Flexplane: An Experimentation Platform for Resource Management in Datacenters

Amy Ousterhout, Jonathan Perry,
Hari Balakrishnan, Petr Lapukhov

Resource Management

- Dozens of new resource management schemes
  - DCTCP, PDQ, RCP, HULL, pFabric, LSTF, D^3, etc.
- Difficult to experiment with in real networks
  - Schemes require changes to hardware routers

Experimentation with Resource Management

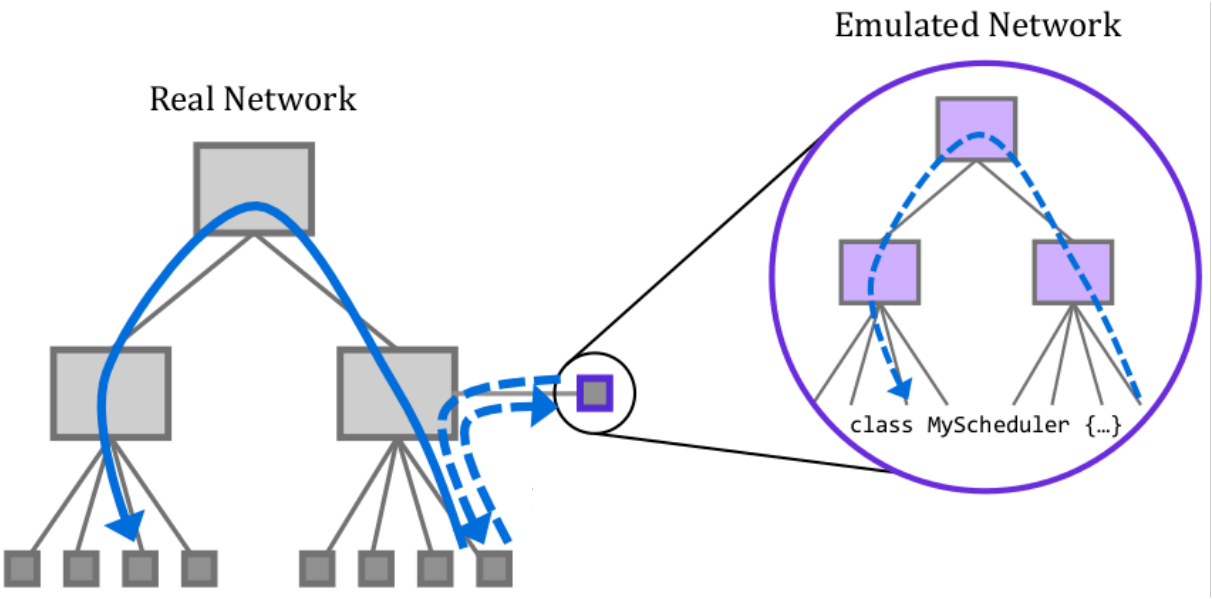
- Experimentation in real networks
  - Software routers - limited throughput
  - Programmable hardware - limited flexibility
- Experimentation in simulation (e.g., ns2)
  - Does not accurately model real network stacks, NICs, and applications

Flexplane: an Experimentation Platform

- Goal: faithfully evaluate resource management schemes
- Flexplane provides:
  - Accuracy – predict behavior of hardware
  - Flexibility – express schemes in C++
  - High throughput – run at hardware rates

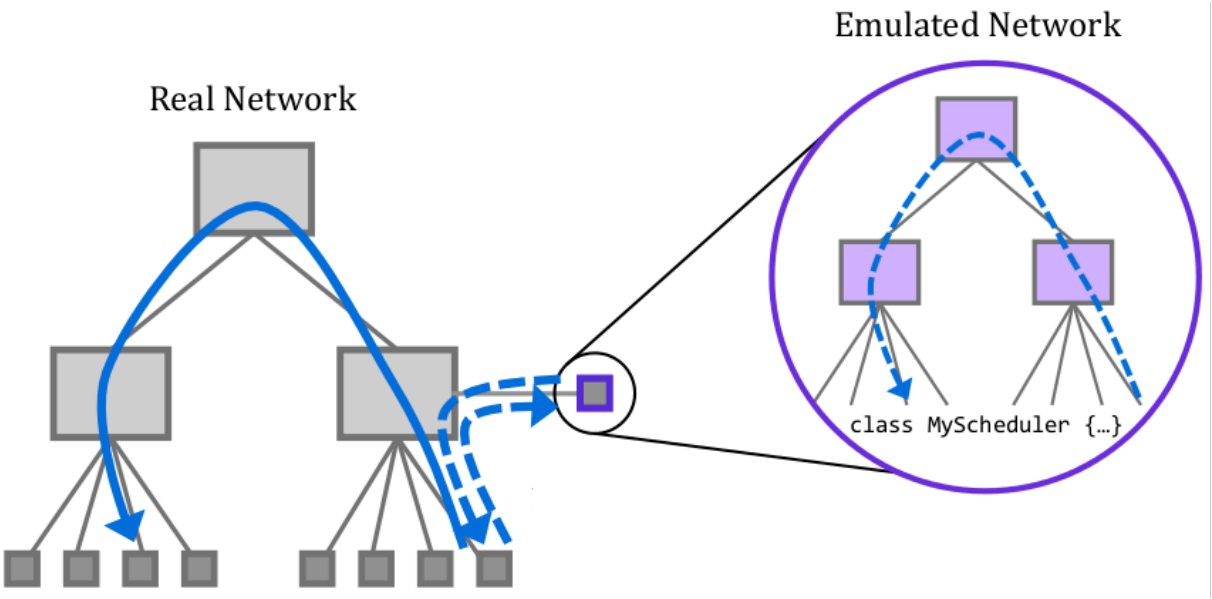
Approach: Whole-Network Emulation

- Emulator maintains a model of the real network
- Users implement schemes in emulated routers
- Packets experience same behavior in emulator as in hardware network running same scheme



Three Steps for Each Packet

- *Convey* abstract packet to emulator
- *Emulate* the network behavior
  - Time divided into timeslots
- *Reflect* behavior onto real network



Accuracy

- Goal: predict behavior of a hardware network

Hardware: $$l = u + q$$

Flexplane: $$l' = r + t_e + q_e + u + q' \leq 4u + q' + q_e$$

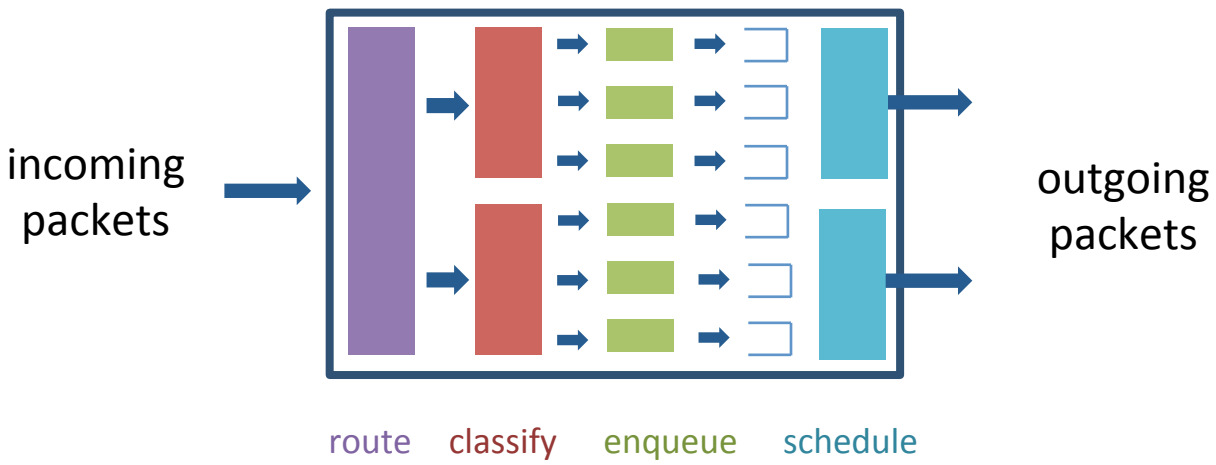
l: latency
u: unloaded delay
q: queuing delay
r: RTT to the emulator

t: transmission delay
s: switch delay
u = *t* + *s*

Flexplane API

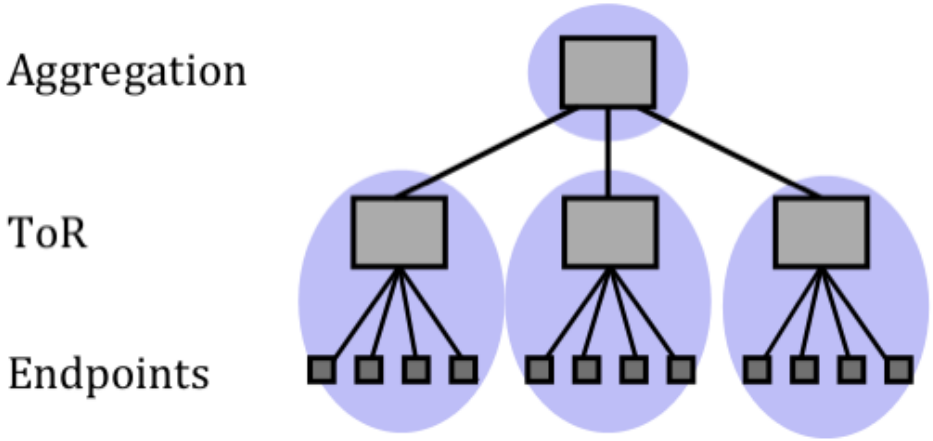
- Decouples framework from schemes

Emulator	int route(AbstractPkt *pkt) int classify(AbstractPkt *pkt, int port) enqueue(AbstractPkt *pkt, int port, int queue) AbstractPkt *schedule(int output_port)



Multicore Emulator Architecture

- Pin network components (routers, endpoints) to cores
  - Router state not shared across cores
- Communication via FIFO queues
- Achieves 761 Gbits/s with 8 cores



Flexplane is Easy to Use

- Implemented several resource management schemes in dozens of lines of code

scheme	LOC
drop tail queue manager	39
RED queue manager	125
DCTCP queue manager	43
priority queueing scheduler	29
round robin scheduler	40
HULL scheduler	60
pFabric QM, queues, scheduler	251

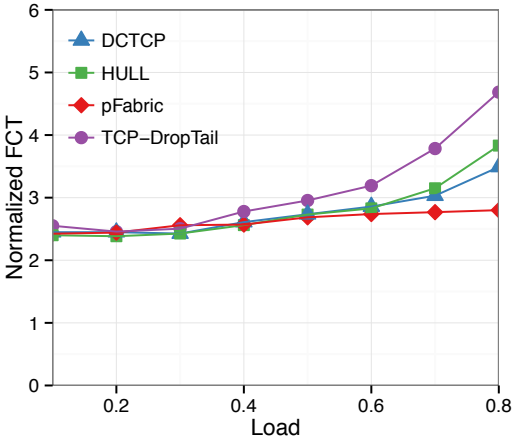
Flexplane is Accurate

- Bulk TCP: 5 senders, 1 receiver
- Throughput 9.2-9.3 Gbits/s vs. 9.4 Gbits/s in hardware
- Similar queue occupancies

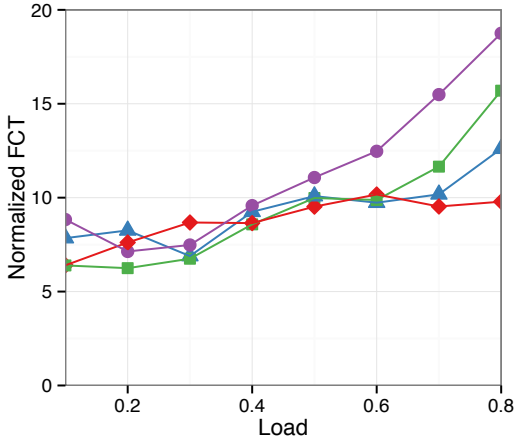
	Queue Occupancies (MTUs)			
	Hardware		Flexplane	
	median	σ	median	σ
DropTail	931	73.7	837	98.6
RED	138	12.9	104	32.5
DCTCP	61	4.9	51	13.0

Flexplane Enables Experimentation

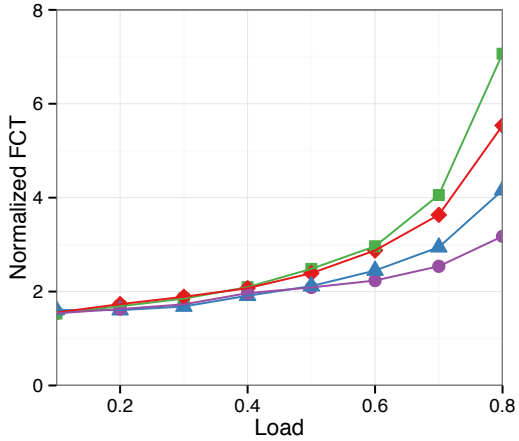
- Reproducible research in real networks



(a) (0, 100KB]: Average



(b) (0, 100KB]: 99th percentile



(c) (10MB, ∞): Average

- Experiment with Spark
  - Results depend on resource management scheme and application