Radon: Network QoS

Andrew Shewmaker
Carlos Maltzahn Katia Obraczka
Scott Brandt Ivo Jimenez

UC Santa Cruz
Graduate Studies in Computer Science

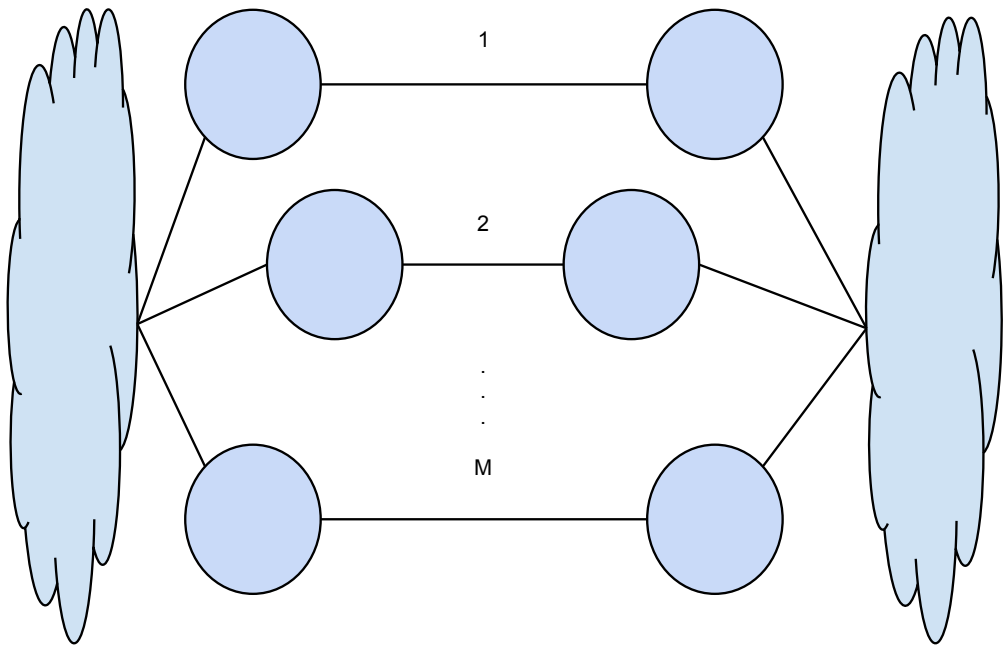
Oct. 22nd, 2013

This research was made possible by LANL, Cisco, and Google.

Radon Targets Highly Contended, Parallel Bottlenecks

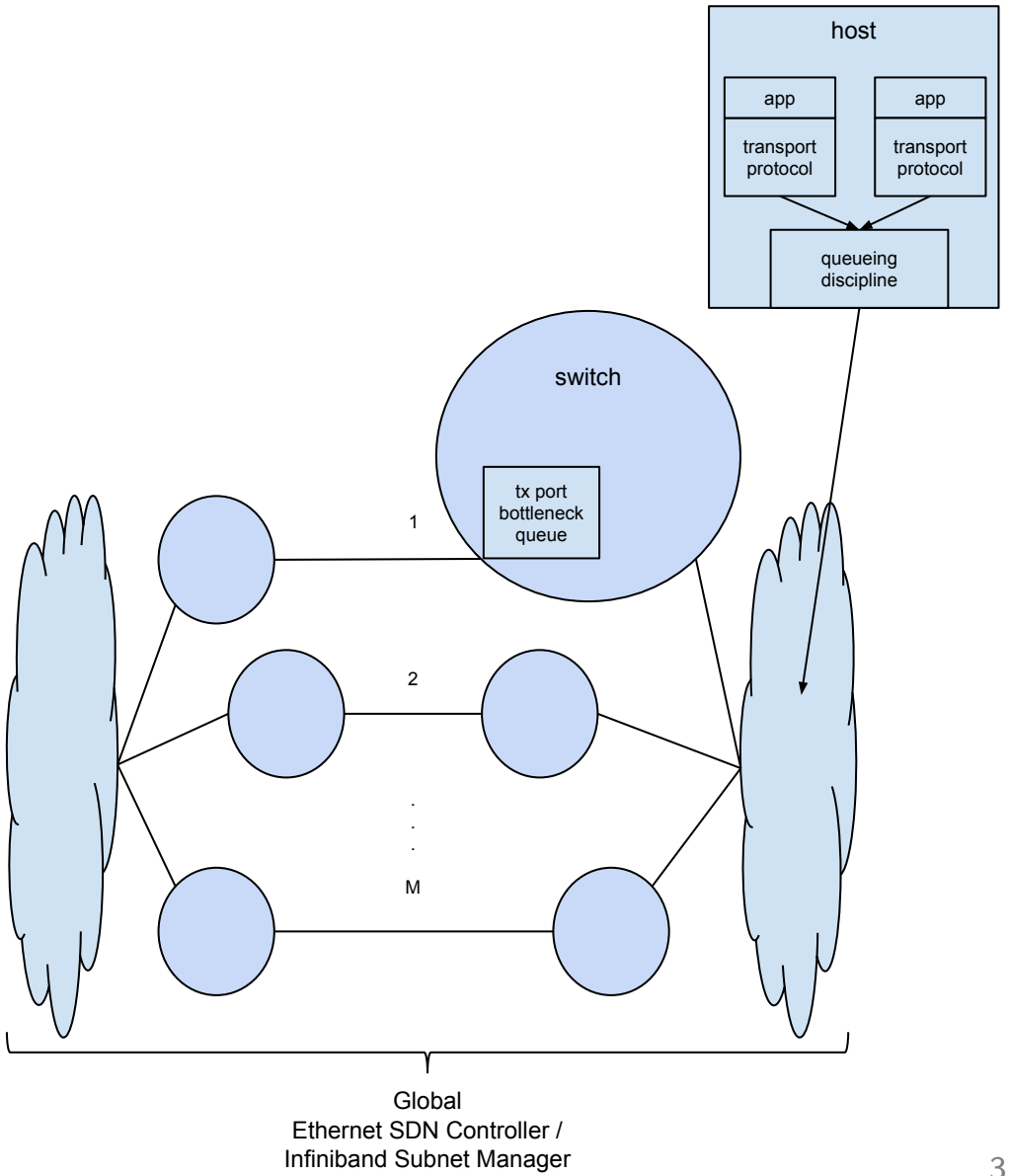
The purpose of Radon (Resource Allocation and Dispatching On Networks) is to maximize utilization of network gateways where bandwidth is precious.

- storage networks
- exascale I/O
- cloud SLAs
- datacenter provisioning
- WAN gateways



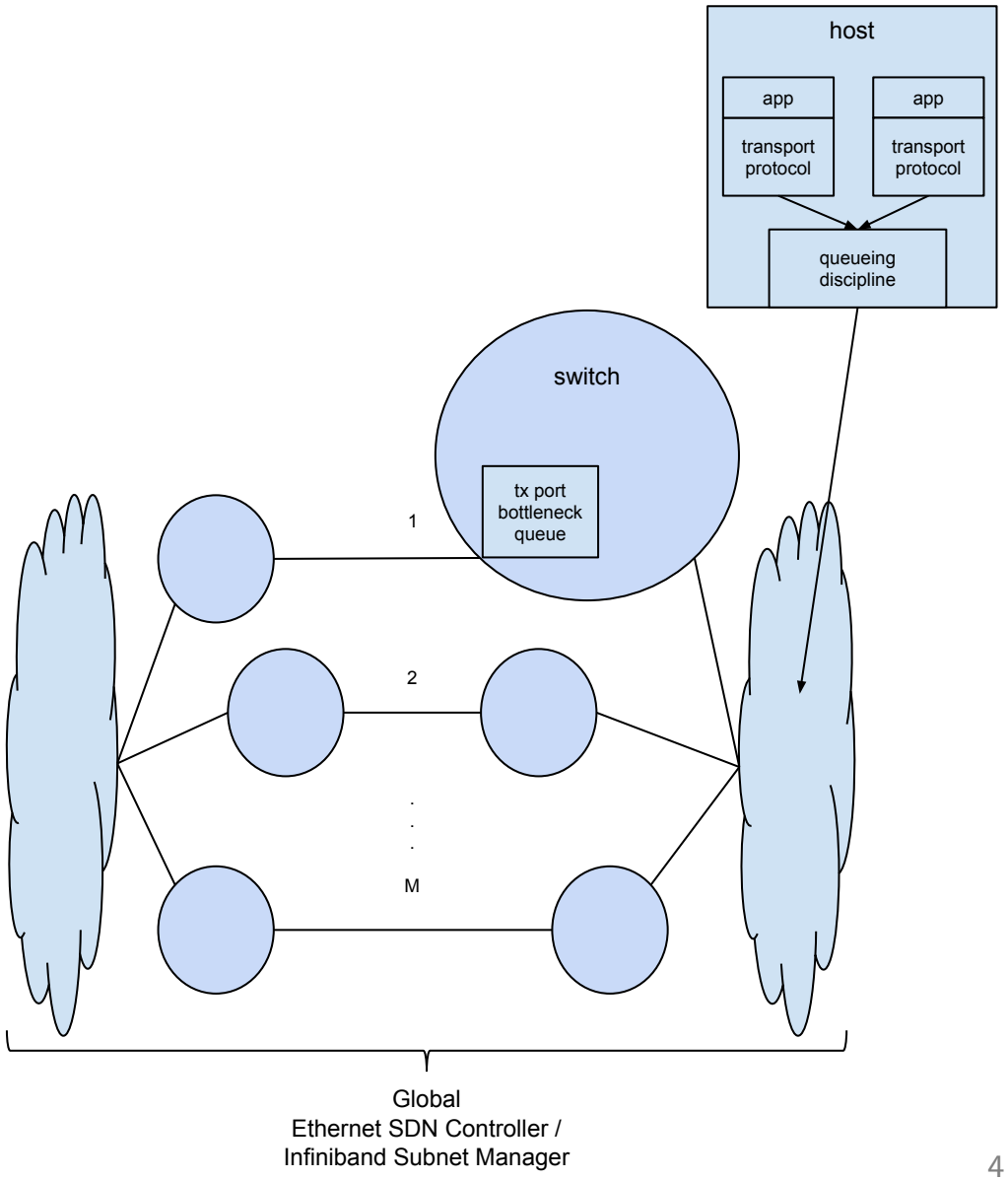
How Much QoS Enforcement At Each Layer?

- globally by a centralized traffic controller
- per flow by the host queueing disciplines
- per application by the transport layer
- per application by the application



Radon's Enforcement Choices

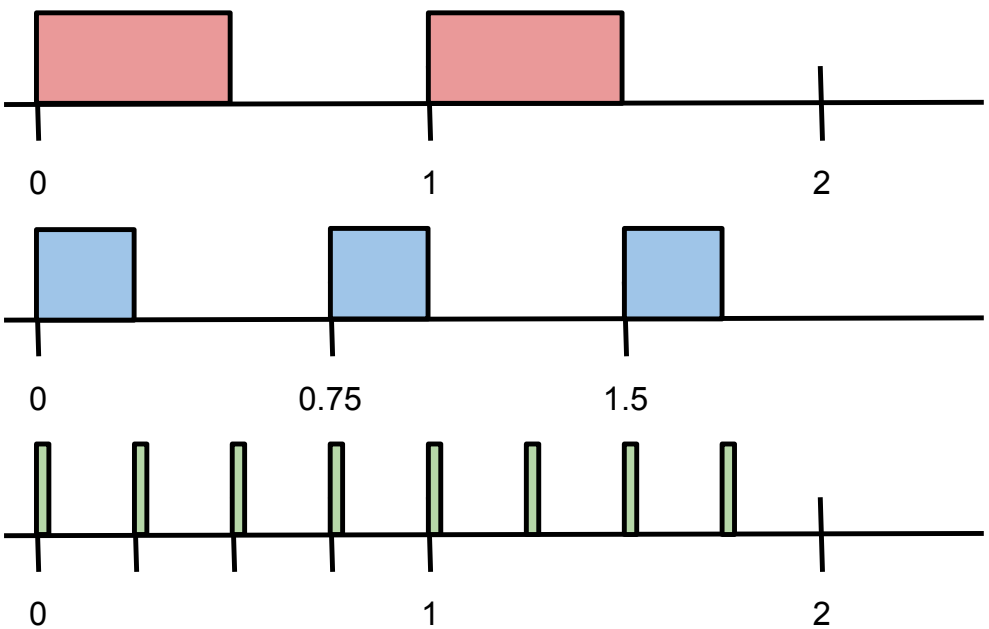
- global route scheduling by a central controller
- per flow rate limiting by the host
- per application rate limiting and congestion avoidance by the transport layer



Don't Use Priorities If You Know What You Want

- RAD (utilization, period) - how much? how often?
- enable arbitrarily fine-grained QoS
- keep meanings consistent in a dynamic environment
- reason across global, flow, and application layers

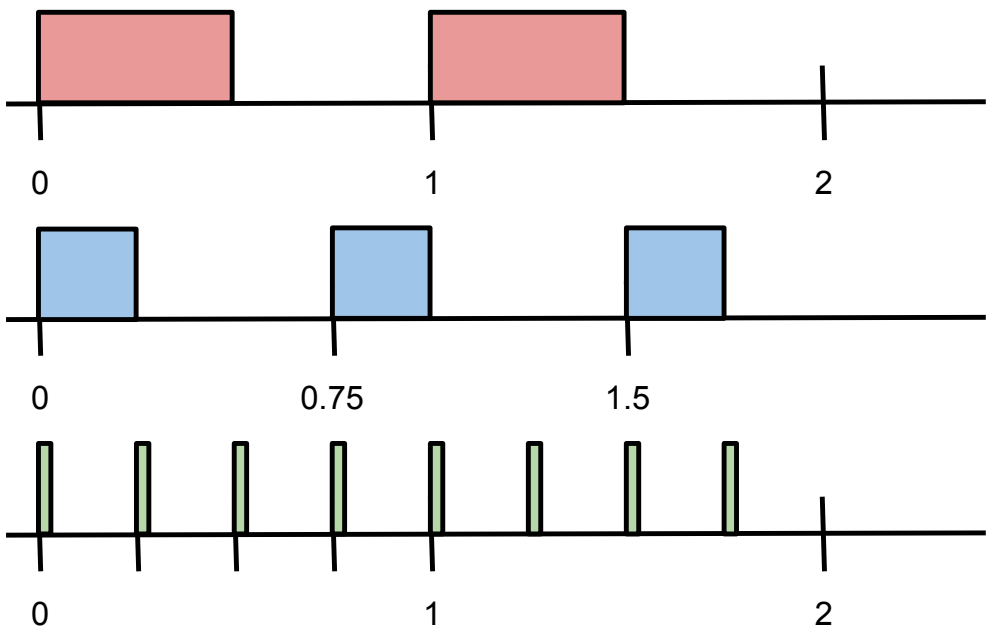
	Priority	RAD reservation
A	2?	(0.5, 1s)
B	3	(0.33, 0.75s)
C	1?	(0.14, 0.25s)



SDN Controller - Admission control

Admit if the sum of all utilizations does not exceed M, for M routes

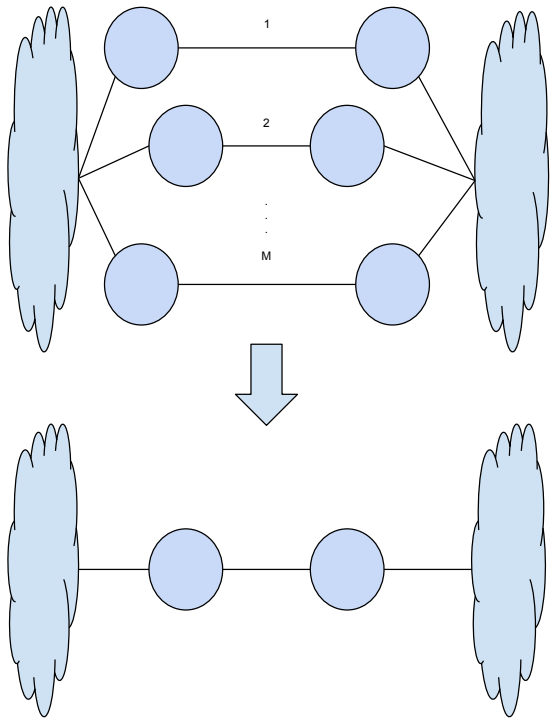
	Priority	RAD reservation
A	2?	(0.5, 1s)
B	3	(0.33, 0.75s)
C	1?	(0.14, 0.25s)



SDN Controller - Route Management

Reduction to ~~Uniprocessor~~ Unipath (RUN)

- Regnier, P., et. al. Multiprocessor scheduling by reduction to uniprocessor: an original optimal approach.
  RTSS 2011 (Best Paper Award)
- reduce N flows on M routes
- schedule single path with well-known algorithms
- transforms back to multipath schedule



Existing Host Queueing

- qdiscs provide strict rate limiting
  - rate-limited for any measured interval
  - not work conserving
- cpu/NUMA mappings
  - head-of-line blocking
- multiqueue scheduler avoids head-of-line blocking
  - round-robin is fair, but not QoS-aware
- network classifier cgroup tags packets for specific qdiscs

Proposed Host Queueing

Reduction to ~~Uniprocessor~~ Uniqueue (RUN)

- limits migrations between queues
- prevents head-of-line blocking
- short term: control groups + strict rate limiting qdiscs
- long term: control groups + flexible RAD-based qdisc

Transport Layer

Quotes from the 2013 ACM Symposium on Cloud Computing:

"Whatever the major protocol will be, it will be called TCP."

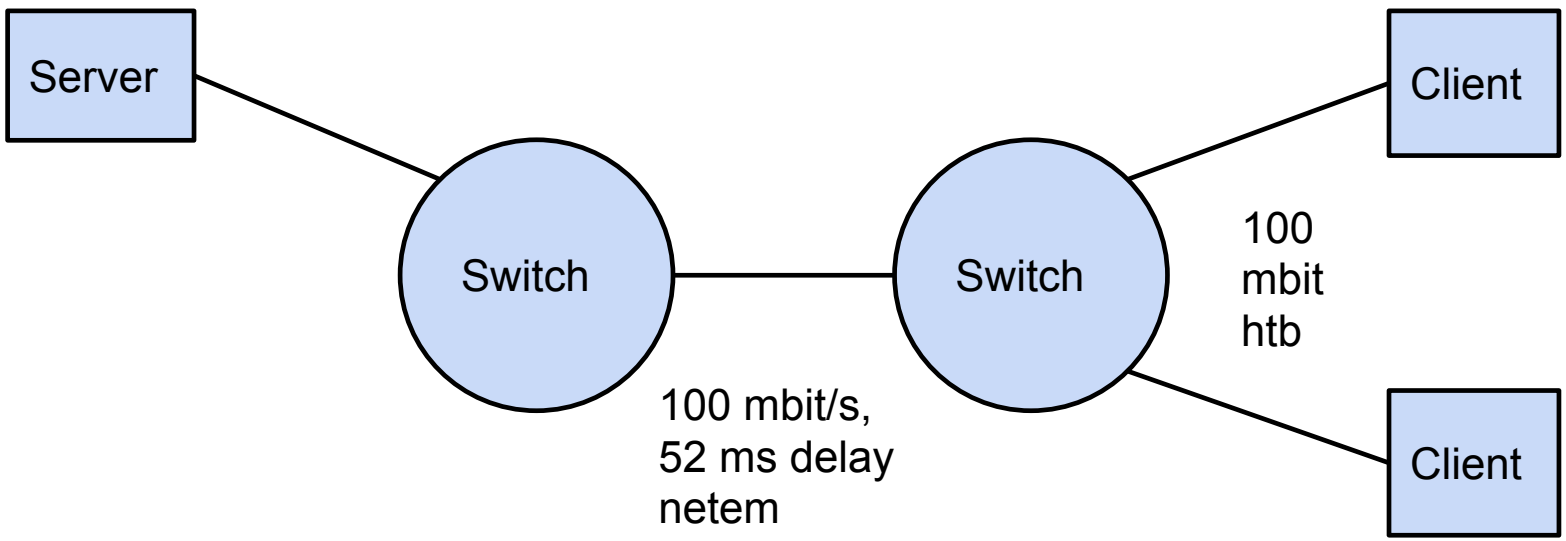
"Buffering ... don't do it."

TCP is here to stay, but it needs to be improved

- work conserving—
  we need something to use up unused bandwidth
- knows about congestion—
  it needs to know more
- needs to be aware of RAD reservations

CUBIC TCP Experiment - Setup

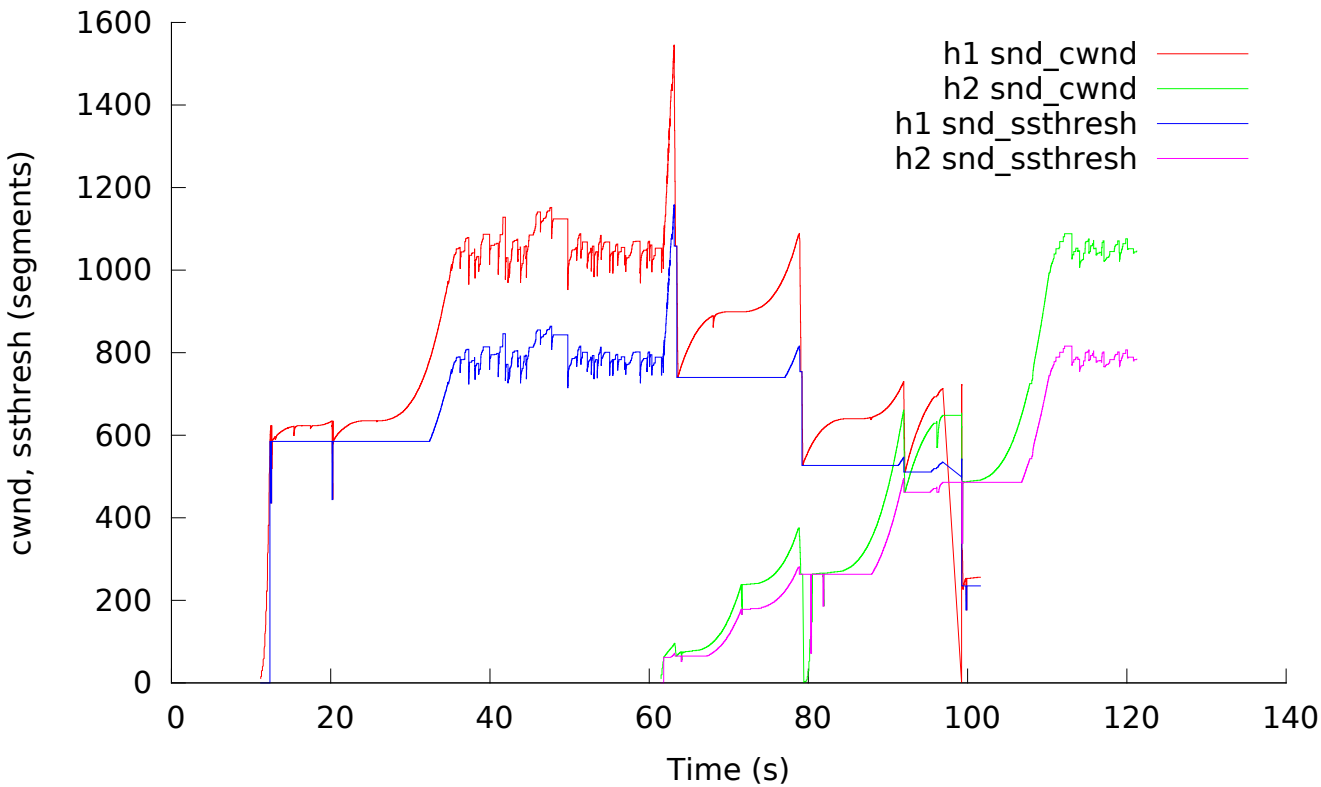
Mininet Setup



This setup replicates Figure 9 from the Hystart Technical report.

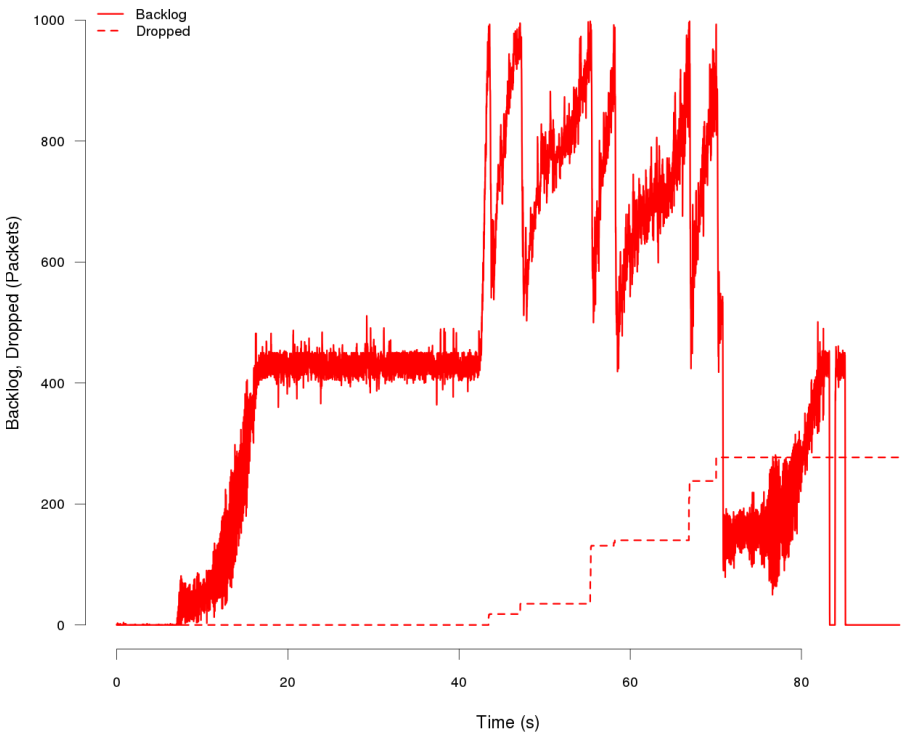
CUBIC TCP Experiment - Congestion Window

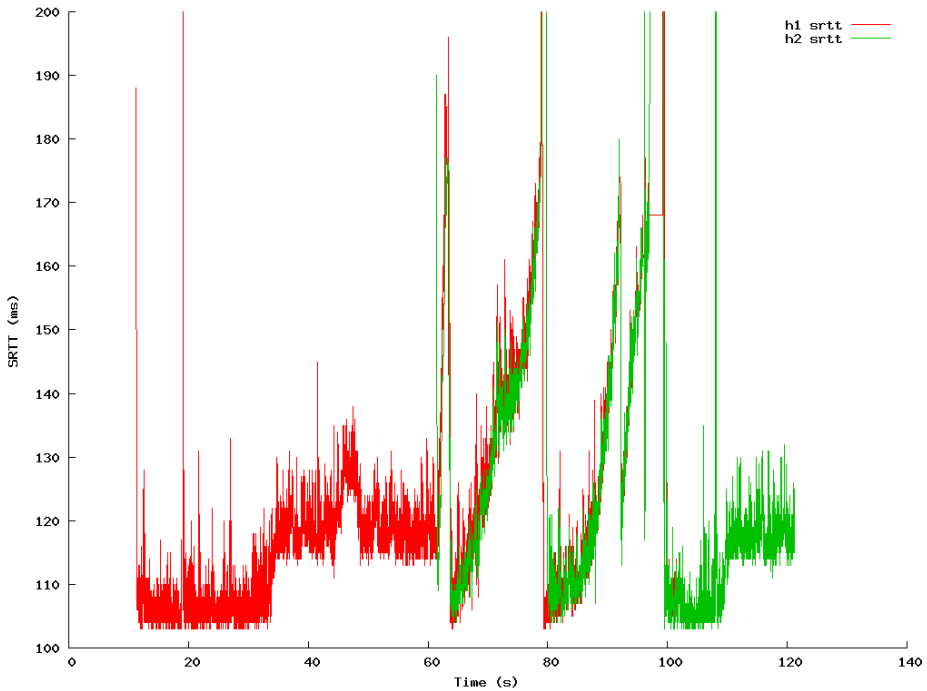
Typical cubic congestion responses.



CUBIC TCP - Bottleneck Queue & Round Trip Times

Standing/filling/dropping bottleneck queue → clear RTT trends





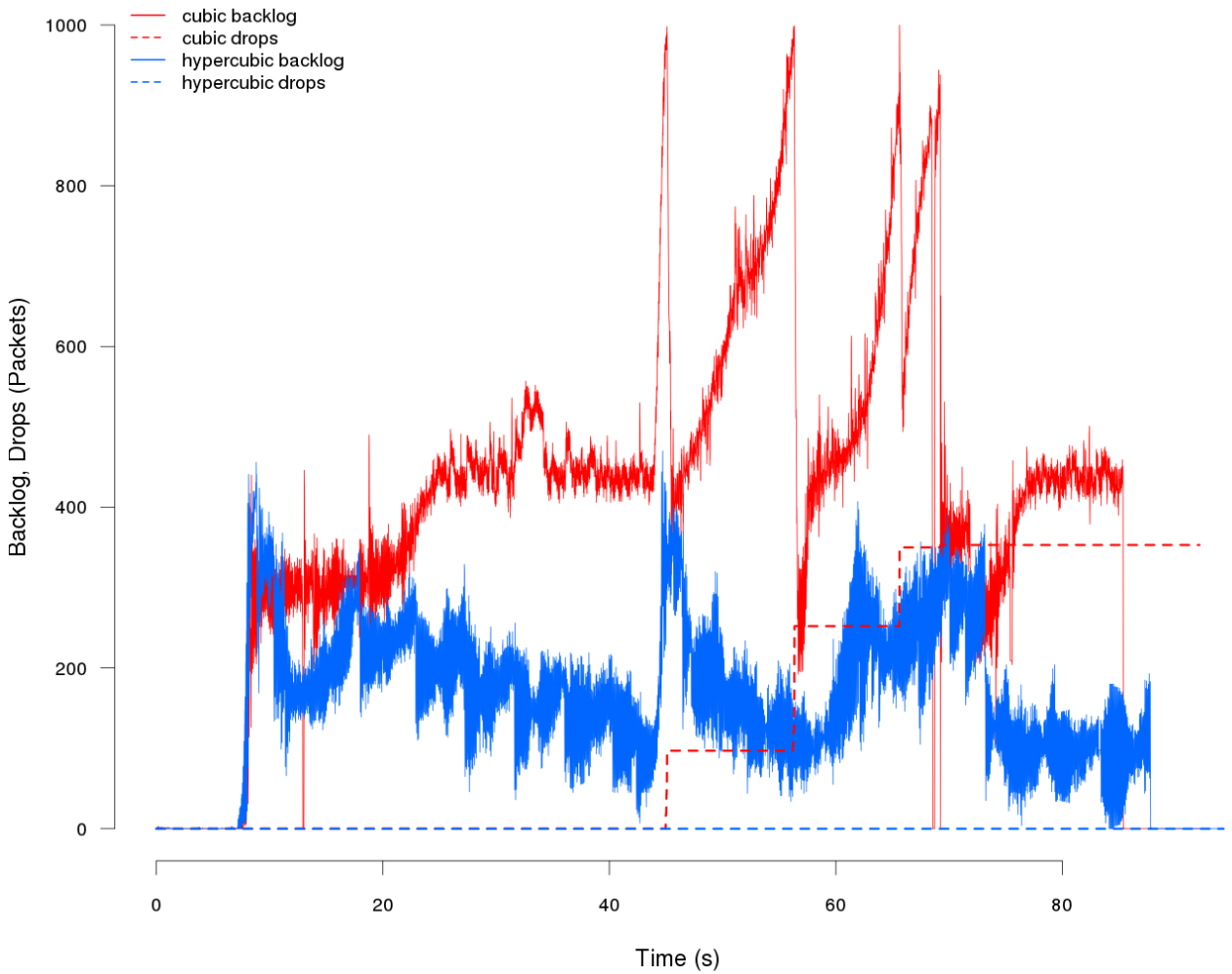
Next Generation TCP: Hypercubic

To Do:

- ✓ use Hystart detection as general congestion signal
- enhance Hystart detection
- add cgroup and RAD reservation awareness

CUBIC TCP Experiment

Hypercubic bottleneck queue < 500 and prevents packet loss



Going Forward

- Reduce To Unipath (RUN)
- Reduce To Uniqueue (RUN)
- fix TCP (Run! Just kidding.)

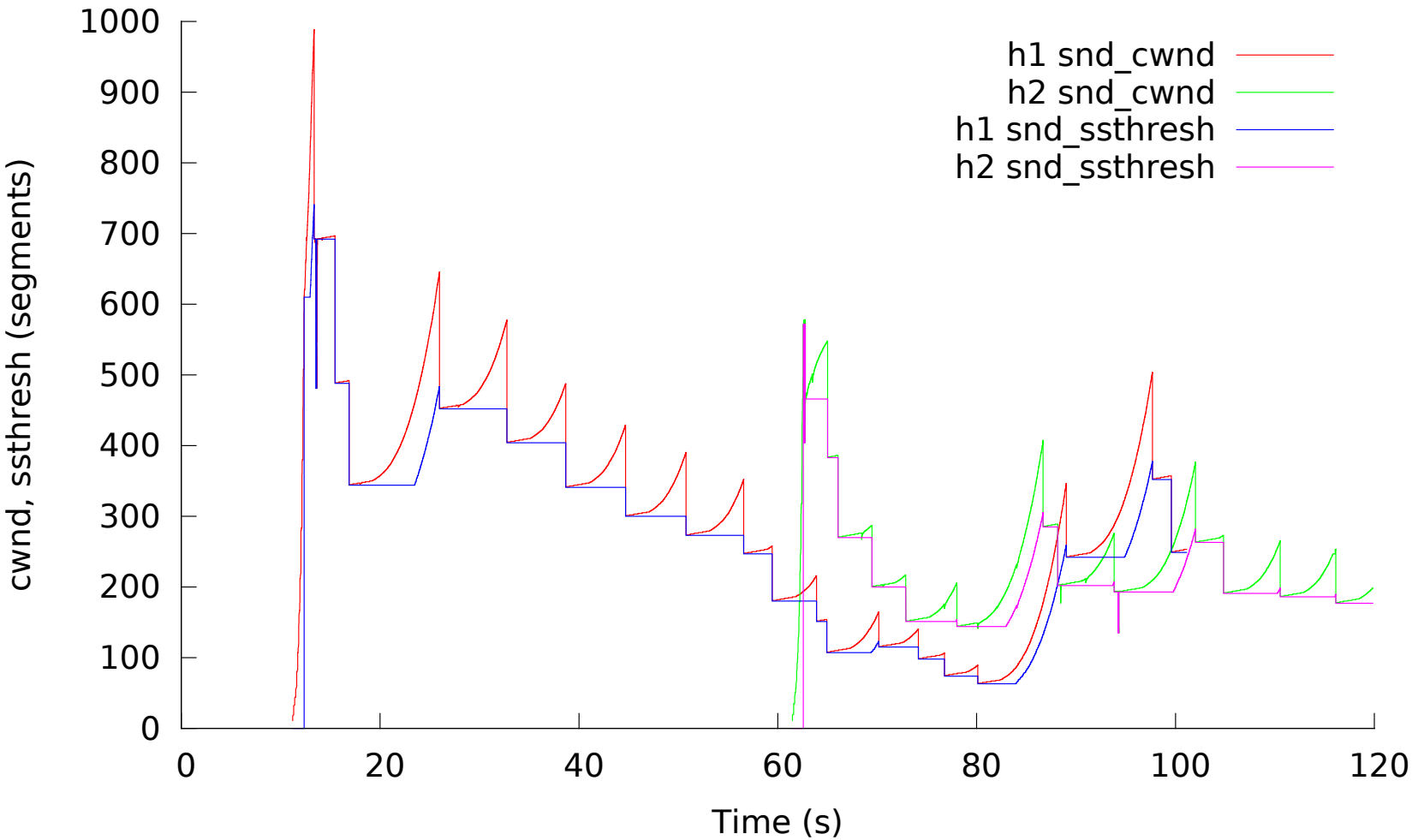
I'm looking for partners willing to try these ideas on a large scale testbed.

Thanks!

- more on RAD
  Dynamic Integrated Scheduling of Hard Real-Time, Soft Real-Time and Non-Real-Time Processes
  http://systems.soe.ucsc.edu/node/88
- original RUN paper by Regnier, P., et. al.
  (RTSS11 Best Paper Award)
  Multiprocessor scheduling by reduction to uniprocessor: an original optimal approach.
  http://systems.soe.ucsc.edu/node/579
- **Run, Fatboy, Run: Applying the Reduction to Uniprocessor Algorithm to Other Wide Resources**
  tech report

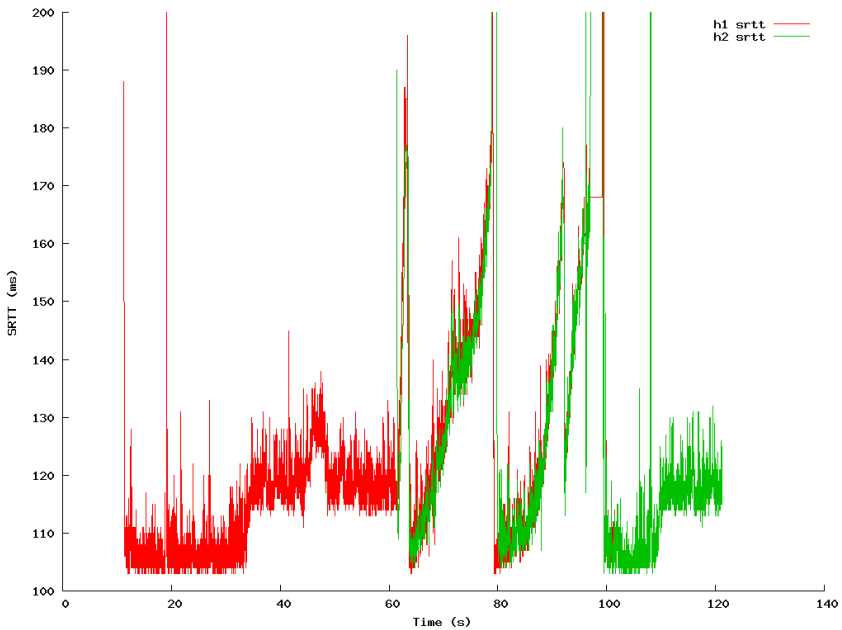
Backup Slides

Hypercubic Congestion Window



Round Trip Times Comparison

CUBIC



Hypercubic