Fastpass

A Centralized “Zero-Queue” Datacenter Network

Ideal datacenter properties

- Burst control
  — memcached, Fine-grained RTO

- Low tail latency
  — pFabric, HULL, DCTCP, D3, Orchestra

- Multiple app/user objectives
  — Hedera, SWAN, MATE, DARD, VL2, …

How to satisfy all these properties simultaneously?

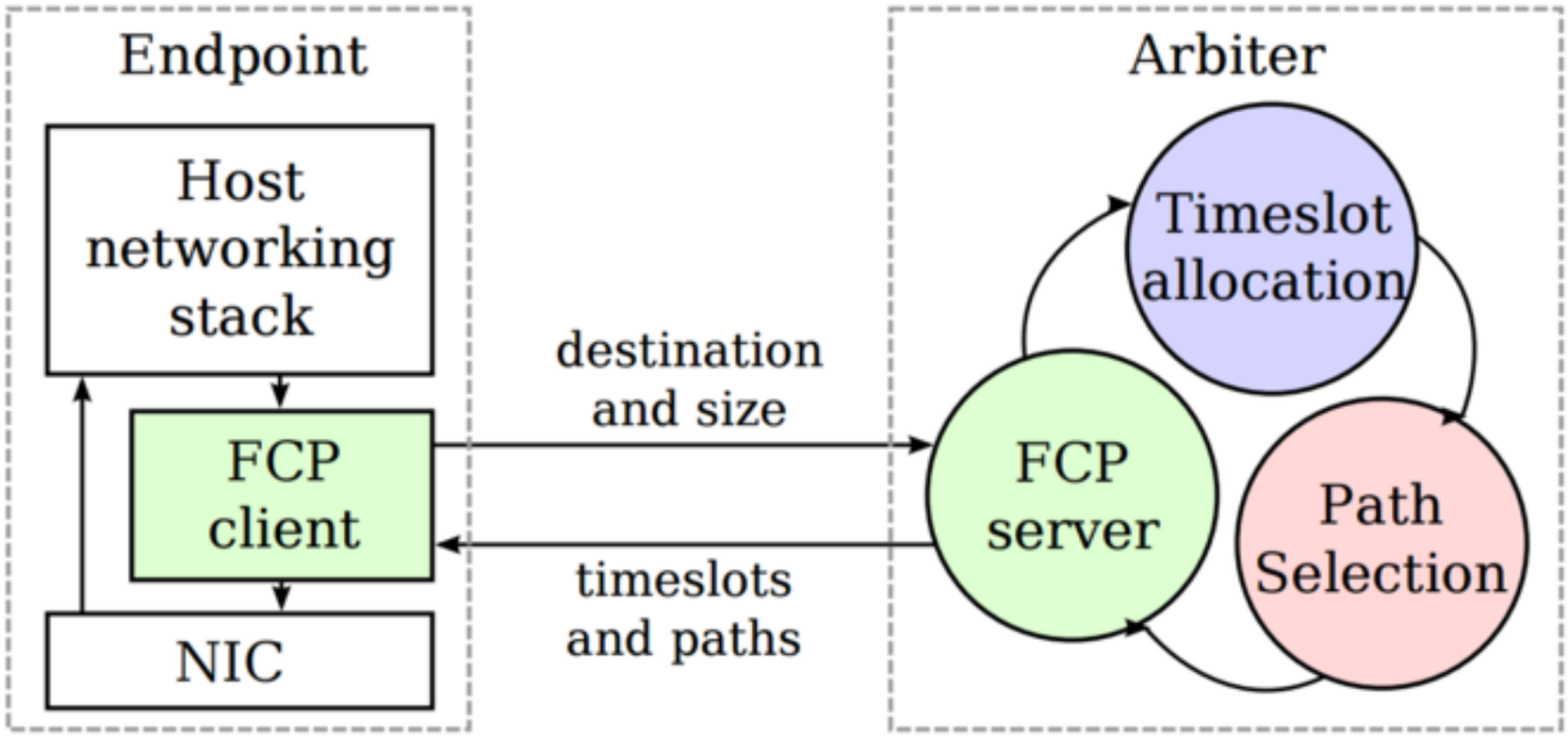
How to design a network with…

- “Zero” network queues
- High utilization
- Multiple resource allocation objectives

Control each packet's timing and path using a centralized arbiter

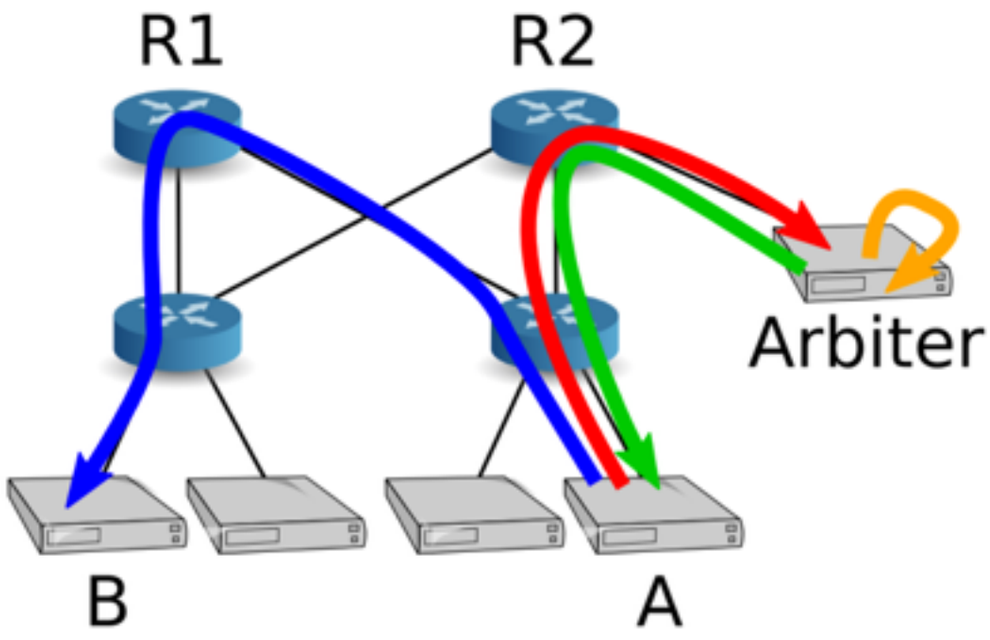
Traditional	Flow control	Congestion control	Update routing tables	Scheduling and queue management	Packet forwarding
SDN	Flow control	Congestion control	Update routing tables	Scheduling and queue management	Packet forwarding
Fastpass	Flow control	Congestion control	Per-packet path selection	Scheduling and queue management	Packet forwarding
	Endpoint		Centralized		Switch

Architecture

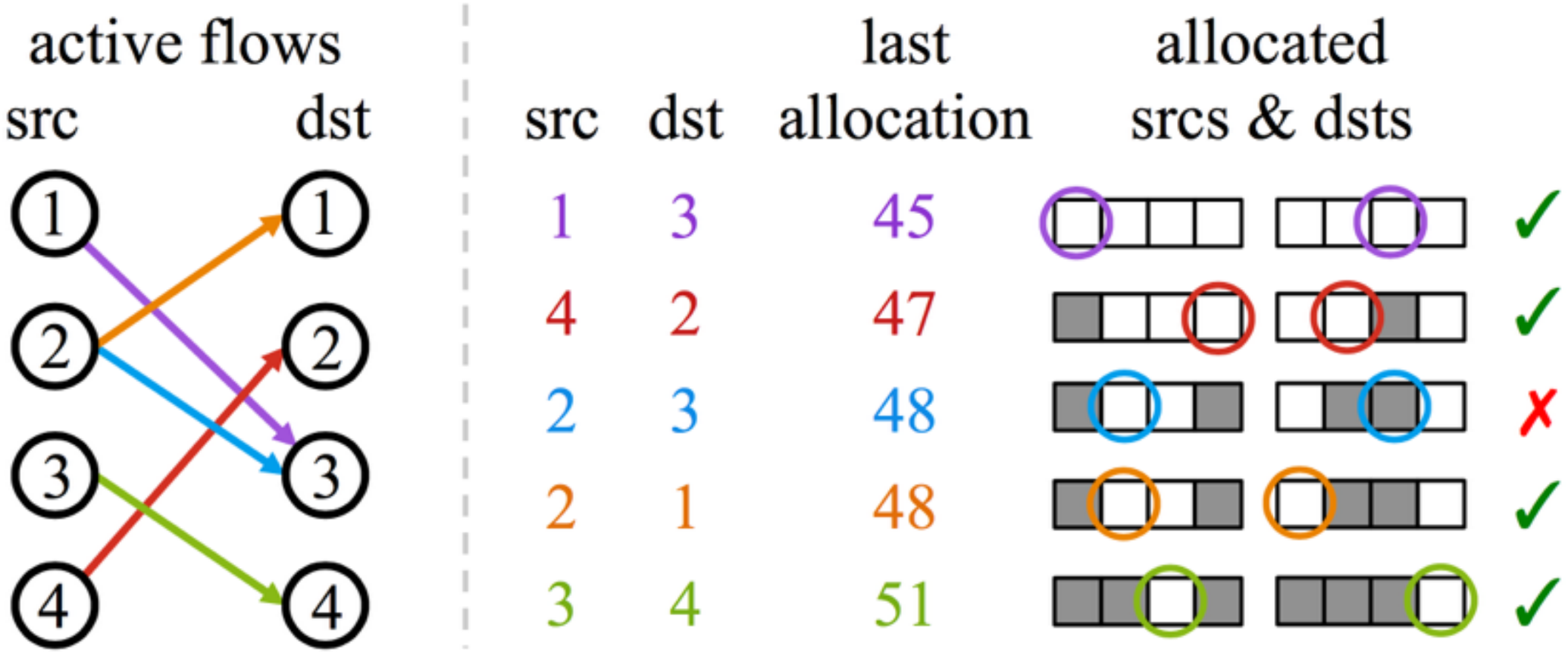


Example: Packet from A to B

5µs	A → Arbiter	"A has 1 packet for B"
1-20µs	Arbiter	timeslot allocation & path selection
15µs	Arbiter → A	"@t=107: A → B through R1"
no queuing	A → B	sends data

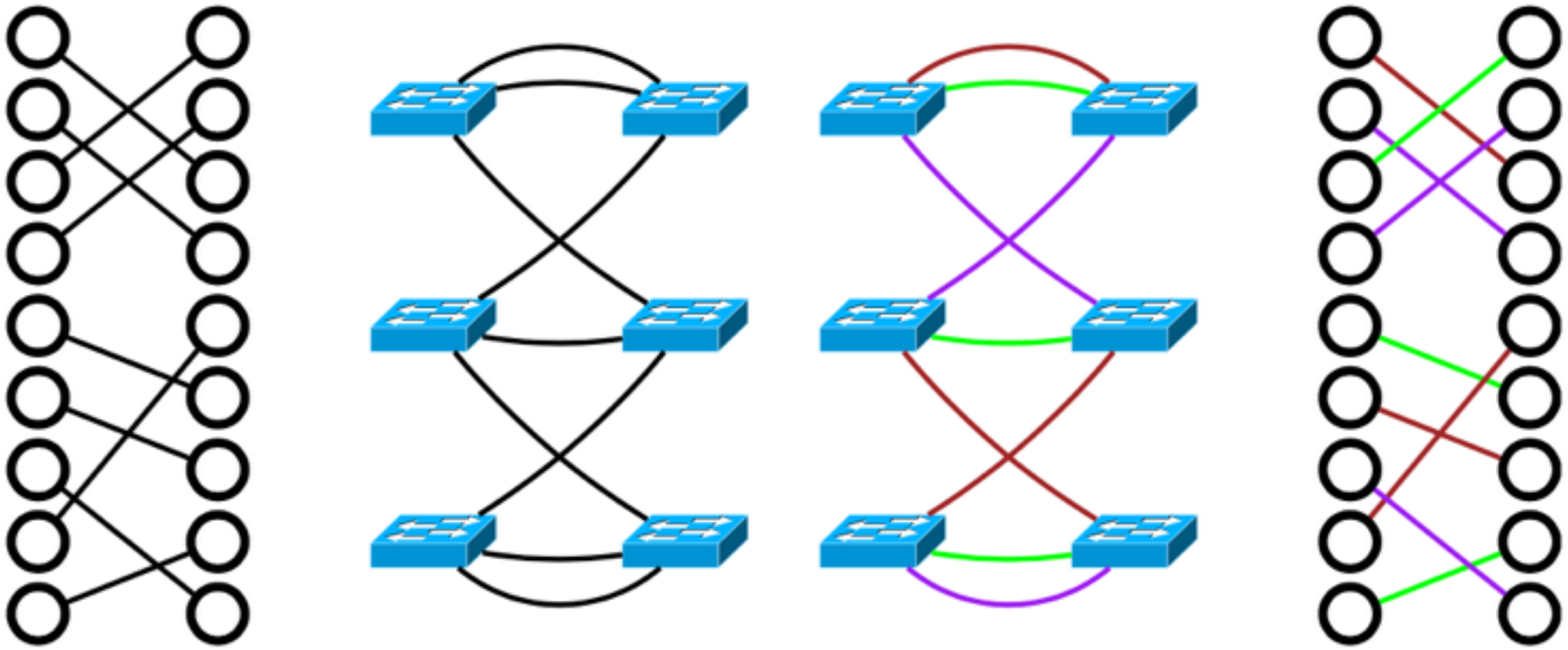


Timeslot Allocation



Ordering of requests used to implement policies.
E.g. LRU for max-min fairness,
lowest remaining MTUs for min-FCT

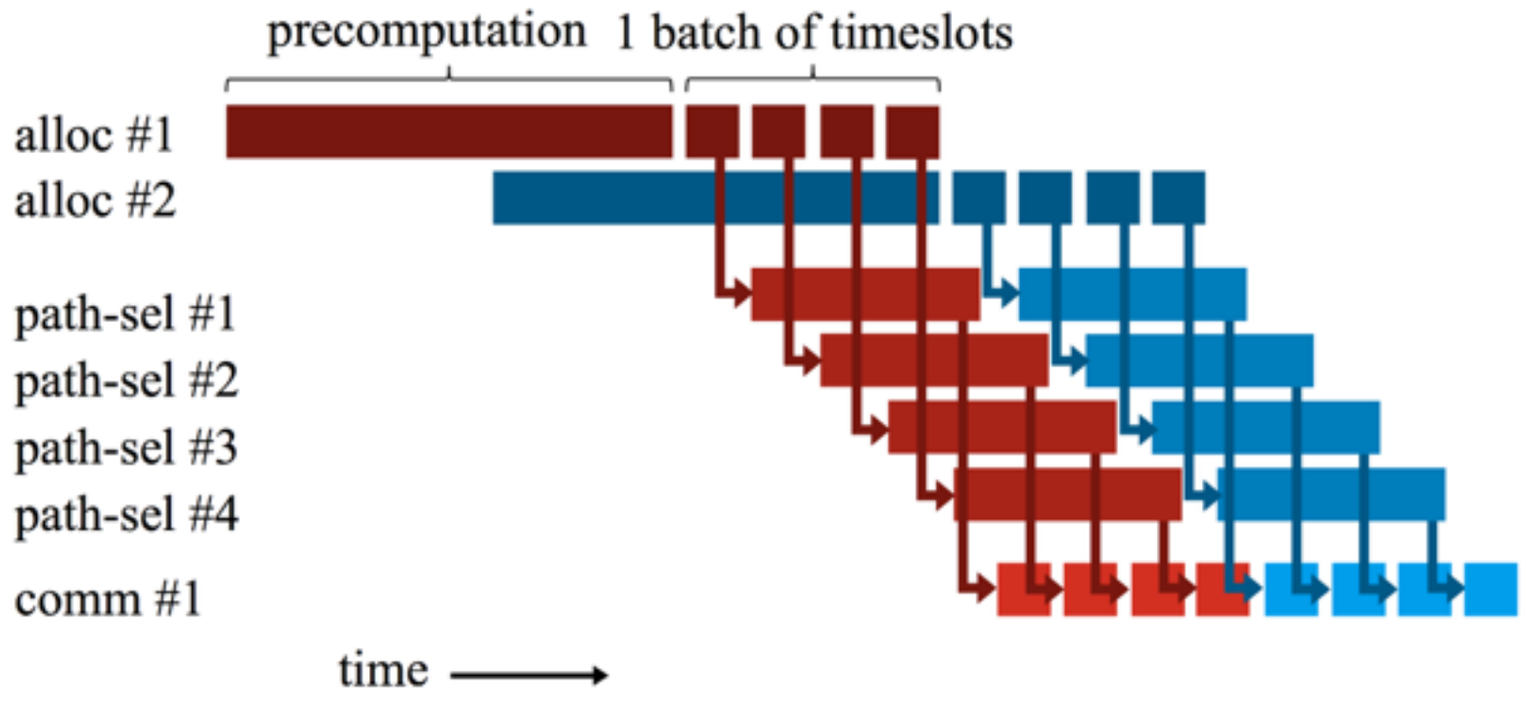
Path selection



Use edge coloring, each color denotes a path

Implementation

- Pipelined execution of tasks over multiple cores



- Clock synchronization using PTP (achieves sub microsecond synchronization)

- Client timing using hrtimers (microsecond scale precision)

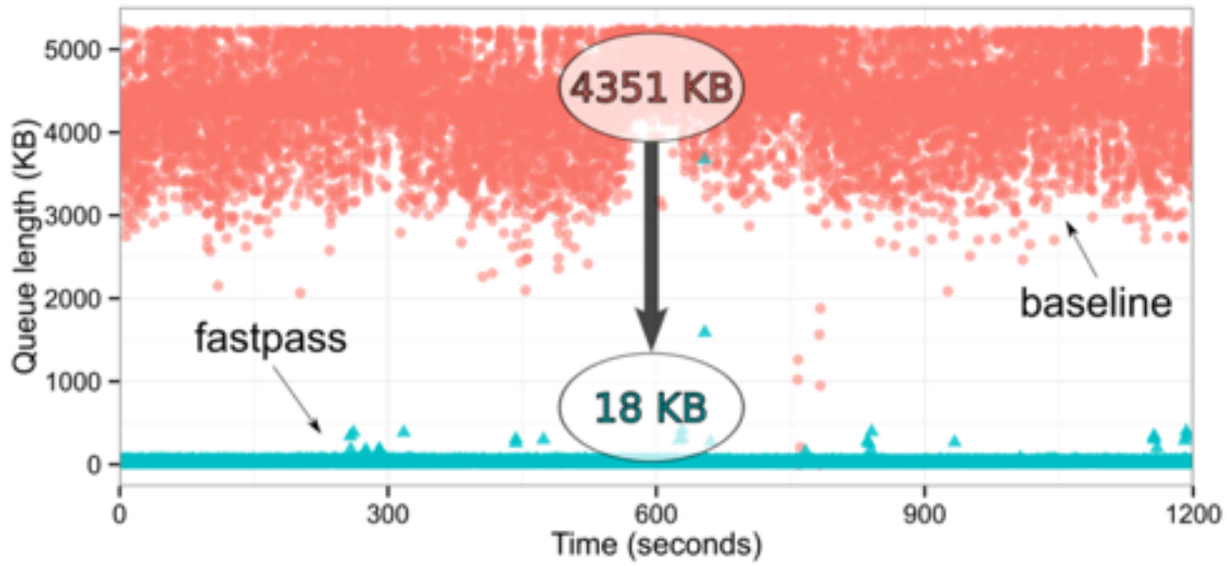
Fault tolerance

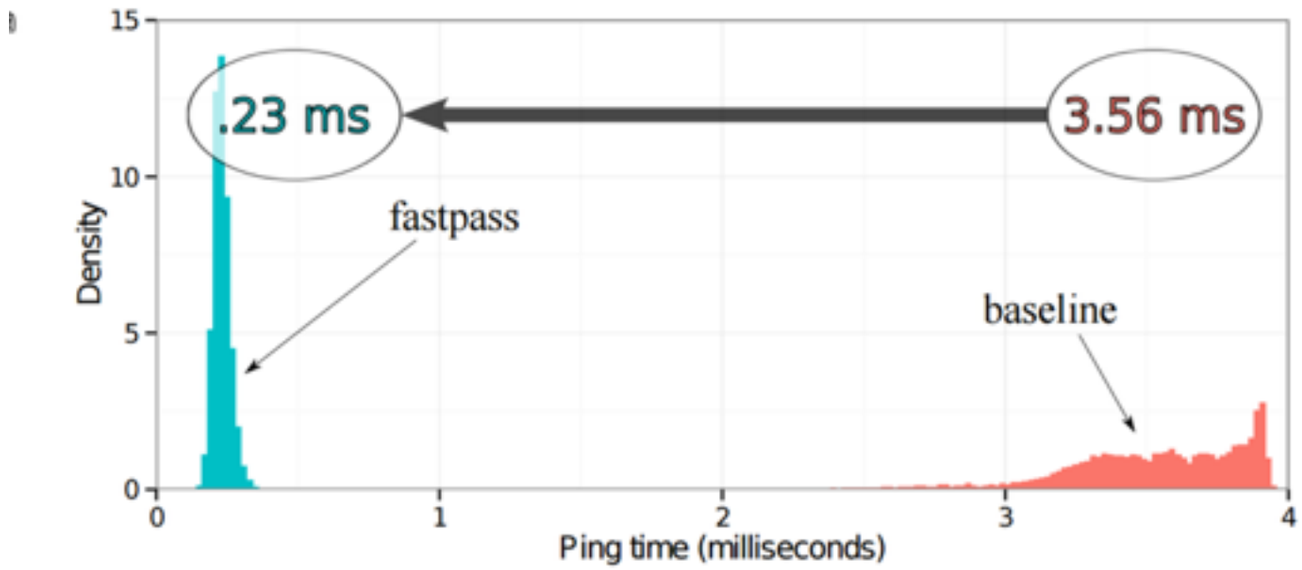
- Arbiter failures
  - — hot backups

- Network failures
  - — packet loss reports

Results:

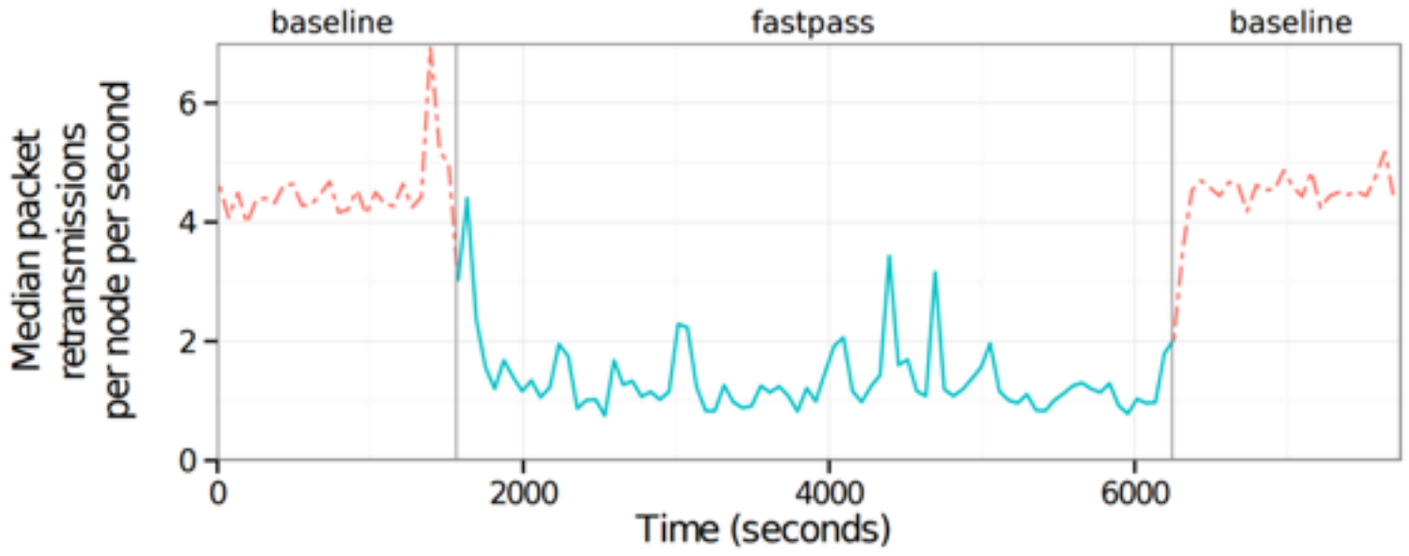
Smaller queues, lower RTTs
with iperf and pings





Results:

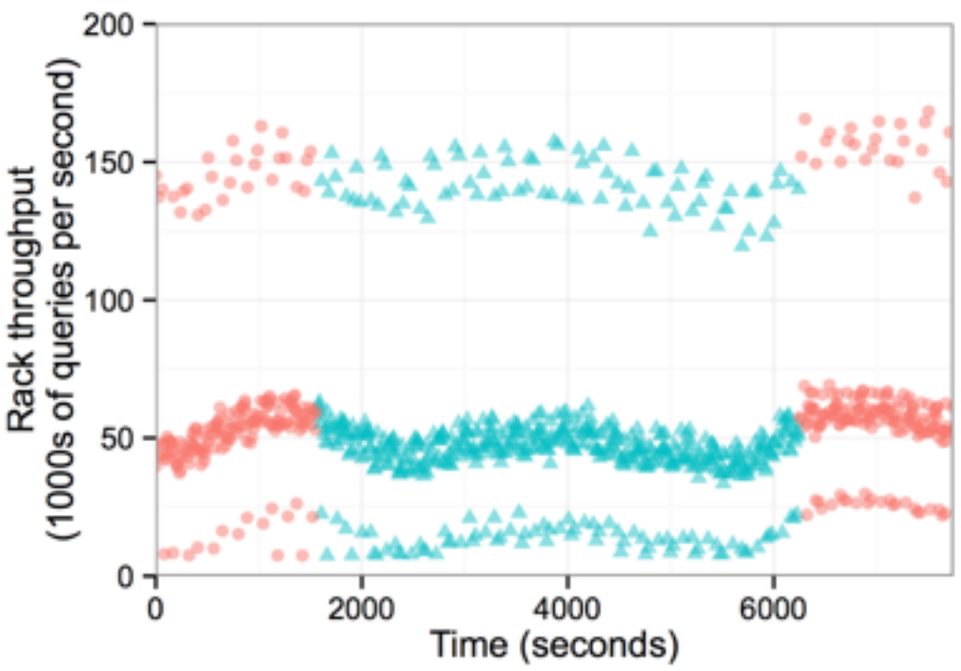
Lesser retransmissions in production…

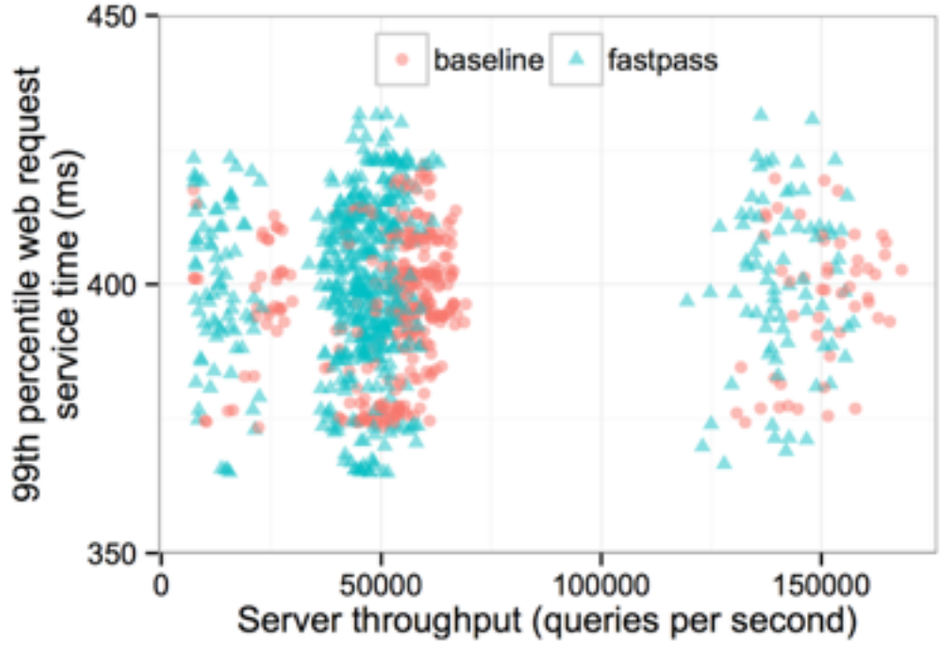


Each server: ~50k QPS

Results:

…But latency and throughput profiles were barely different





Issues

- Not really zero-queue — simply relocated to the endpoints and arbiter
- How to scale?
  - — Several arbiters would need to cooperate
  - — Precise time synchronization required
- How useful is Fastpass in practice?
  - — End-to-end delay at varying load
  - — Experimental setup had only single ToR