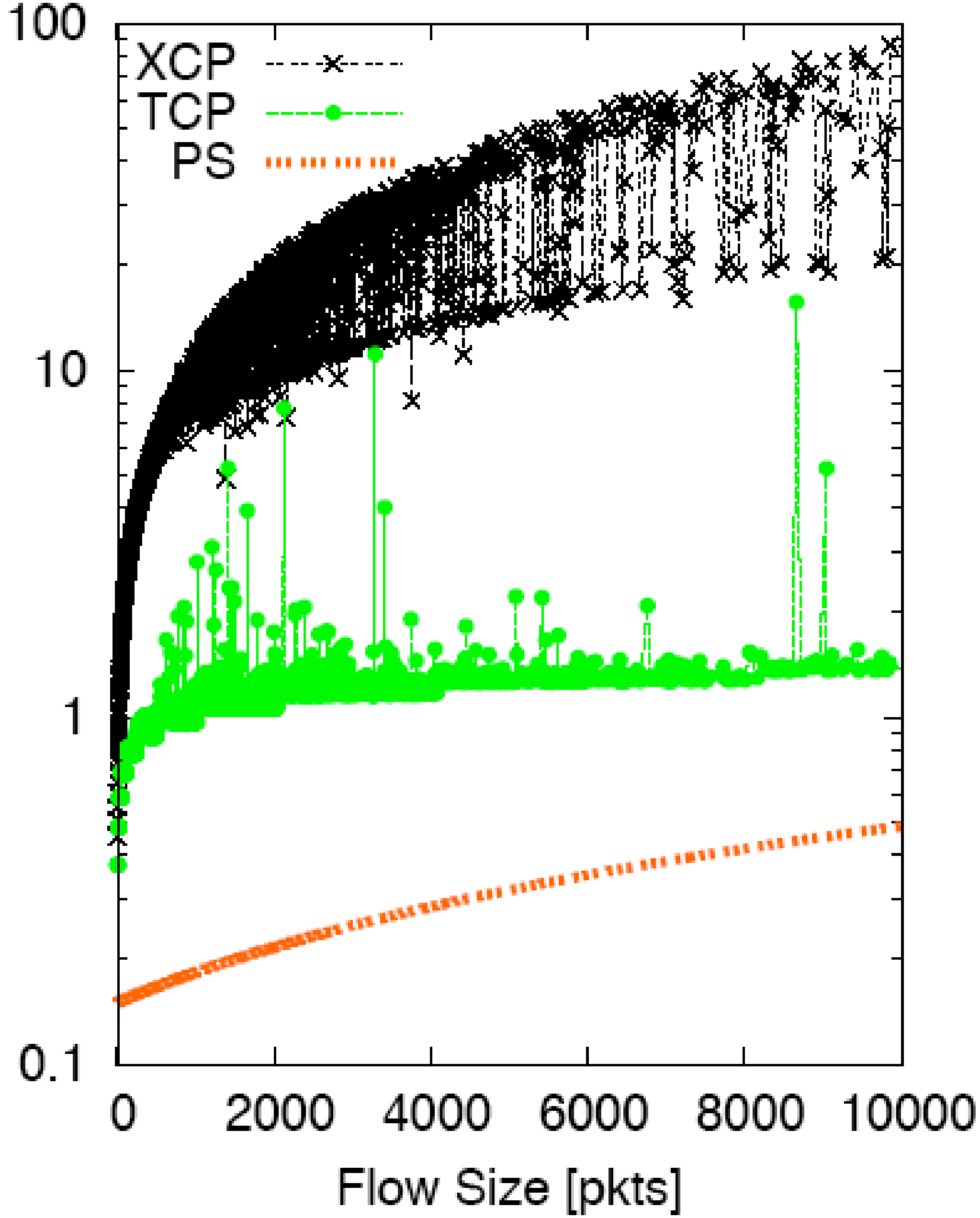




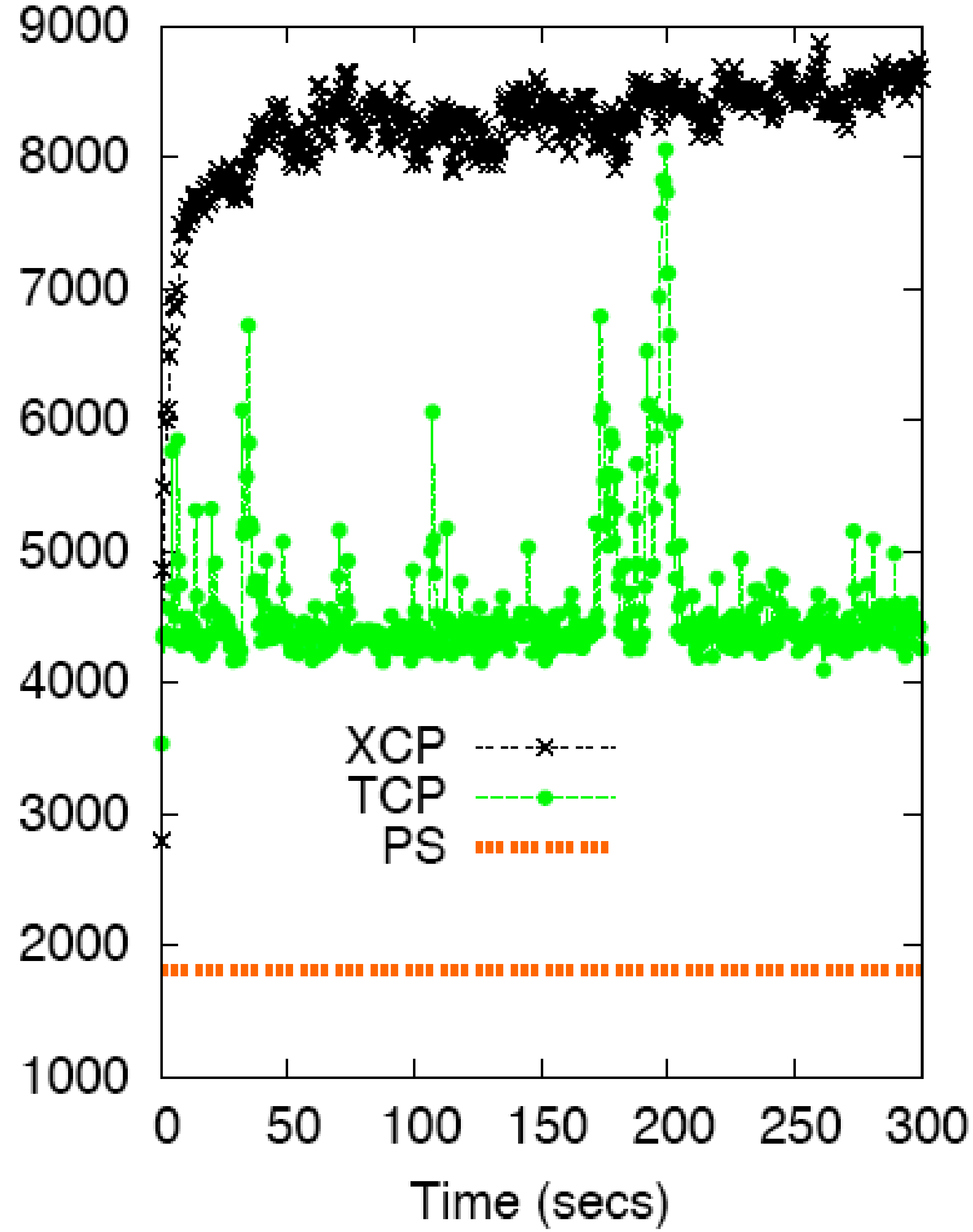
XCP's Weakness: Poor Average Case

Common Internet scenario

Flow Duration (secs) vs. Flow Size

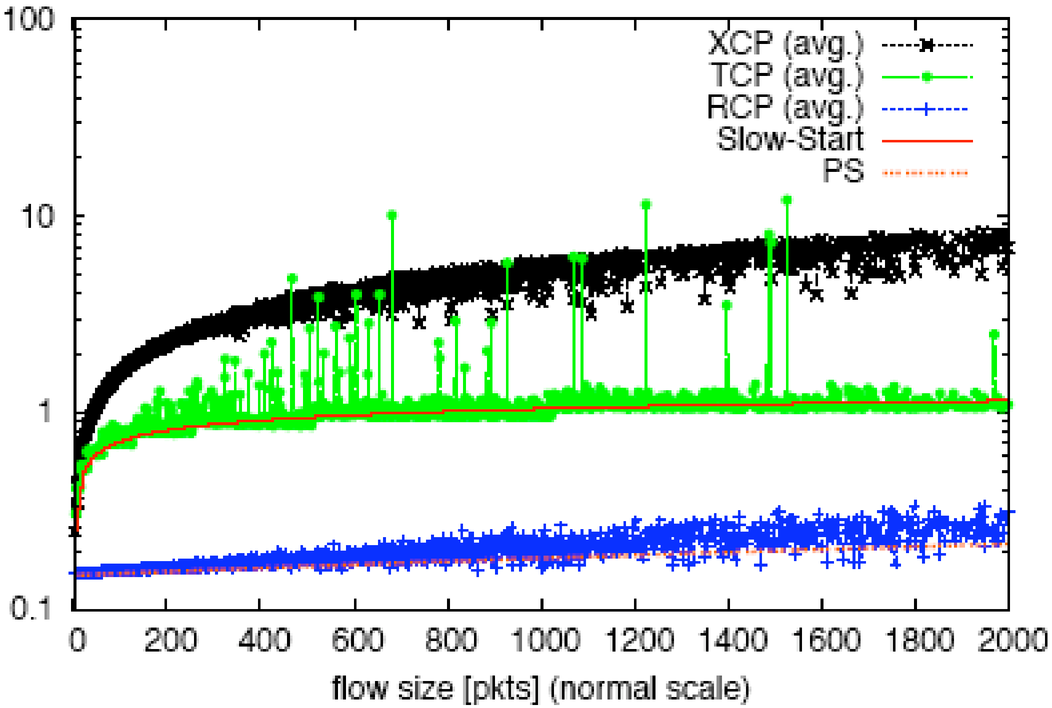


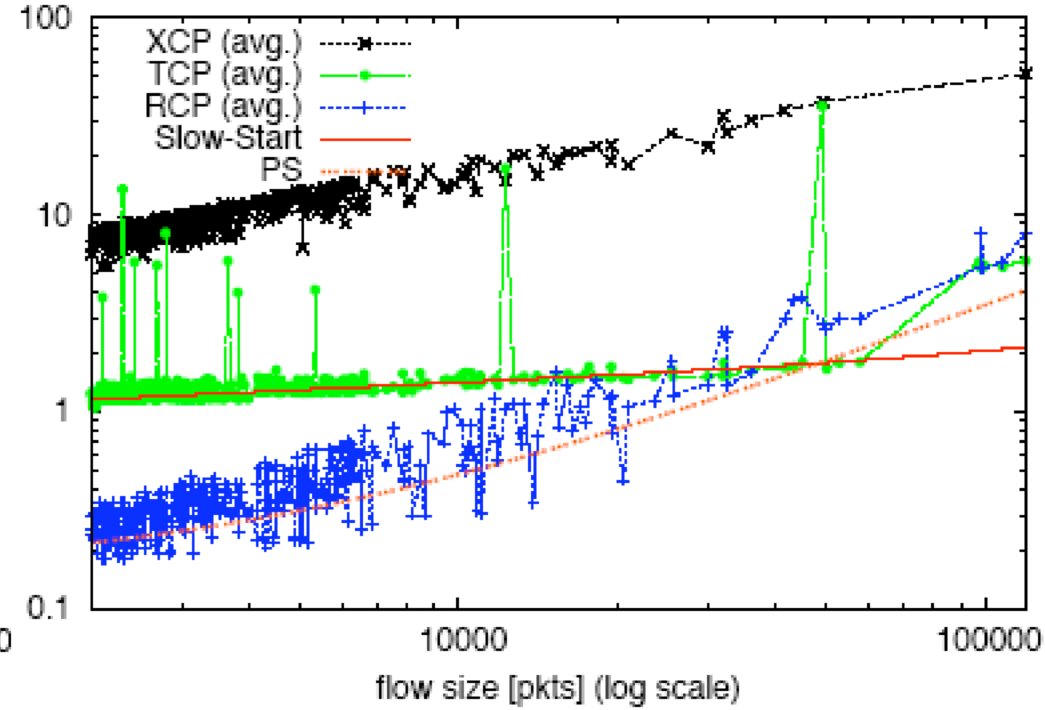
Active Flows vs. time



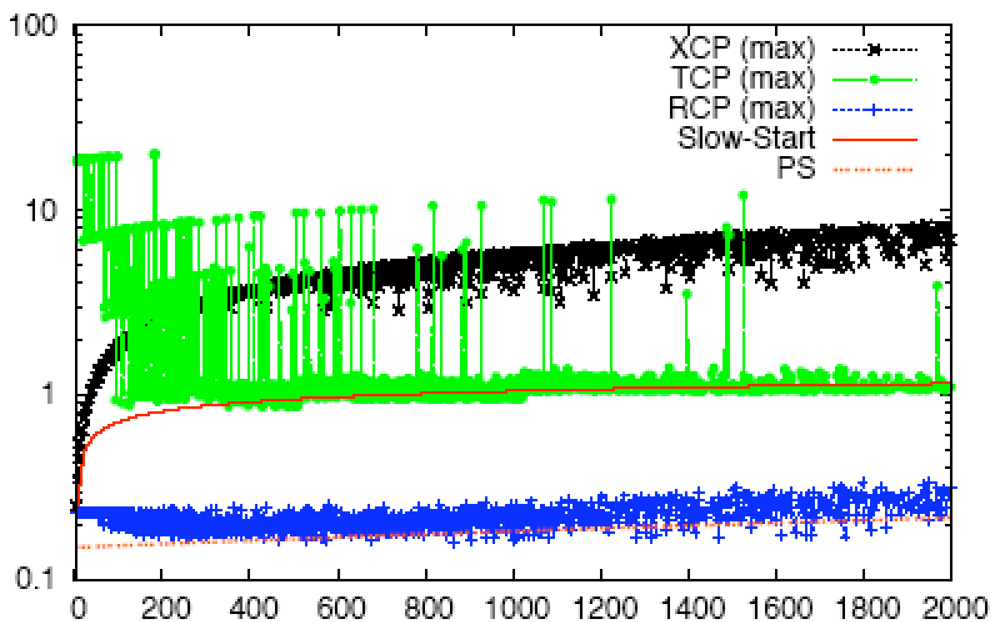
RCP’s Strength: Good Average Case

Flow completion times close to ideal Processor Sharing



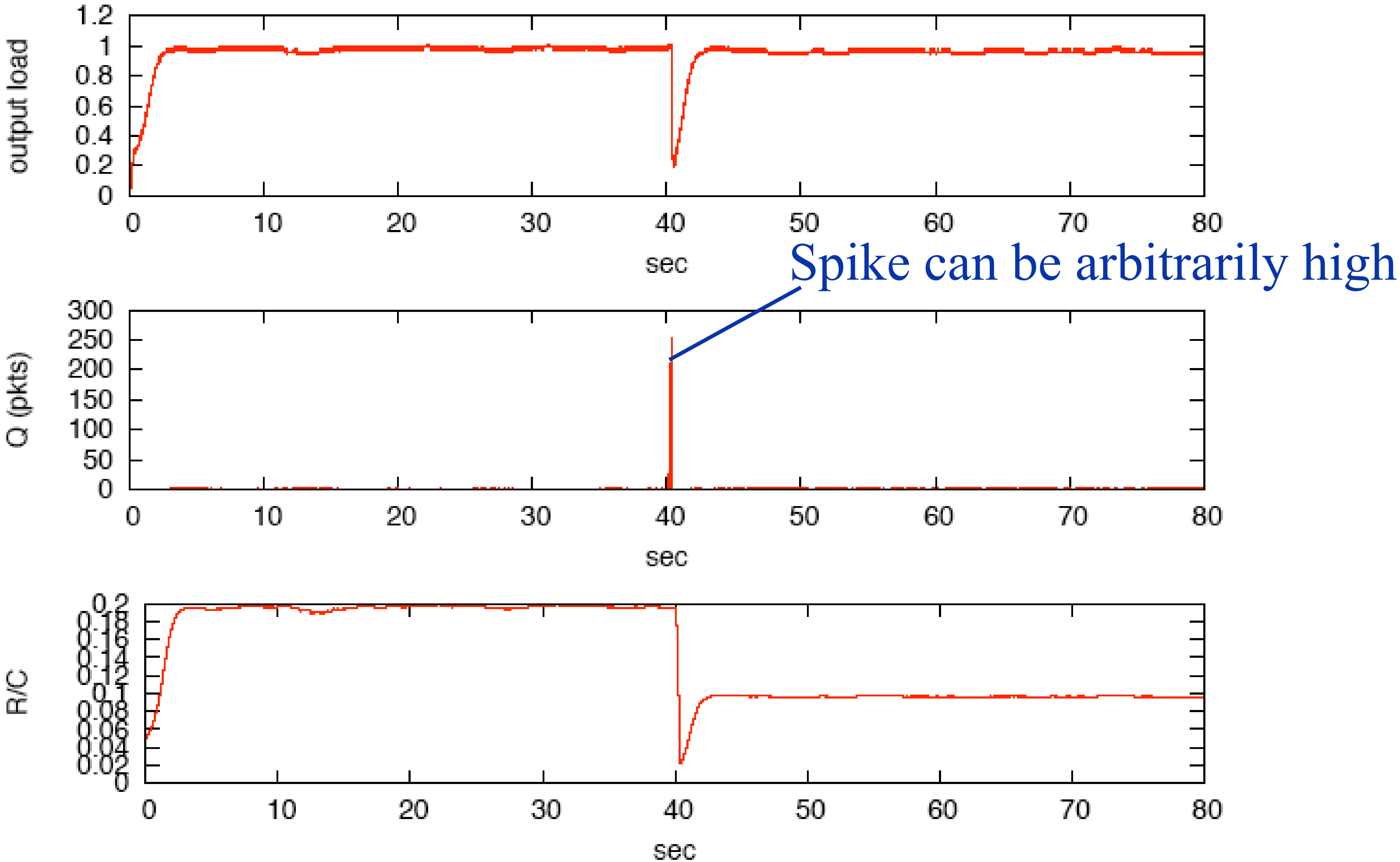


Max. FCT



RCP's Weakness: Unbounded Worst Case

A lot of flows starting at once: $N \times R(t) >> C$



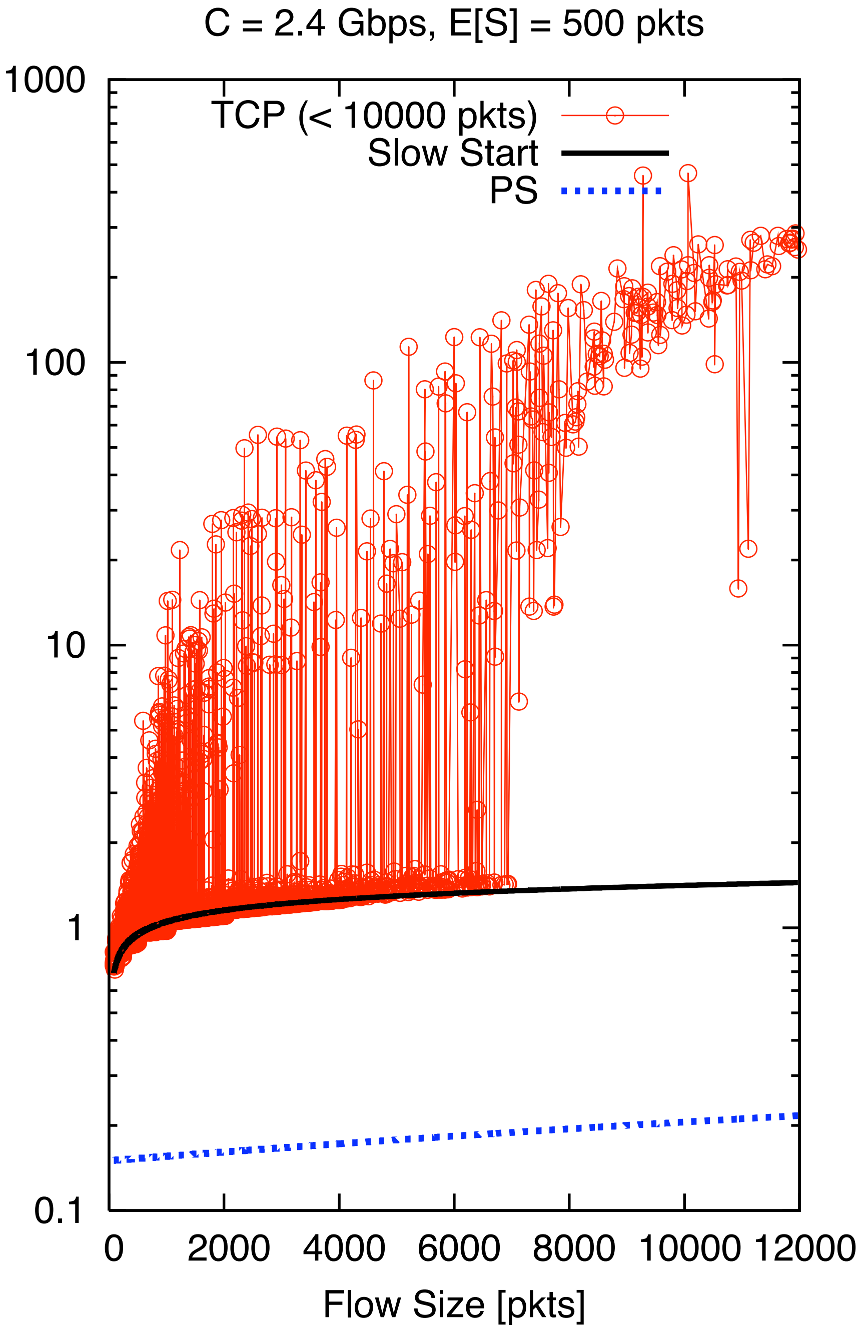
Question

Can we have the best of the two ?

-- good average case behavior like RCP

-- zero losses under any traffic pattern like XCP

Why care about losses anyway?



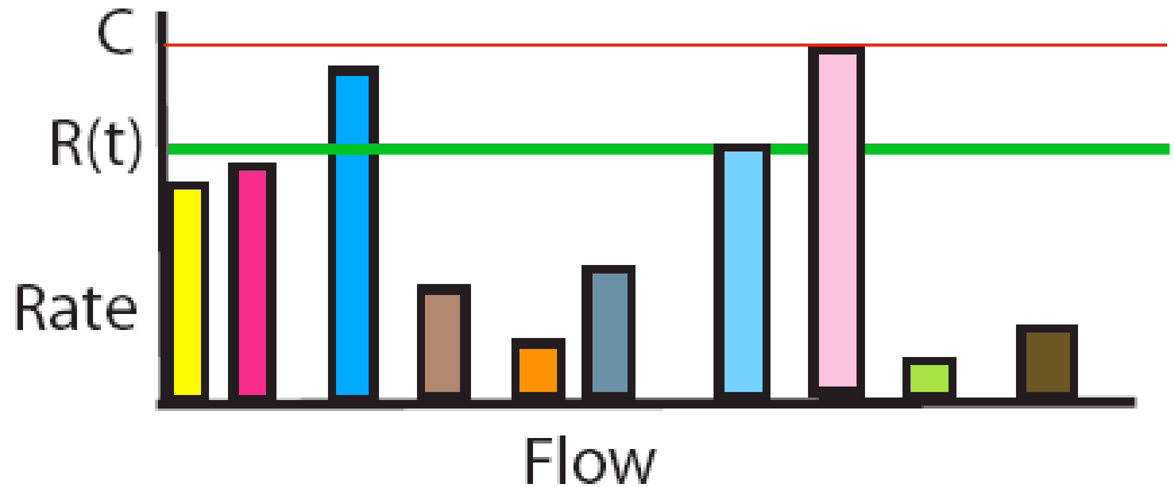
- Losses (and large queues) == unpredictability in the network
  - timeouts and retransmissions
  - flows last many more RTTs
- Example of TCP flows with and without losses
- Stronger abstraction of a network without losses under any traffic pattern

Intuition for achieving zero loss

RCP Rate Equation:

$$R(t) = R(t-T)[1 + \frac{\frac{T}{d_0}(\alpha(C - y(t)) - \beta\frac{q(t)}{d_0})}{C}]$$

Flow Snapshot:



$$N \times R(t) >> C$$

$$\sum_{i=1}^{N} R_i(t) = y(t) \qquad \sum_{i=1}^{N} [R(t) - R_i(t)]^+ = \text{unbounded!}$$

Aggregate bound = $C - y(t) + (B - q(t))/d_0$

worst case buffer occupancy

Using XCP Equations for achieving zero loss

A flow packet: $cwnd_i, rtt_i, feedback_i$

$$feedback_i = p_i - n_i$$

Negative feedback:

Computed from RCP equation

$$\Delta throughput_i = \max(0, \frac{cwnd_i}{rtt_i} - R(t))$$

$$n_i = \frac{\max(0, \frac{cwnd_i}{rtt_i} - R(t))}{\frac{cwnd_i}{rtt_i}}$$

Using XCP Equations for achieving zero loss

Positive feedback:

$$\Delta throughput_i = \max(0, R(t) - \frac{cwnd_i}{rtt_i}) = \frac{\Delta cwnd_i}{rtt_i}$$

$$p_i \propto \frac{\Delta cwnd_i}{\sharp\text{pkts in control interval } \bar{d}}$$

Computed from RCP eqn.

$$p_i = \xi_p \times \frac{rtt_i^2}{cwnd_i} \times \max(0, R(t) - \frac{cwnd_i}{rtt_i})$$

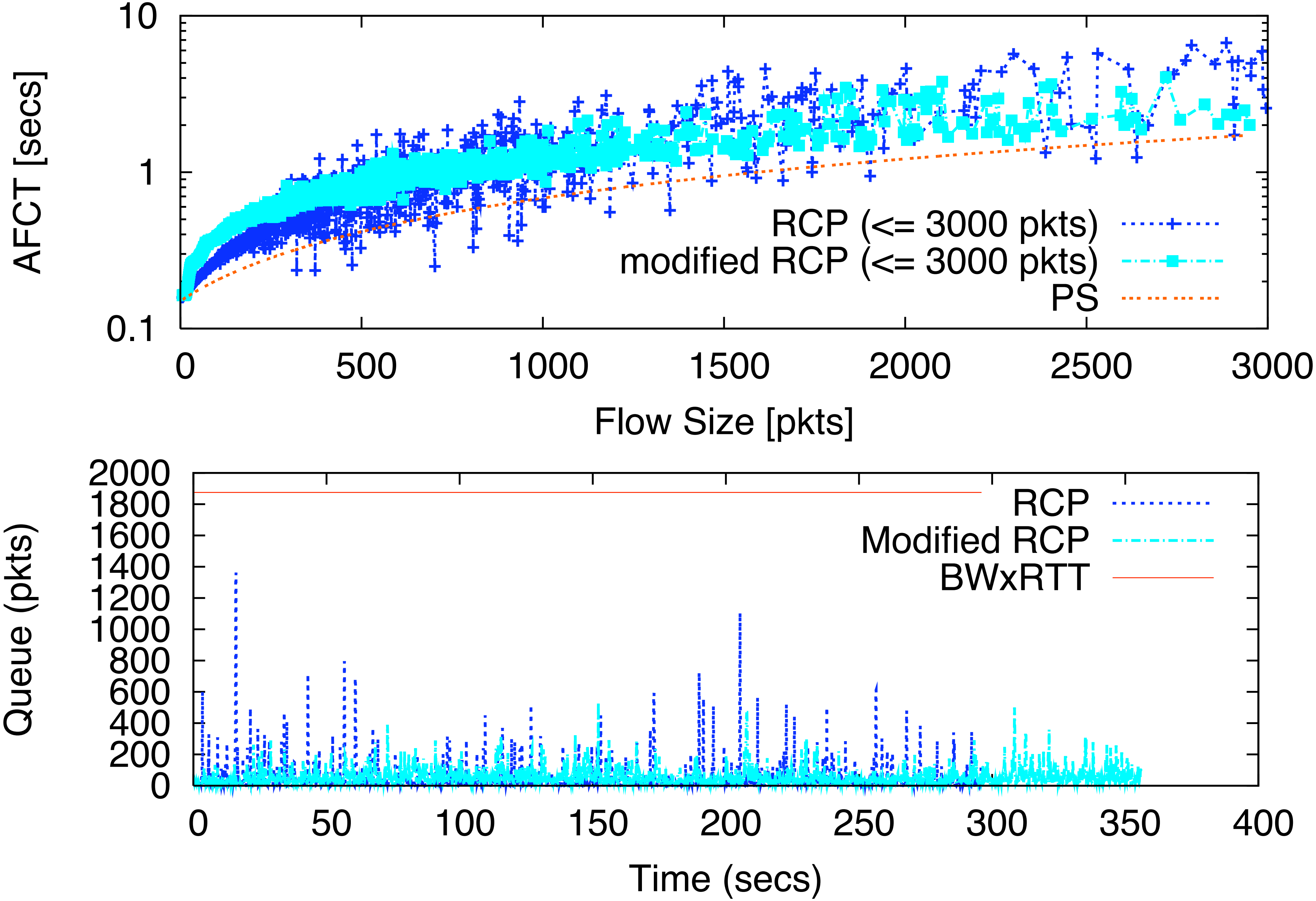
$$[C - y(t)]\bar{d} + (B - q(t))$$

$$\frac{\phi}{\bar{d}} = \sum_{i=1}^{L} \frac{p_i}{rtt_i} \qquad \xi_p = \frac{\phi}{\bar{d}\,[\sum^{L} \frac{rtt_i}{cwnd_i} \times \max(0, R(t) - \frac{cwnd_i}{rtt_i})]}$$

$$p_i = min(\frac{[R(t) - \frac{cwnd_i}{rtt_i}]^+}{\frac{cwnd_i}{rtt_i}}, \xi_p \frac{rtt_i^2}{cwnd_i} \max(0, R(t) - \frac{cwnd_i}{rtt_i}))$$

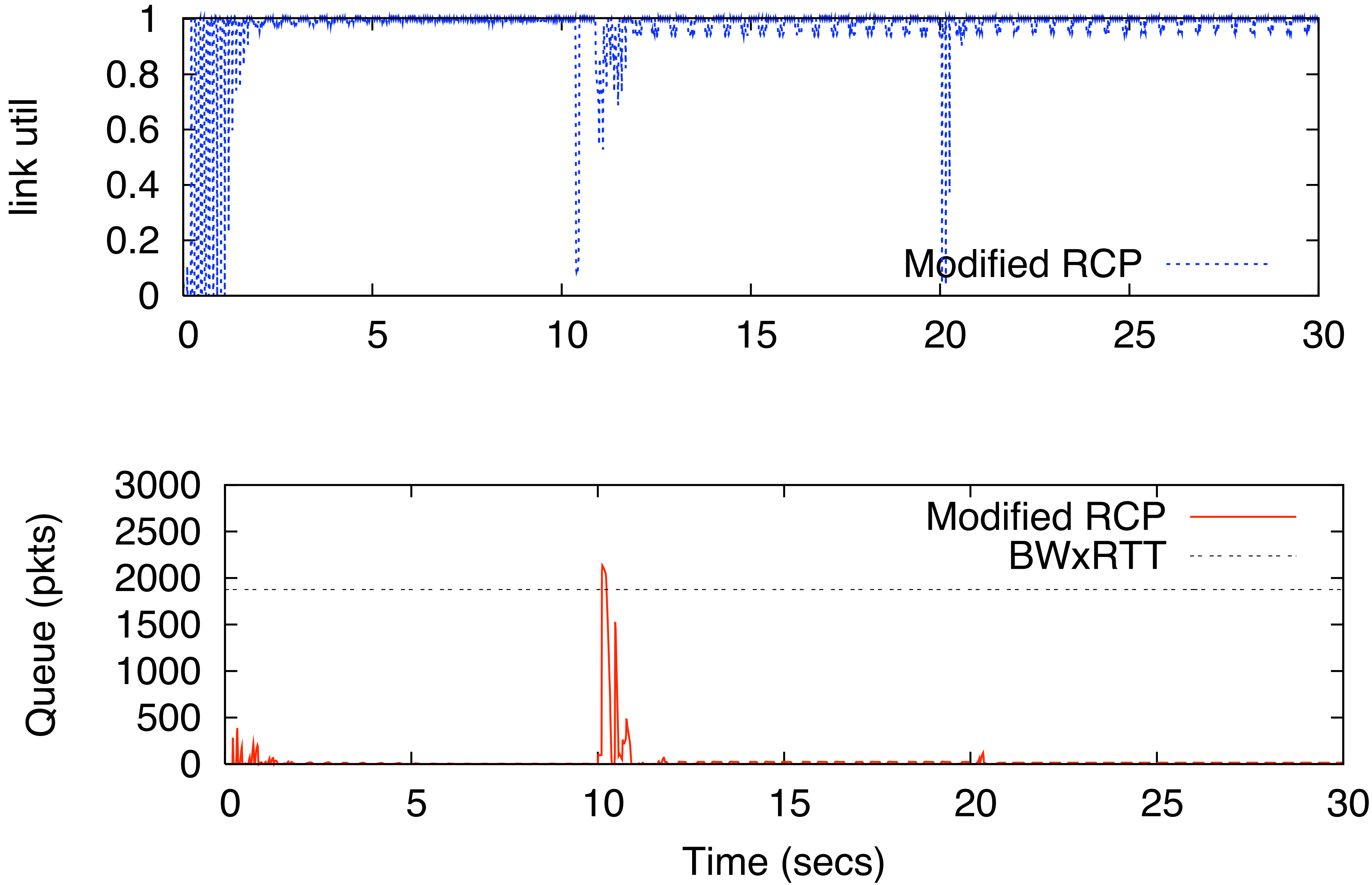
Modified RCP: Average Case Behavior

Flow completion times reasonably close to idea PS



Modified RCP: Worst Case Behavior

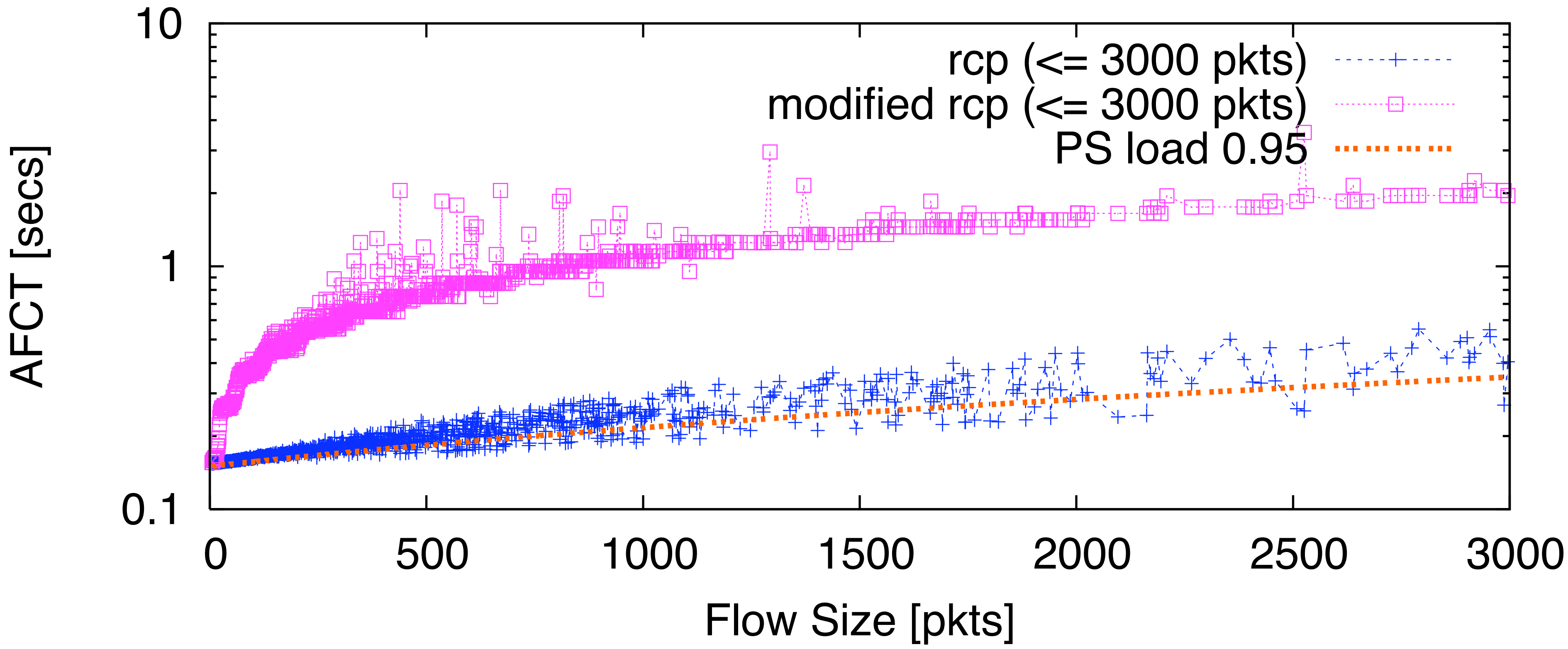
Bounded worst case



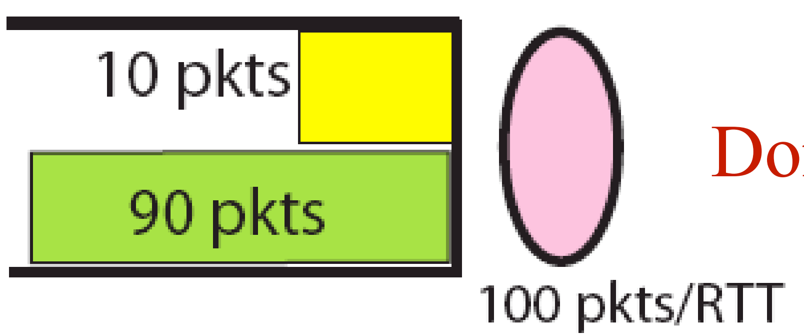
The complete truth is...

Not-so-close to PS for some scenarios

C=2.4 Gbps, load=0.95, rtt=0.1, pareto shape=1.2



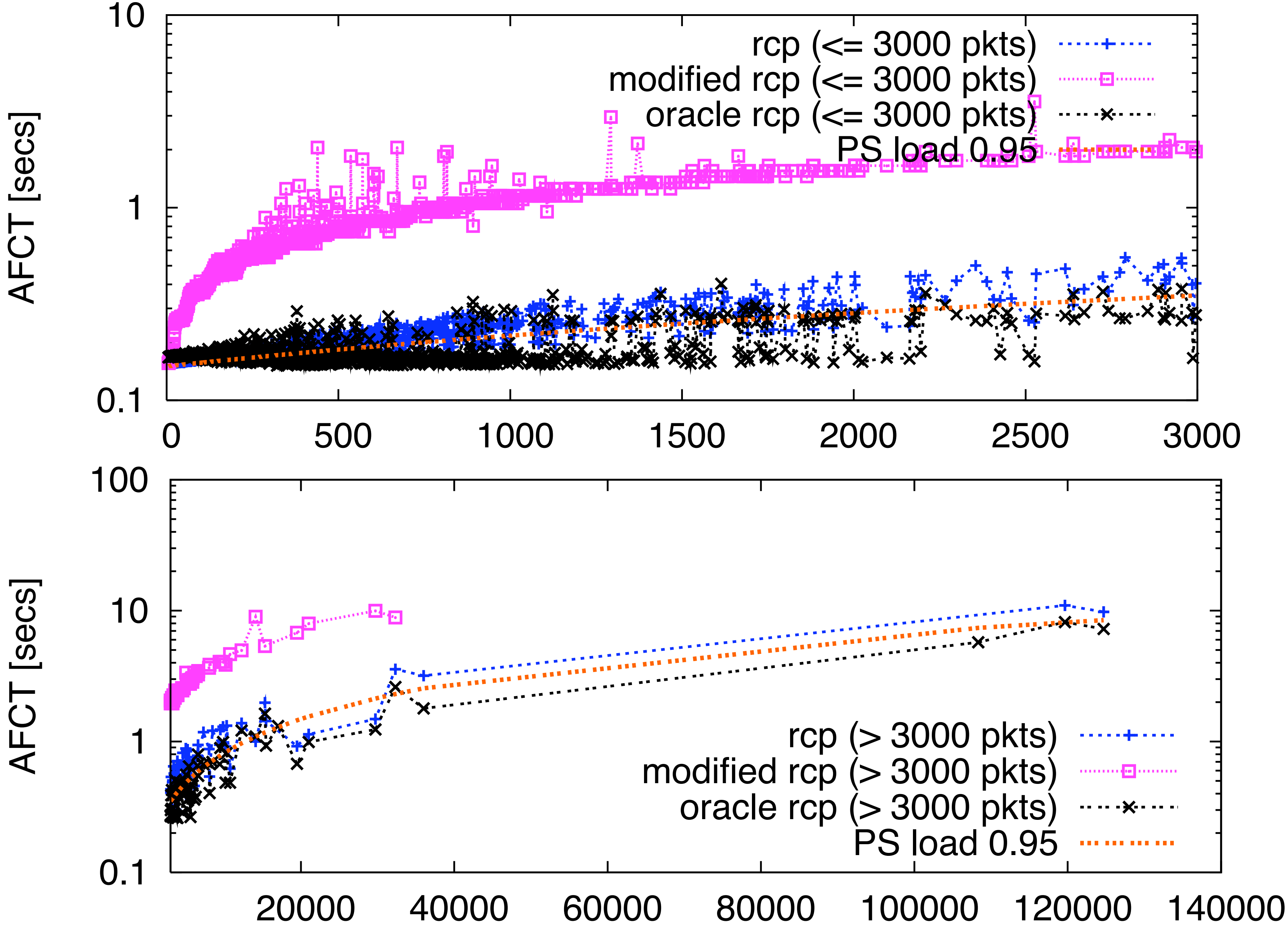
Why ?



Don’t know flow sizes

If you know flow sizes (desired rate_i)

$$p_i = \xi_p \times \frac{rtt_i^2}{cwnd_i} \times \max(0, \min(\text{desired rate}_i - \frac{cwnd_i}{rtt_i},\ \ R(t) - \frac{cwnd_i}{rtt_i}))$$



Simpler Implementation

Positive feedback:

$$\Delta throughput_i = \max(0, R(t) - \frac{cwnd_i}{rtt_i}) = \frac{\Delta cwnd_i}{rtt_i}$$

$$p_i = \frac{\Delta cwnd_i}{\sharp\text{control pkts in interval } \bar{d}}$$

$$\sharp\text{control pkts} \propto \frac{\epsilon\ cwnd_i}{rtt_i} \qquad \frac{1}{cwnd_i} \leq \epsilon \leq 1$$

$$p_i = \xi_p \times \frac{rtt_i^2}{\epsilon\ cwnd_i} \times \max(0, R(t) - \frac{cwnd_i}{rtt_i})$$

Conclusion

- RCP = Fast + unbounded worst case
- modified-RCP = not-so-Fast + bounded worst case
- oracle-RCP = Fast + bounded worst case