Fastpass

A Centralized "Zero-Queue" Datacenter Network

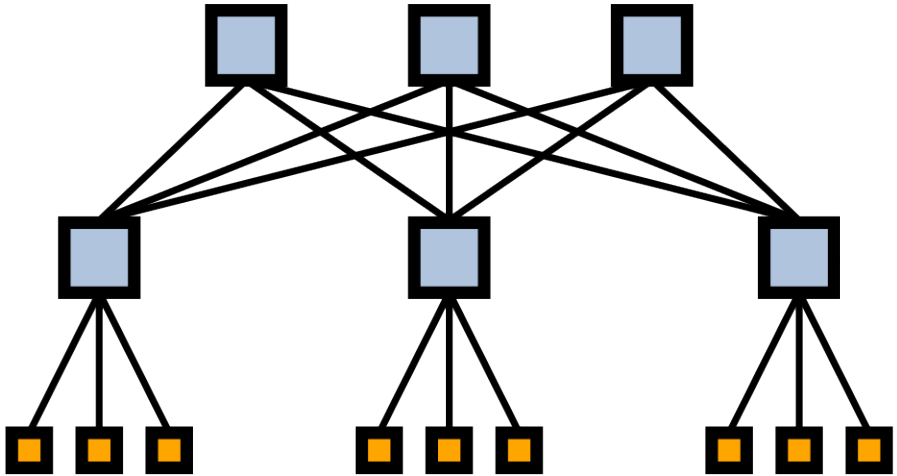
Road Network vs. Rail Road Network

Topology

Core

ToR

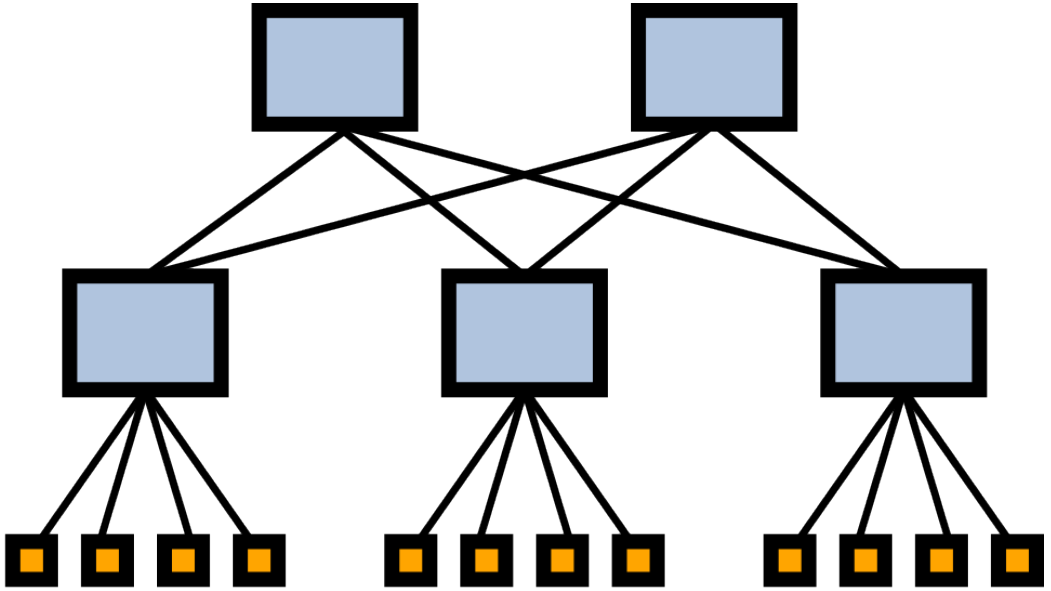
Endpoints



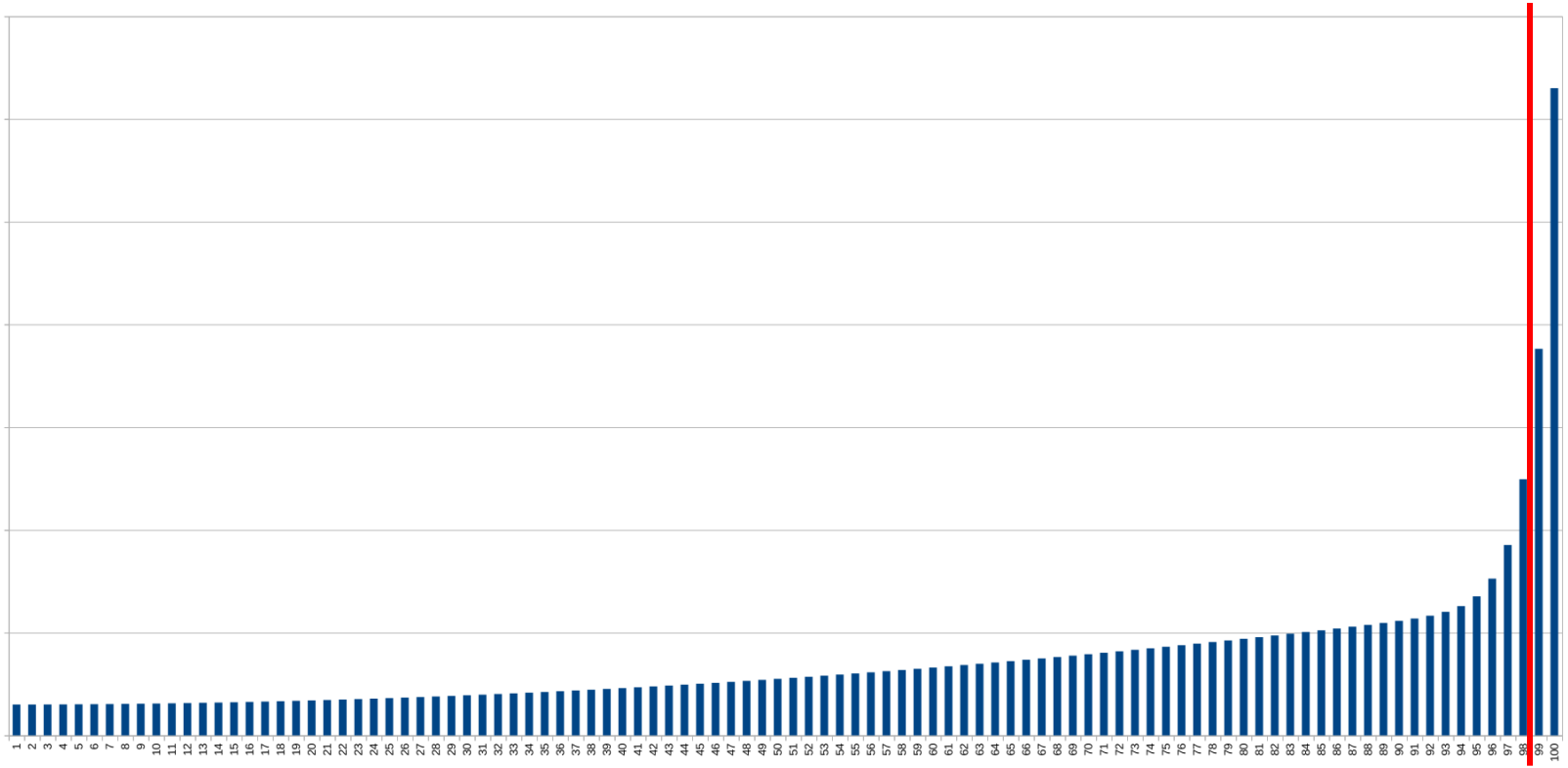
Current network architectures

Routing decision

Congestion control



99th Percentile



99th Percentile

100 billion hits per day

≈ 1 billion hits are in the 99th percentile

https://www.facebook.com/notes/facebook-engineering/scaling-facebook-to-500-million-users-and-beyond/409881258919

99th Percentile

$$0.99^{10} = 90.4\%$$

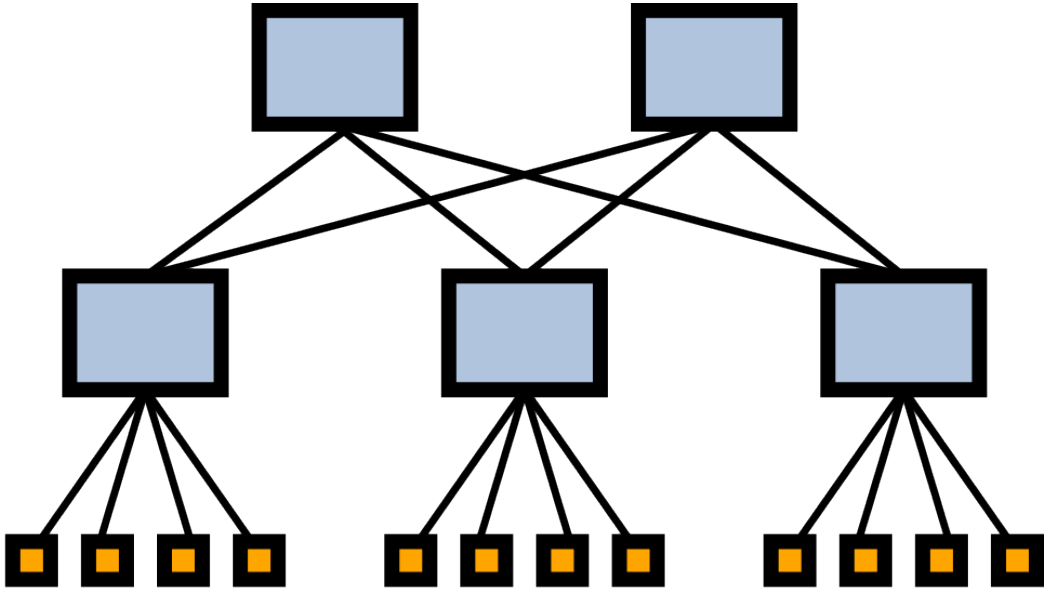
$$0.99^{20} = 81.7\%$$

$$0.99^{50} = 60.5\%$$

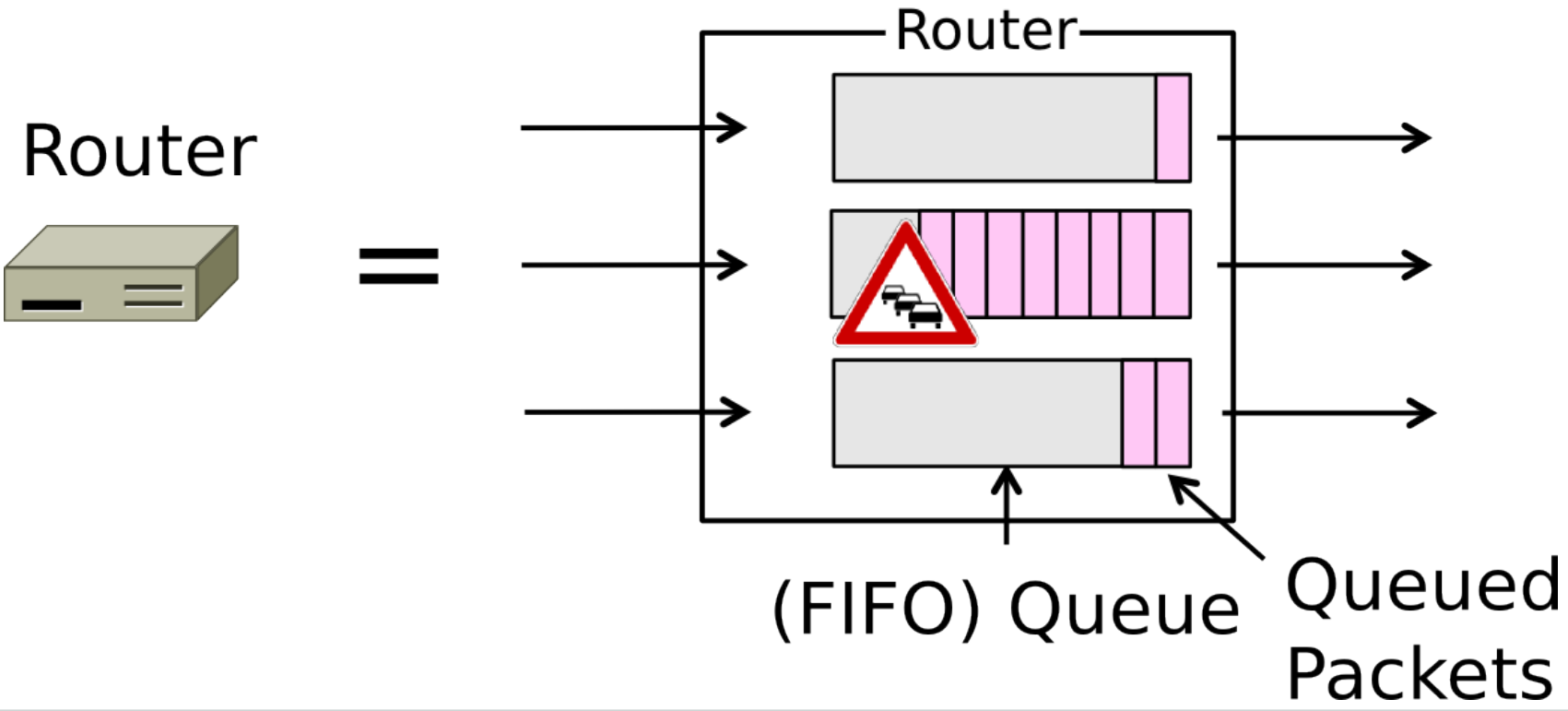
Current network architectures

Routing decision

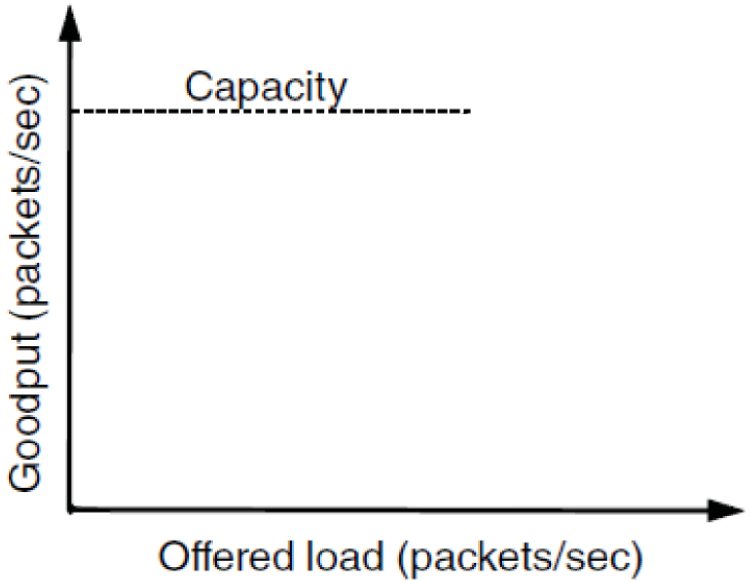
Congestion control

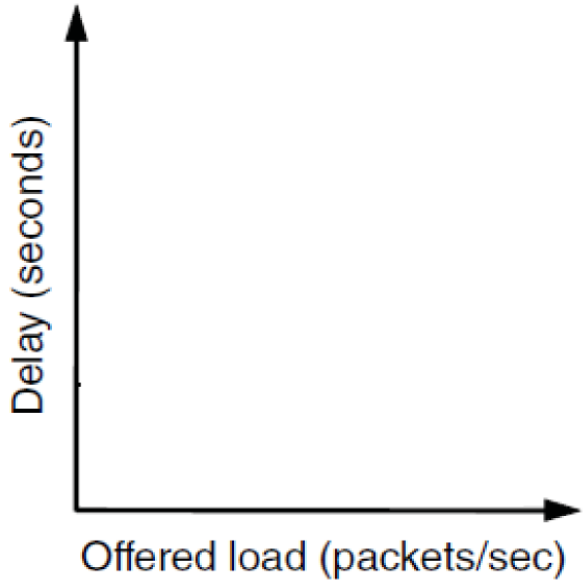


Queueing

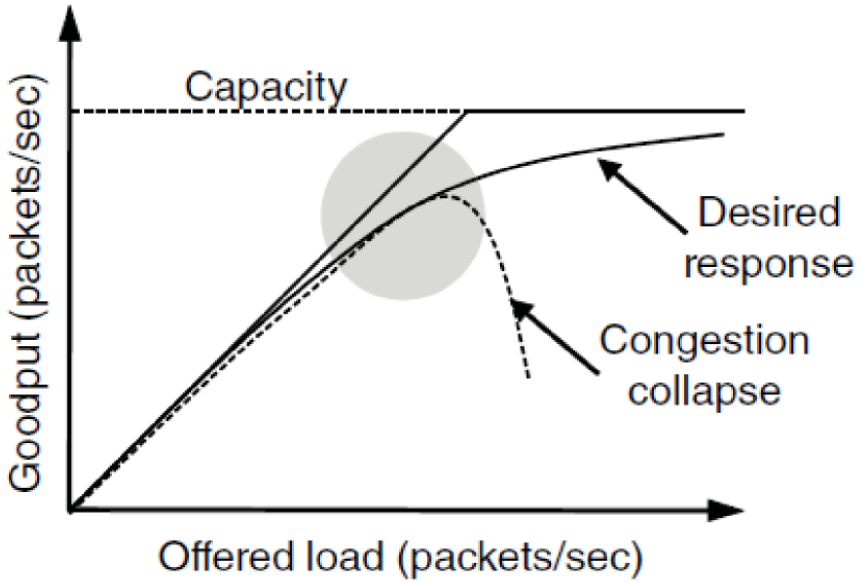


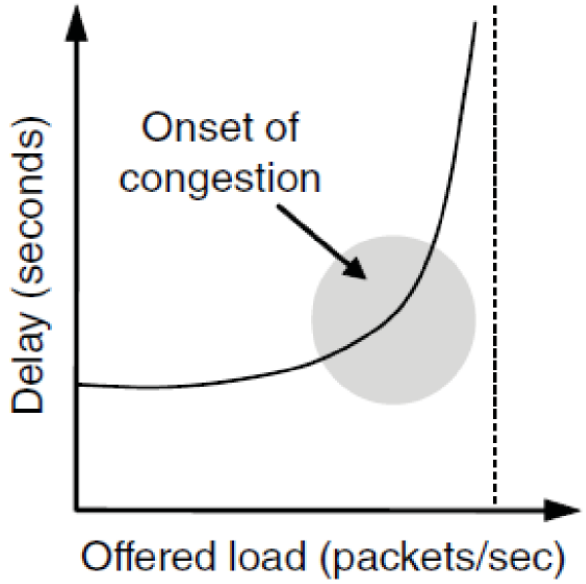
Congestion


Capacity
Goodput (packets/sec)
Offered load (packets/sec)

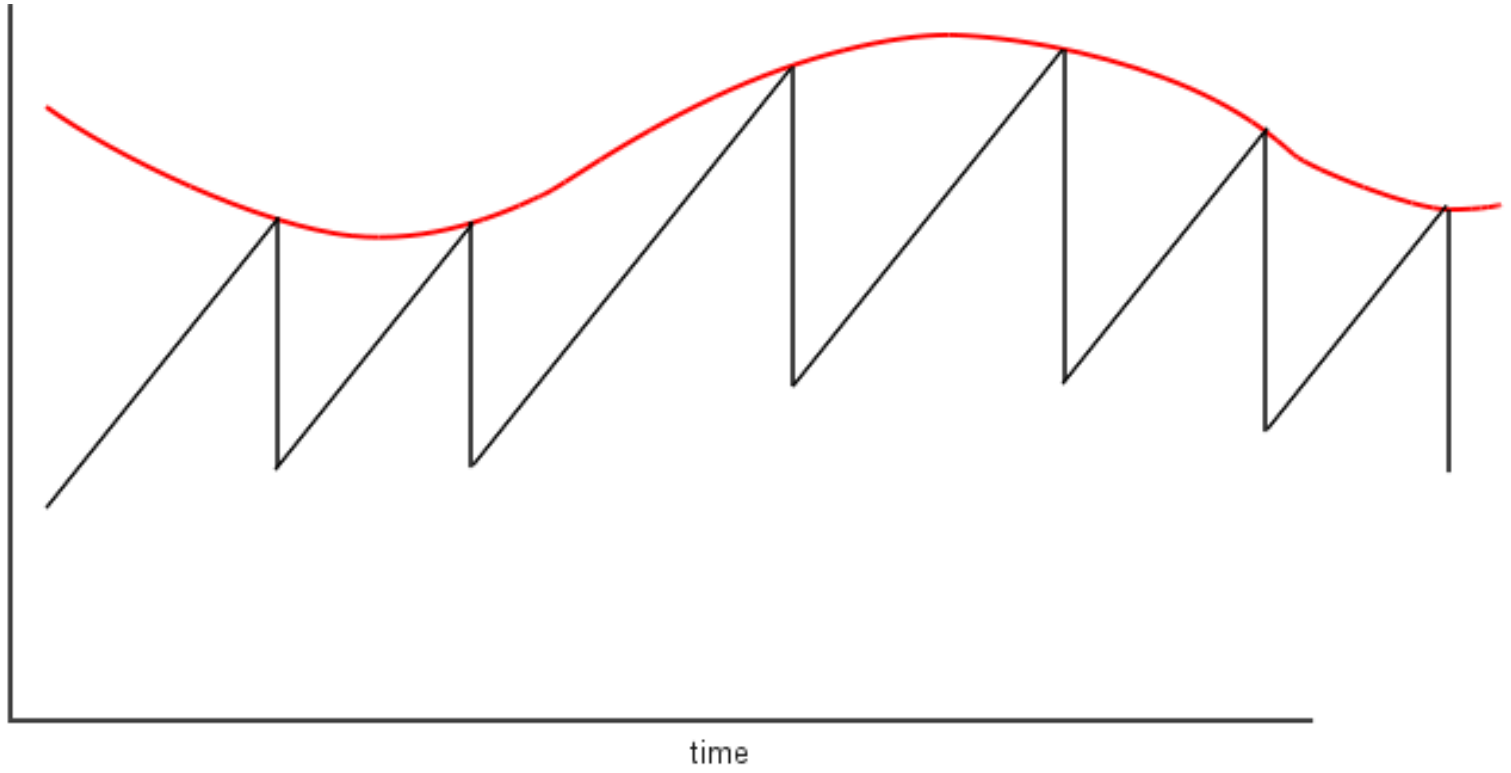

Delay (seconds)
Offered load (packets/sec)

Congestion


Capacity
Goodput (packets/sec)
Desired
response
Congestion
collapse
Offered load (packets/sec)


Delay (seconds)
Onset of
congestion
Offered load (packets/sec)

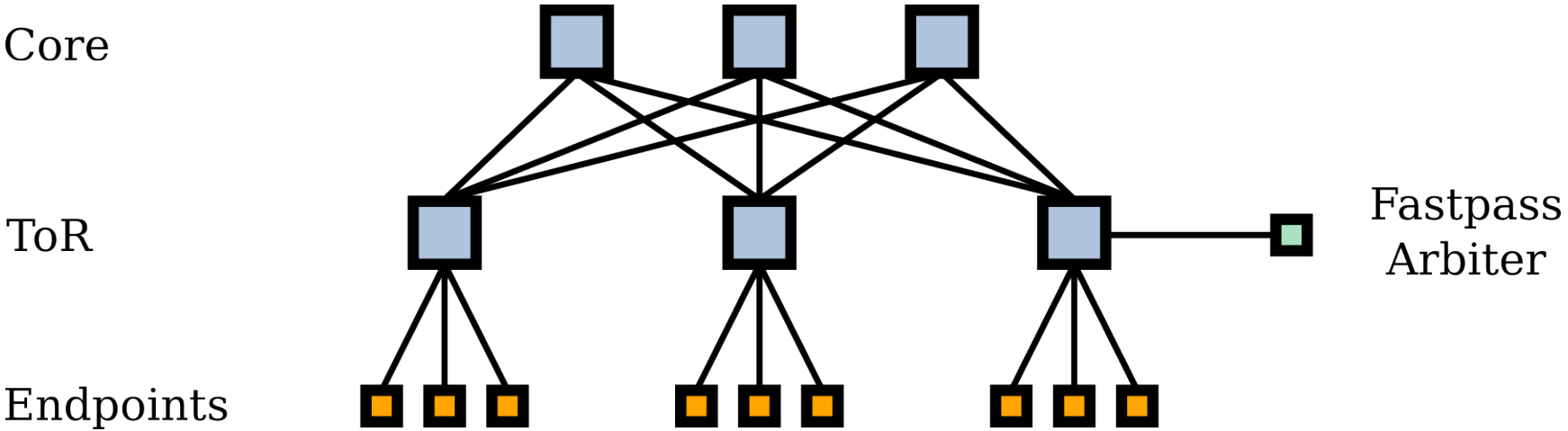
Additive increase/multiplicative decrease



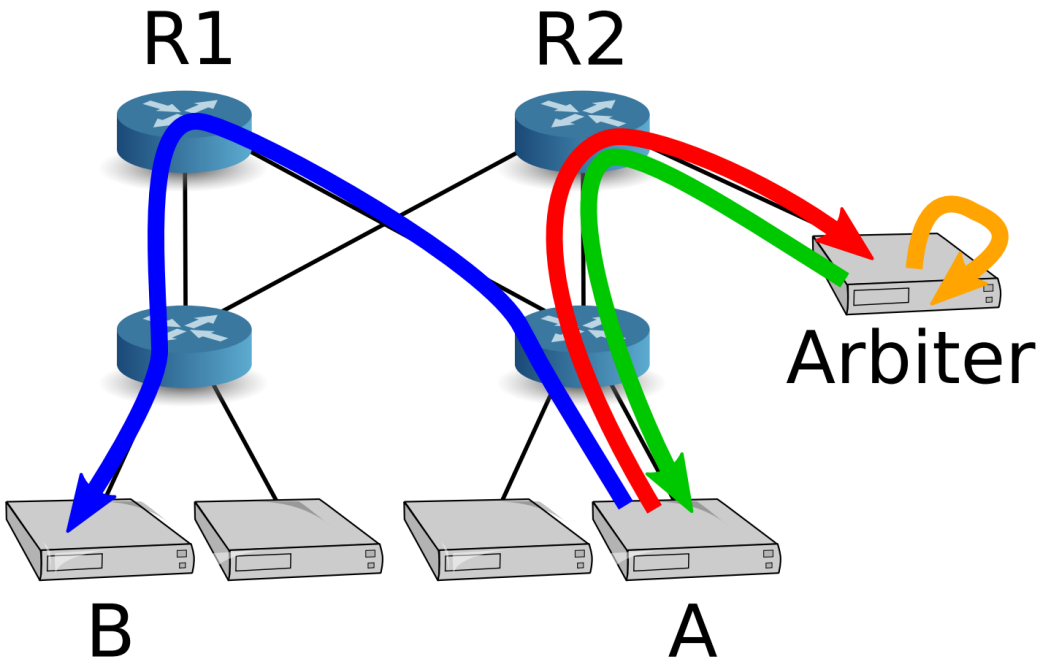
Goals of Fastpass

- No queuing
- High utilization
- Support multiple resource allocation objectives

Fastpass

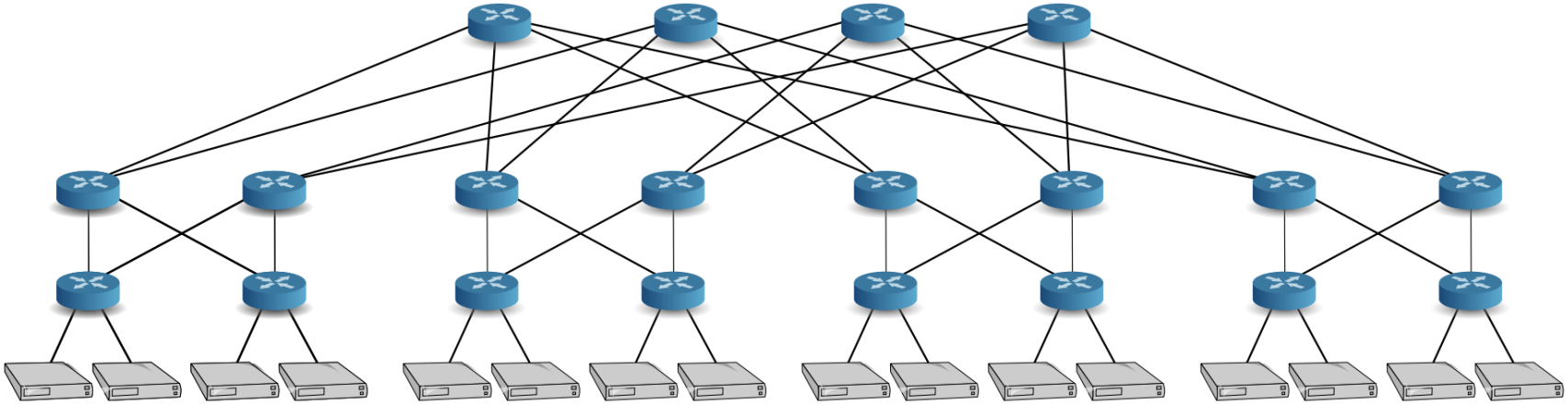


Send package from A to B



5µs	A → Arbiter
1-20µs	Arbiter
15µs	Arbiter → A
no queuing	A → B

Scheduling and path selection



Arbiter treats network as a big switch

How fast must the allocation algorithm be?

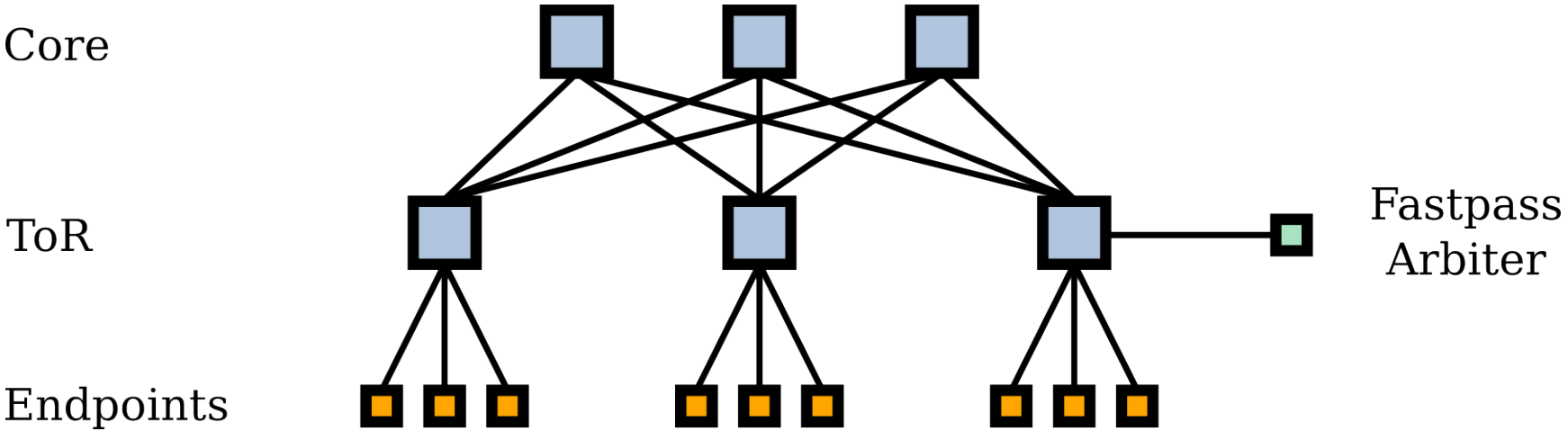
If we assume that we have a 10Gbit/s link and a MTU of 1500 byte

10 gigabits = 1 250 000 000 bytes

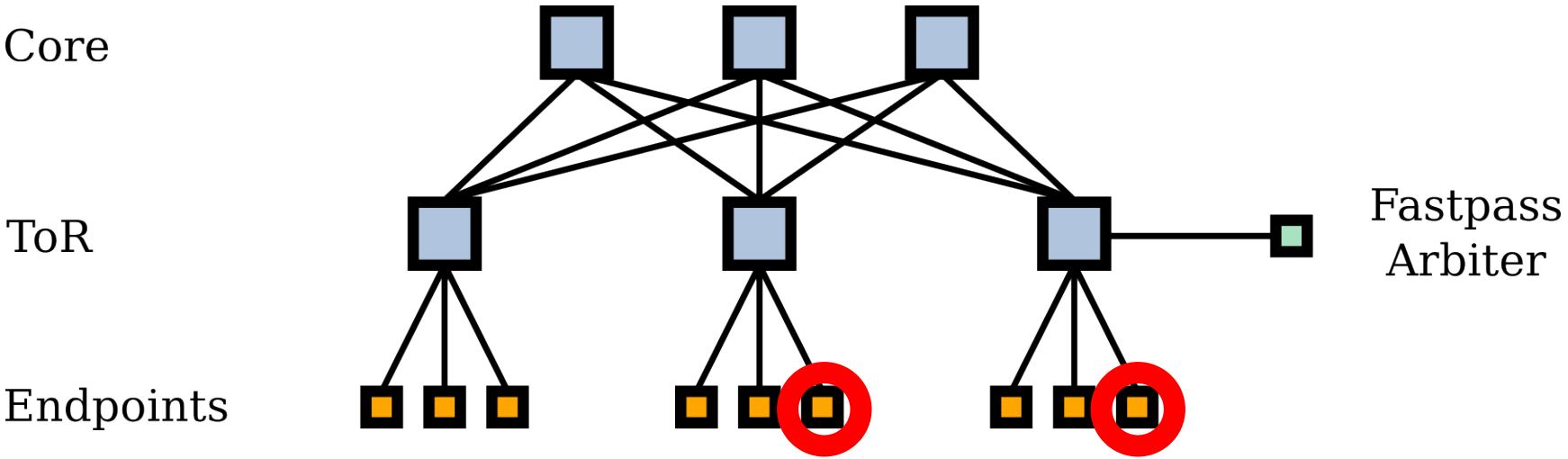
1 250 000 000 bytes / 1500 bytes = 833333 slots per second

0.0000012 second per slot = **1.2** µs per slot

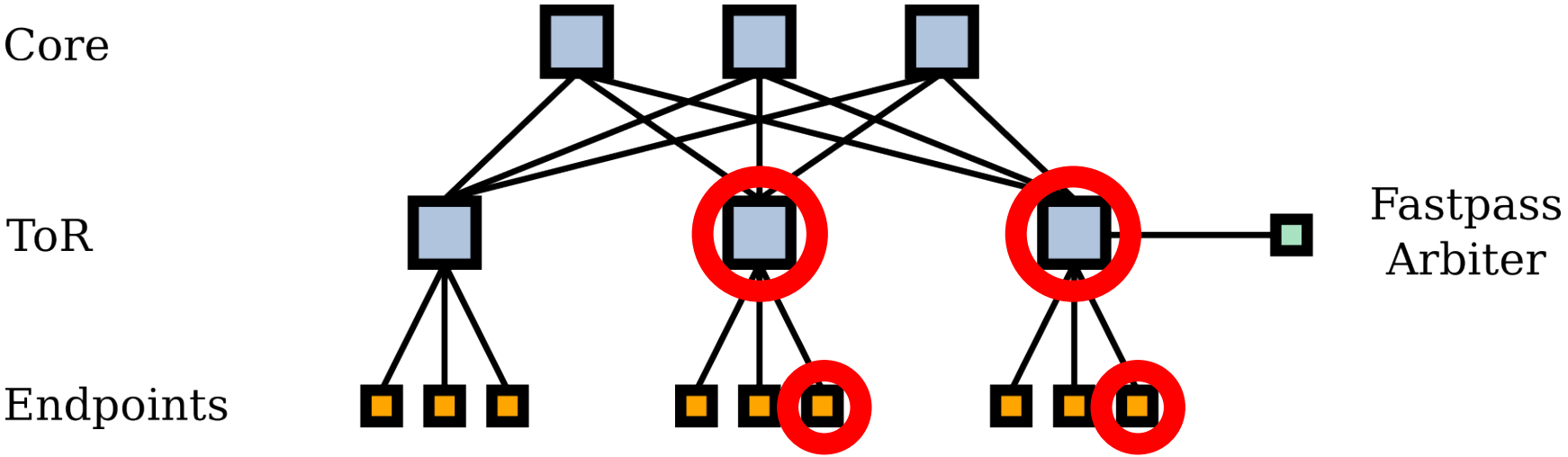
Path selection



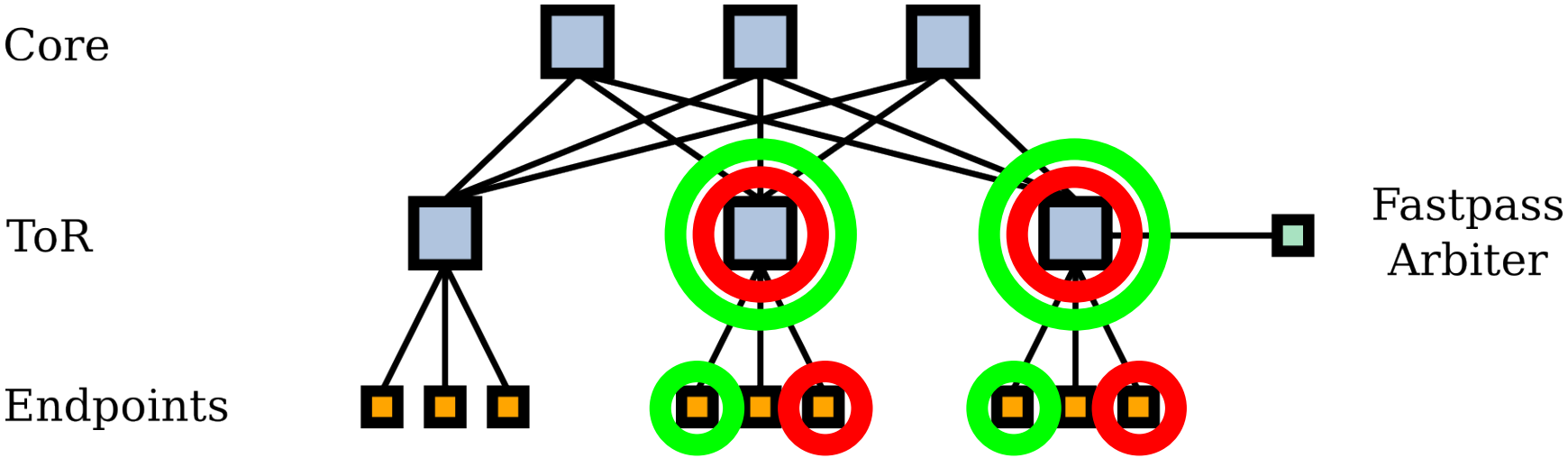
Path selection



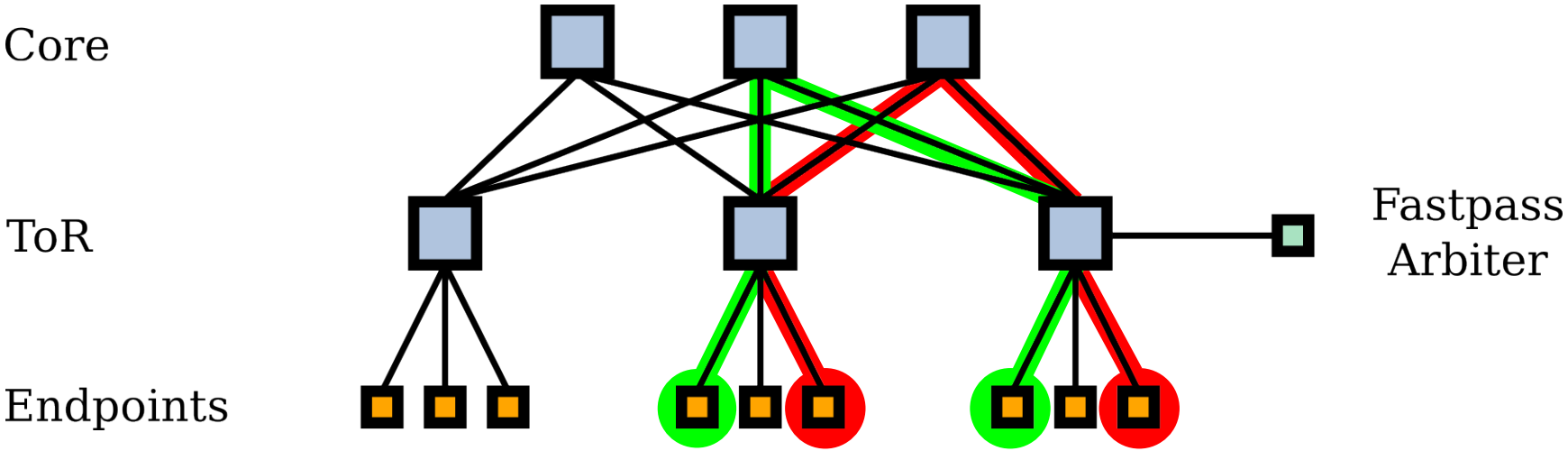
Path selection



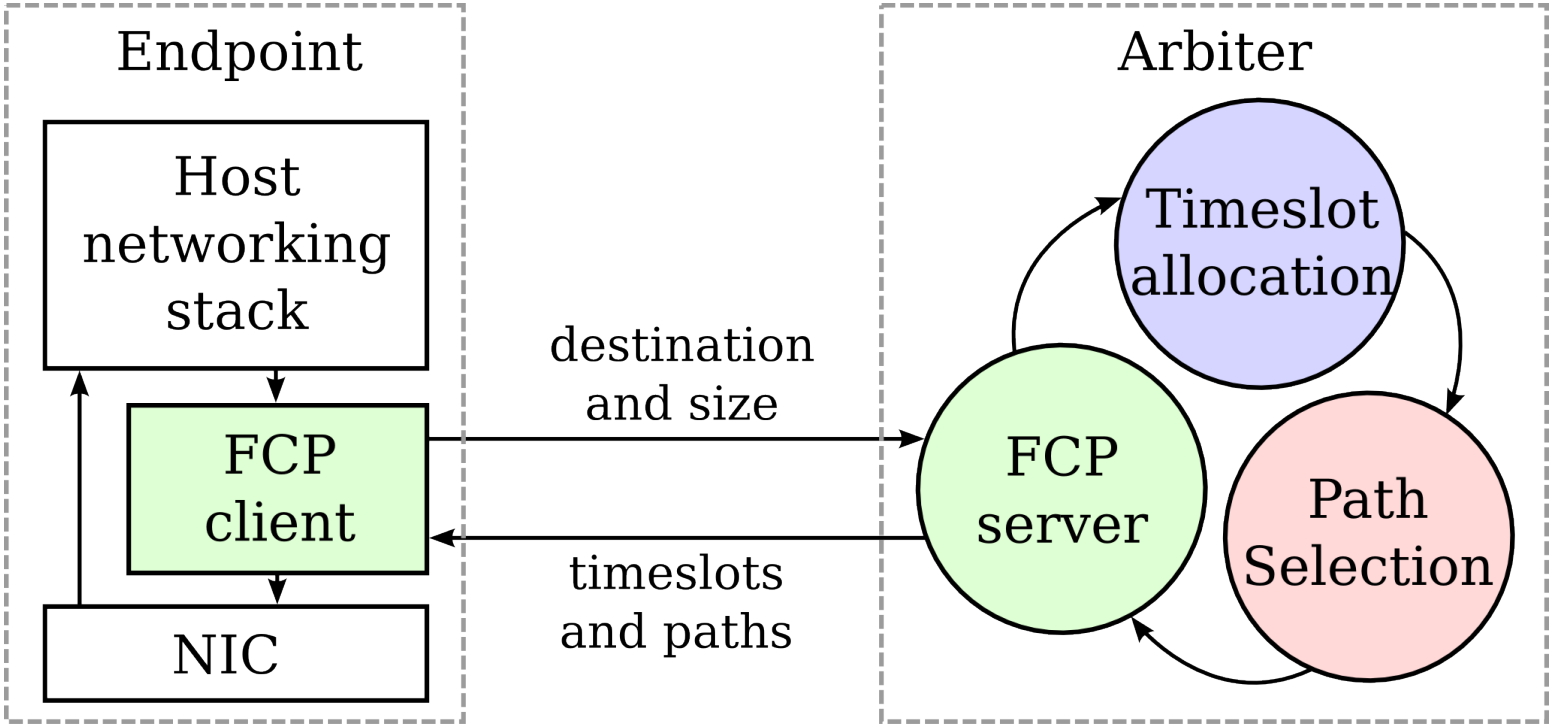
Path selection



Path selection

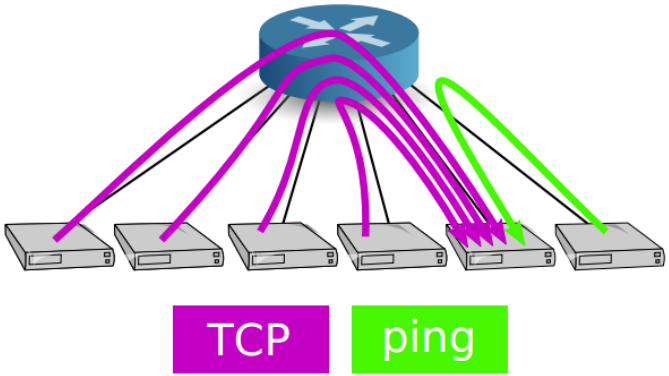


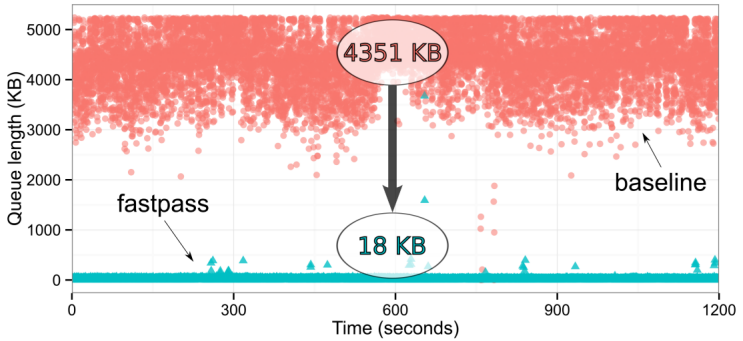
Implementation

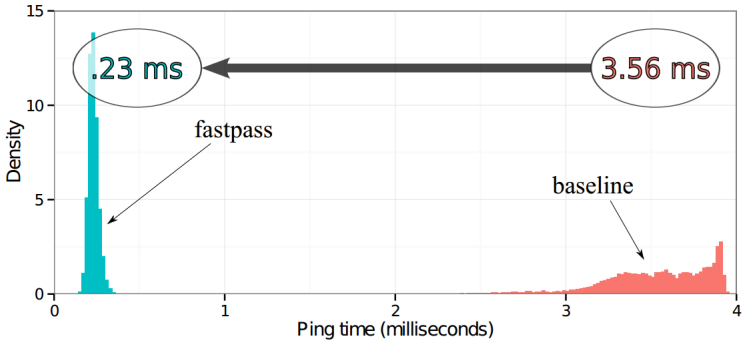


Evaluation

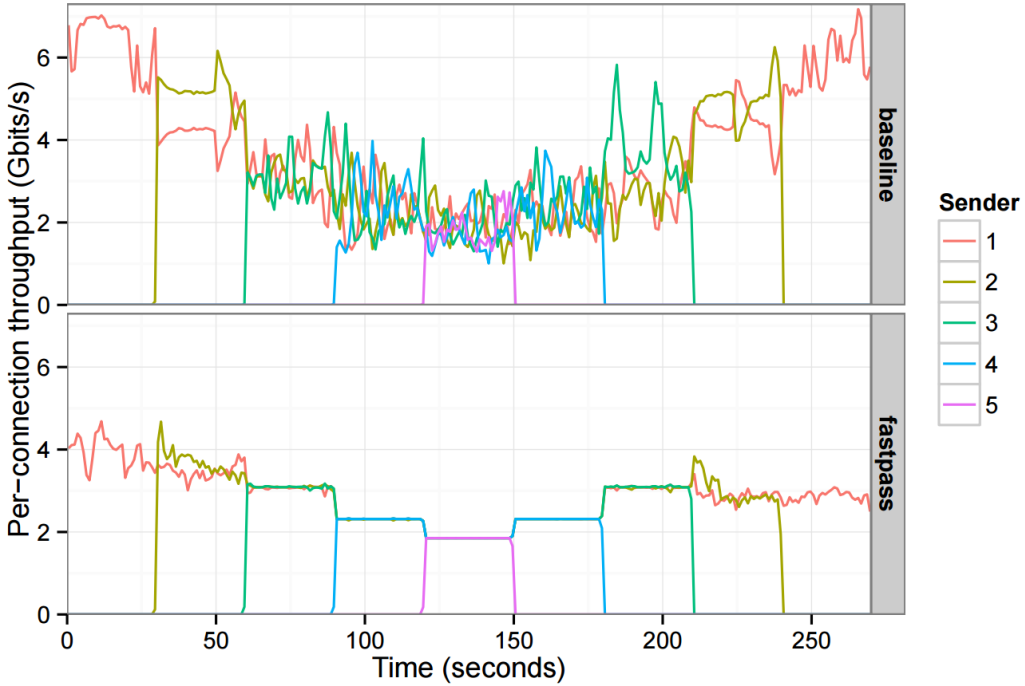
Baseline: 9.43 Gbits/s
Fastpass: 9.28 Gbits/s



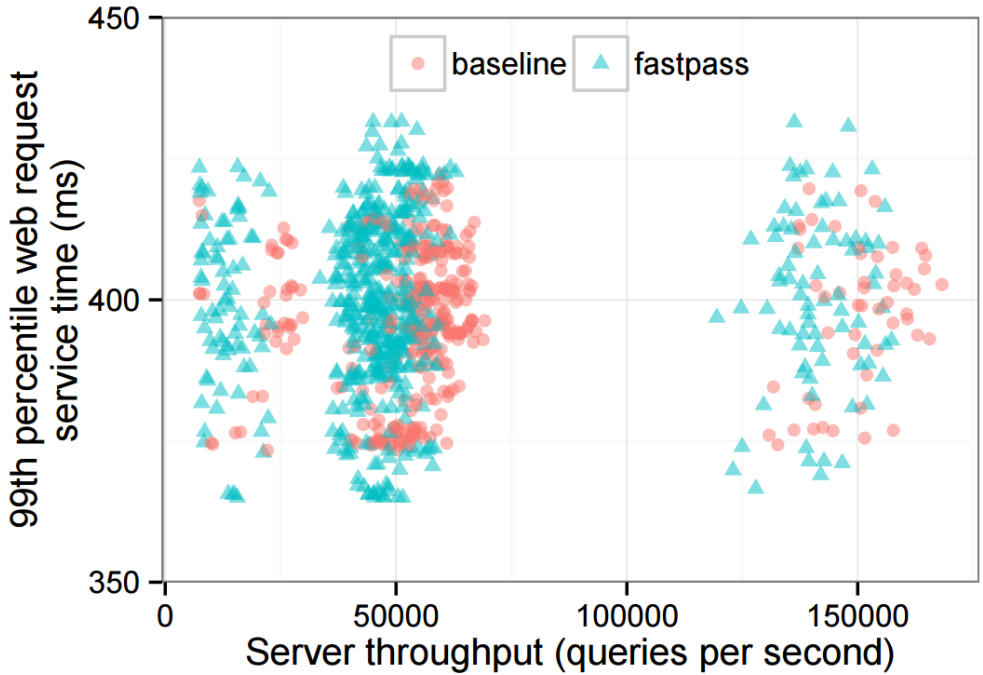




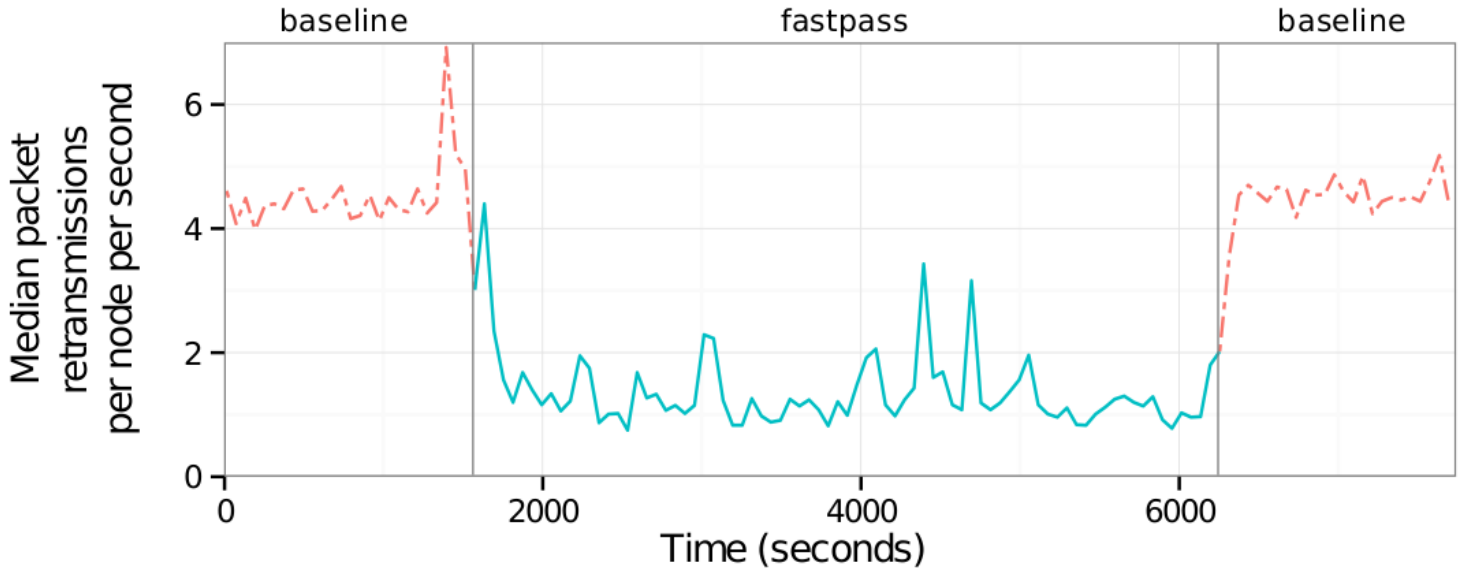
Per−connection throughput



99th percentile



Evaluation


baseline
fastpass
baseline
Median packet retransmissions per node per second
0
2
4
6
0
2000
4000
6000
Time (seconds)

Thank you for your attention.

Sources

http://spcl.inf.ethz.ch/Teaching/2014-osnet/

http://fastpass.mit.edu/Fastpass-SIGCOMM14-Perry.pdf

http://conferences.sigcomm.org/sigcomm/2014/doc/slides/52.pdf