# Buffers in All-Optical Packet Switches

Vijay Sivaraman

University of New South Wales, Sydney, Australia



UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

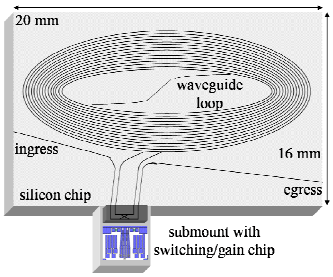
## Outline

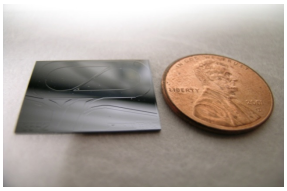
- Question: What if (optical) routers have (close to) zero buffers?
- Impact:
  - Performance: high loss, low throughput, …
  - Behavior of congestion control (TCP)
  - Interaction of open- and closed-loop traffic
- Mitigation techniques:
  - Traffic conditioning (shaping, pacing, …)
  - Forward Error Correction (FEC)
  - Routing, grooming, …
- Directions for future work?

UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

# The Problem

- Electronic buffers: RAM cheap but needs real-estate, consumes energy, complicates design
- Optical buffers: Signal degradation limits storage capacity, integrated optics for 10-20 packets?
  - 1250 Byte packet at 10 Gbps needs 1 usec of buffering



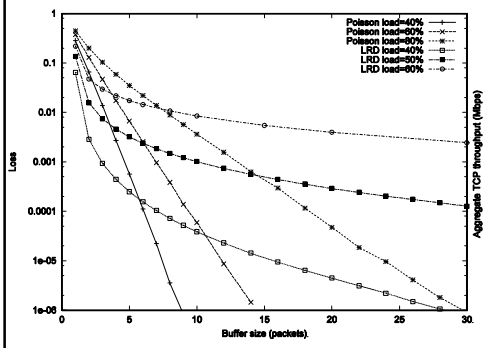



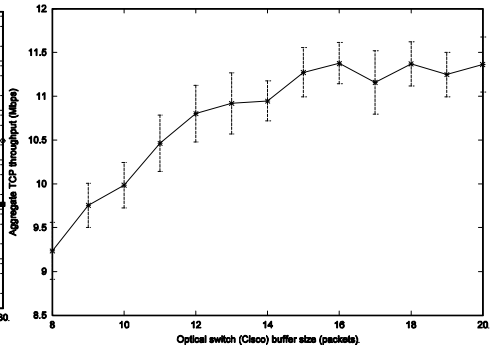
[Burmeister and Bowers, UCSB]

UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

# Impact on Performance

- Loss (open-loop traffic):
  - Load: 40 – 80%
- Throughput (TCP traffic):
  - Load: ~60%

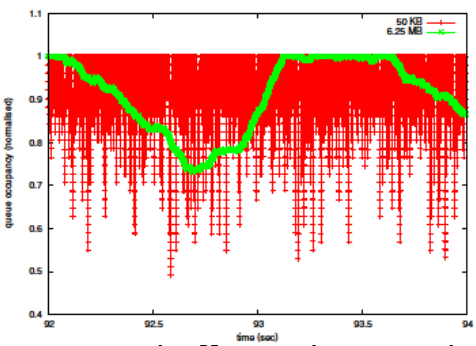


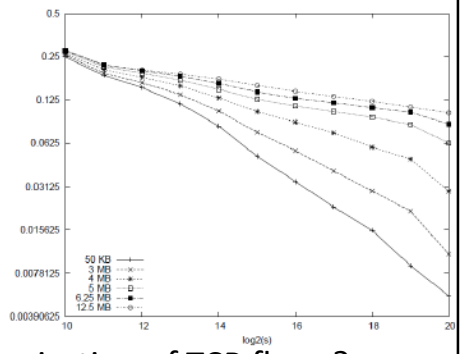




UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

# Nature of TCP Traffic

- How does TCP react to buffers in the network?
  - 2000 TCP flows sharing a bottleneck link



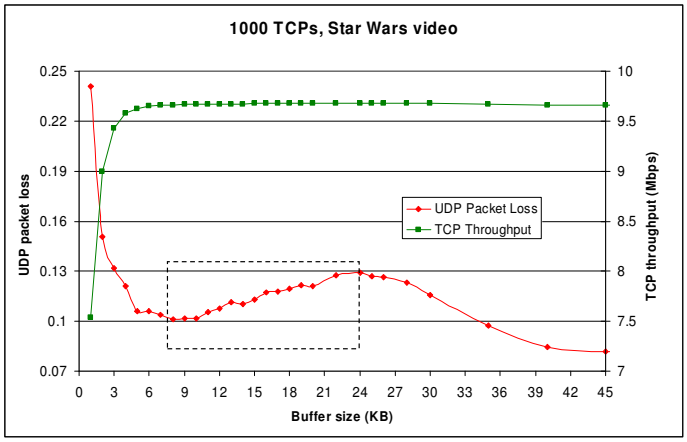




- Large buffers induce synchronization of TCP flows?
- Small buffers make aggregate TCP traffic Poisson?

UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

# Interaction of Open-Loop and TCP

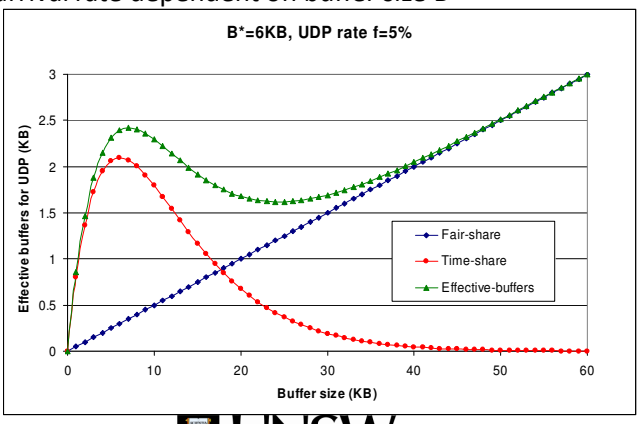
- Loss is non-monotonic for open-loop traffic:




UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

# Explaining Non-Monotonic Loss

- Effective buffers $B_{eff} = fB + (1-f)Be^{(-B/B^*)}$
  - "space-share" and "time-share" components
- Markov chain (M/M/1/B) based model
  - TCP arrival rate dependent on buffer size B




UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

# Mitigation Techniques

- Traffic conditioning
  - Shaping
  - Pacing
- Packet-level forward error correction (FEC)
- Load-balanced routing
- Traffic grooming, wavelength conversion, …

UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

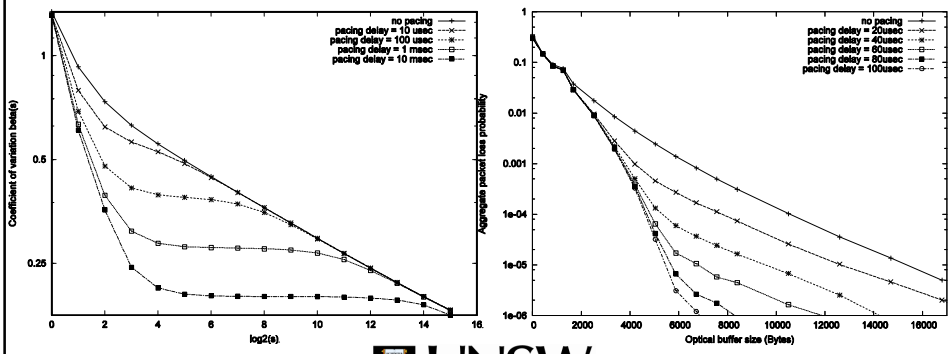
# Traffic Conditioning

- Host TCP pacing: window released over RTT
  - Requires all hosts to upgrade stack
- Edge shaping: release packets at set rate
  - Static rate: what is the best rate to use?
  - Dynamic rate: traffic bursty at sub-RTT timescales
- Edge pacing: release packets in smoothest way
  - subject to upper bound on delay
  - Locally optimal for bounded penalty

UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

# Edge Traffic Pacing

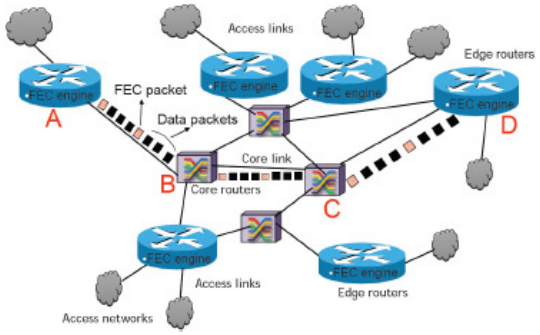
- Traffic burstiness
  - Drops at short time-scale
  - Reconverges at large time-scale
- Network loss (NSFNet)
  - Orders of magnitude lower
  - Small bounded delay penalty




UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

# Packet-Level FEC

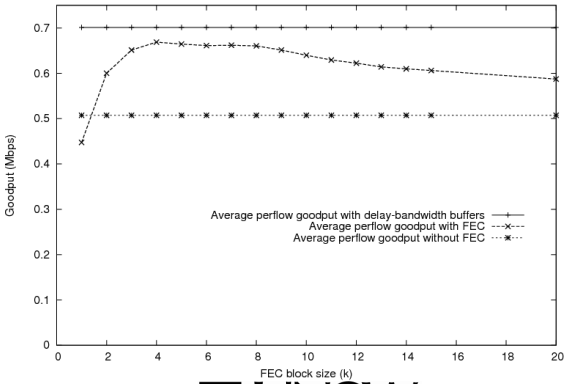
- Premise 1: Links are not heavily loaded
- Premise 2: Core losses are random, not bursty
- Simple FEC scheme:
  - Edge inserts one FEC packet for every *k* data packets




UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

# Efficacy of FEC

- Edge-to-edge loss $L_e = L_c\,[1-(1-L_c)^k]$
  - Core loss $L_c$ estimated using Poisson assumption
- Single-hop network
  - Optimum FEC strength maximizes per-flow TCP goodput




UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

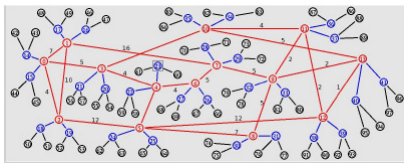
# FEC in Multi-Hop Networks

- Unfair for long-hop flows:
  - Additive core loss over hops rapidly degrades goodput
  - Need stronger FEC for longer-hop flows
- Global optimization problem
  - Minimize maximum edge-to-edge loss over all flows
  - Subject to integer FEC strength $k_i$ for each flow $i$
- Practical heuristic
  - Maximize fairness index
  - Flows of hop-length $h$ assigned FEC strength $k_h$
  - Brute-force search over space ($k_1$, $k_2$, …, $k_H$)

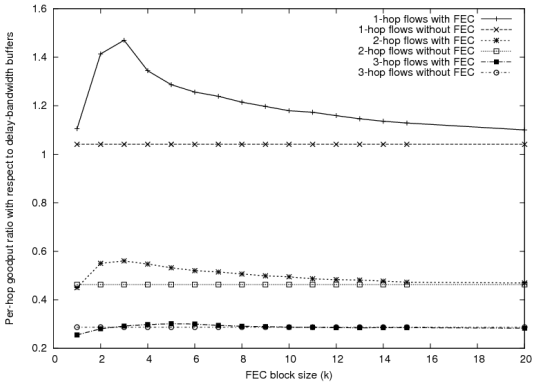
UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

# FEC on NSFNet Topology

- Optimized FEC uses $k_1$=19, $k_2$=4, $k_3$=2
  - achieves good Fairness Index FI=0.96



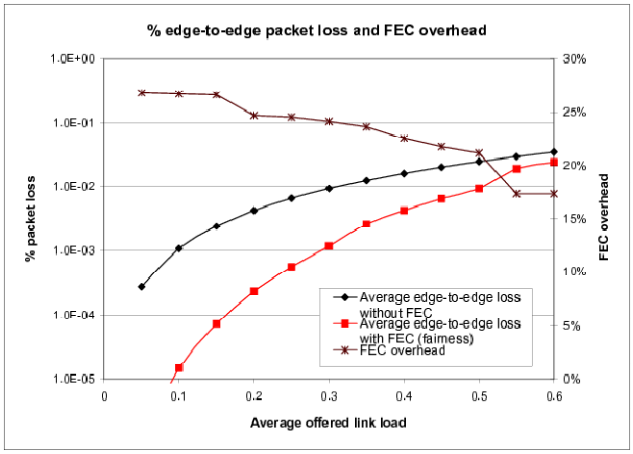
| Network setting | Average goodput (Mbps) | | | Fairness Index |
|---|---|---|---|---|
| | 1-hop flows | 2-hop flows | 3-hop flows | |
| No FEC | 1.571 | 0.667 | 0.391 | 0.78 |
| k = 3 for all flows | 2.219 | 0.807 | 0.397 | 0.70 |
| $k_1$ = 19, $k_2$ = 4, $k_3$ = 2 | 1.090 | 0.760 | 0.596 | 0.96 |
| delay - bandwidth buffers | 1.509 | 1.440 | 1.359 | 1 |




UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

# When is FEC Good?

- Low to moderate loads




UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA

# Directions for Future Work

- If small-buffer links are bottleneck:
  - Is TCP (agg./indiv.) performance acceptable?
  - New congestion control algorithms?
- If small-buffer links are non-bottleneck:
  - New/combo technique(s) for loss mitigation
  - Impact of parameters such as access link speeds
- Impact of buffers on energy consumption
- Analytical models for aggregate TCP traffic
  - Parameters: buffer size B, TCP flows N, …
- Experimental work to validate behaviour

UNSW THE UNIVERSITY OF NEW SOUTH WALES SYDNEY • AUSTRALIA