Extending SDN to the Data Plane

Anirudh Sivaraman, Keith Winstein, Suvinay Subramanian, Hari Balakrishnan

M.I.T.

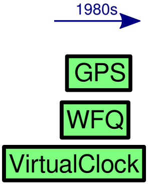
http://web.mit.edu/anirudh/www/sdn-data-plane.html

Switch Data Planes today

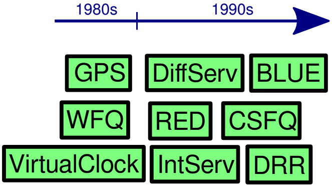
Two key decisions on a per-packet basis:

- **Scheduling**: Which packet to transmit next?
- **Queue Management**: How long can queues grow? Which packet to drop?

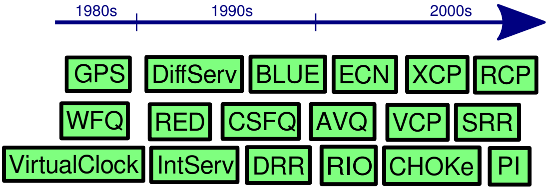
The long lineage of in-network algorithms



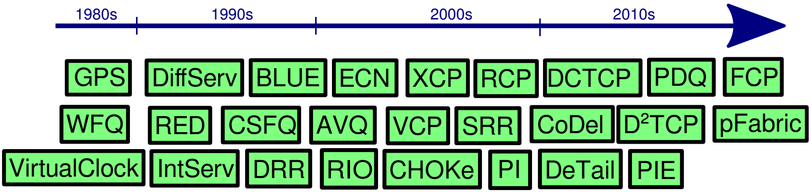
The long lineage of in-network algorithms



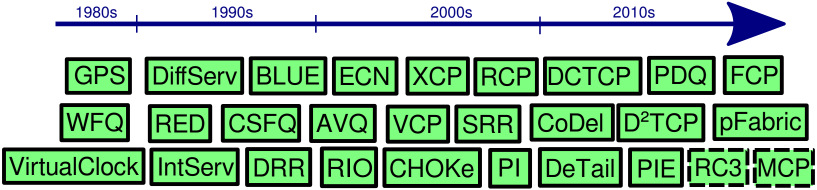
The long lineage of in-network algorithms



The long lineage of in-network algorithms



The long lineage of in-network algorithms



The Data Plane is continuously evolving

- Each scheme wins in its own evaluation.
- Quest for a “silver bullet” in-network method.

We disagree: There is no silver bullet!

- Different applications care about different objectives.
- Applications use different transport protocols.
- Networks are heterogeneous.

Our work:

- Quantify non-universality of in-network methods.
- Extend SDN to the Data Plane to handle in-network diversity.

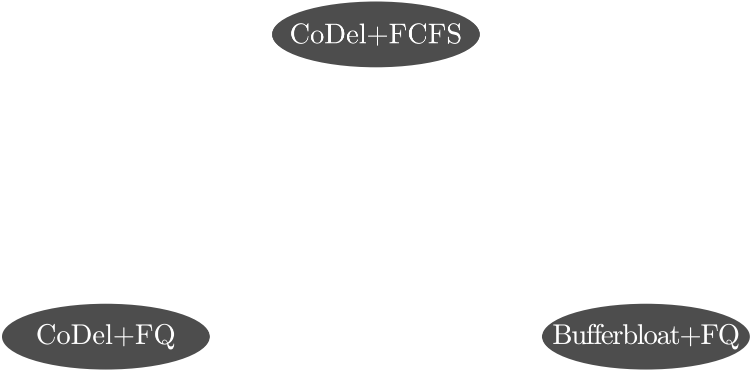
Quantifying "No Silver Bullet": Network Configurations

Configuration	Description
CoDel+FCFS	One shared FCFS queue with CoDel
CoDel+FQ	Per-flow fair queueing with CoDel on each queue (Nichols 2013)
Bufferbloat+FQ	Per-flow fair queueing with deep buffers on each queue

Quantifying "No Silver Bullet": Workloads and Objectives

Workload	Description	Objective
Bulk	Long-running bulk transfer flow	Max. throughput
Web	Switched flow with ON/OFF periods	Min. 99.9 %ile flow completion time
Interactive	Long-running interactive flow	Max. $\frac{\text{throughput}}{\text{delay}}$ (power)

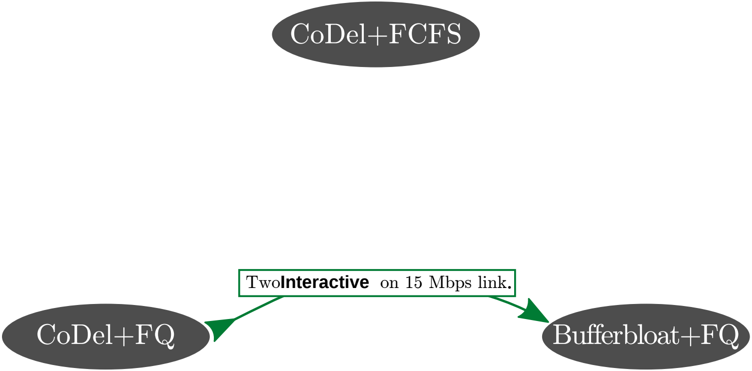
Quantifying "No Silver Bullet"



Quantifying “No Silver Bullet”

Nwk config.	Avg. throughput, delay	Power
Bufferbloat+FQ	7.47 Mbps, 62165 ms	0.12 $Mbit/s^2$
CoDel+FQ	6.55 Mbps, 76.5 ms	85.6 $Mbit/s^2$

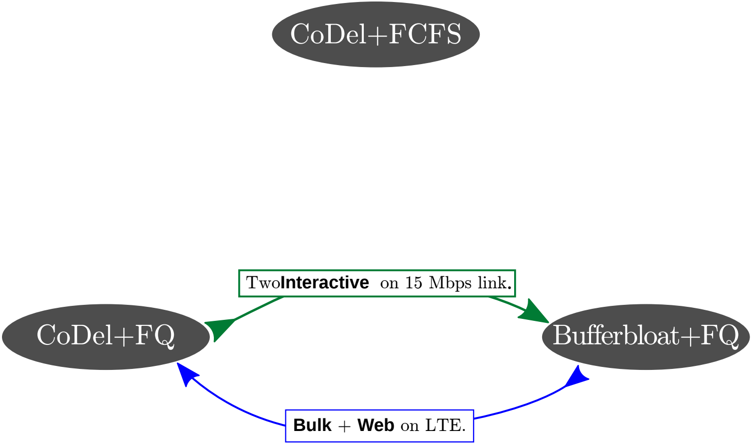
Quantifying "No Silver Bullet"



Quantifying "No Silver Bullet"

Nwk config.	Bulk Throughput	Web Tail FCT
Bufferbloat+	11.22 Mbps	20.94 secs
CoDel+FQ	3.92 Mbps	43.72 secs

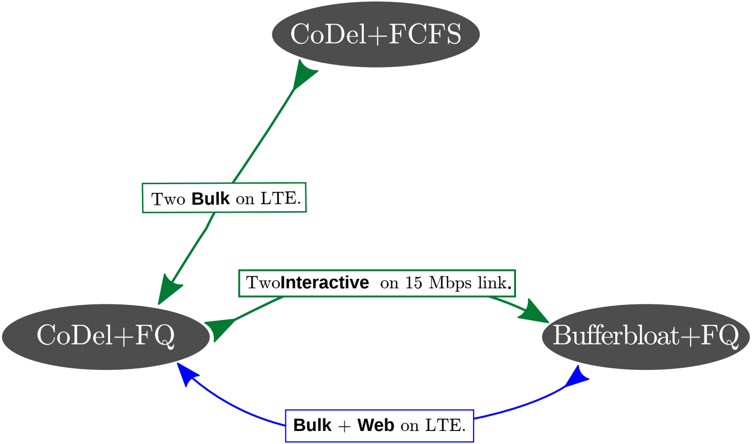
Quantifying “No Silver Bullet”



Quantifying “No Silver Bullet”

Nwk config.	Avg. throughput
CoDel+FC	2.00 Mbps
CoDel+FQ	1.90 Mbps

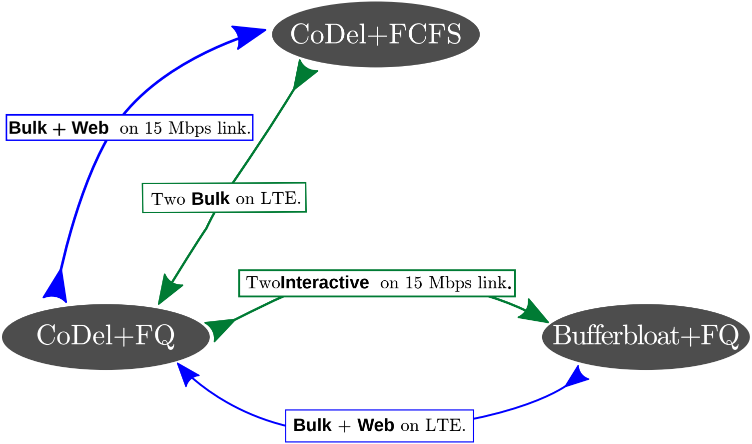
Quantifying “No Silver Bullet”



Quantifying “No Silver Bullet”

Nwk config.	Bulk Throughput	Web Tail FCT
CoDel+FCFS	9.48 Mbps	22.25 secs
CoDel+FQ	9.48 Mbps	18.71 secs

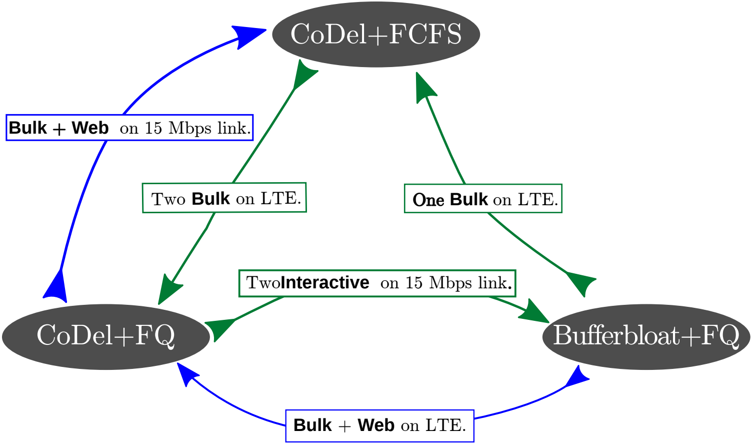
Quantifying “No Silver Bullet”



Quantifying “No Silver Bullet”

Nwk config.	Bulk throughput
Bufferbloat+FQ	11.96 Mbps
CoDel+FCFS	4.35 Mbps

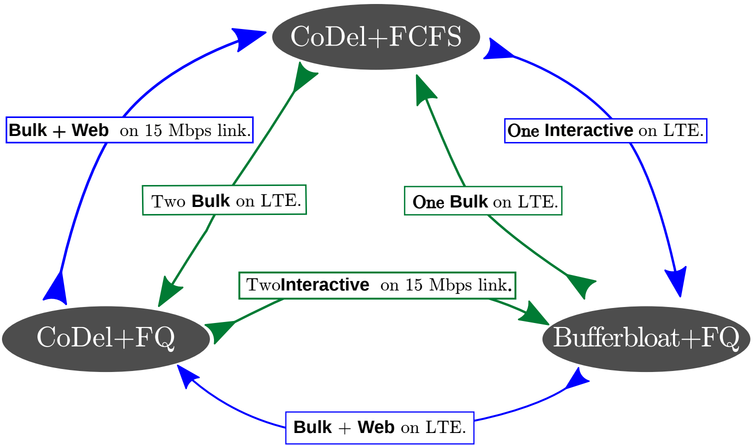
Quantifying “No Silver Bullet”



Quantifying “No Silver Bullet”

Nwk config.	Interactive throughput, delay	Power
Bufferbloat+FQ	11.96 Mbps, 46028 ms	0.26 $Mbit/s^2$
CoDel+FCFS	4.35 Mbps, 83.2 ms	52.28 $Mbit/s^2$

Quantifying "No Silver Bullet"



Why is no single data plane configuration the best?

- Bufferbloat gives the best throughput on variable-rate links.
- FCFS is preferable to Fair Queuing with homogenous objectives.
- Fair Queuing is preferable with heterogeneous objectives.

So what should the network designer do?

- Don't strive for the best in-network behaviour.
- Instead, architect for evolvability.
- Conceptually, extend SDN to include the data plane as well.

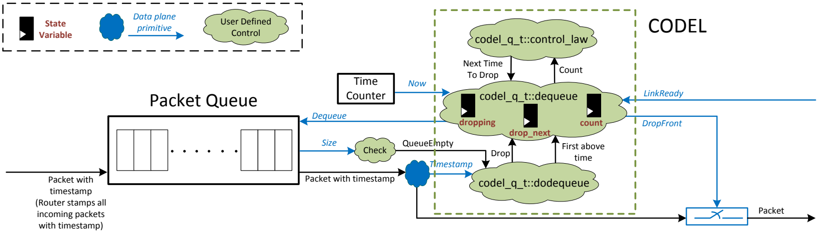
Flexibility without sacrificing performance

- Provide interfaces only to the head and tail of queues
- Operators specify only queue-management/scheduling logic
- No access to packet payloads.

Building such a data plane in four parts

- Hardware gadgets
  - Random number generators (RED, BLUE)
  - Binary tree of comparators (pFabric, SRPT)
- I/O interfaces
  - Drop/mark head/tail of queue
  - Interrupts for enqueue/dequeue
  - Rewrite packet fields
- State maintenance
  - Per-flow (WFQ, DRR)
  - Per-dst address (PF)
- A domain-specific instruction set
  - Expresses control flow
  - Implements new functions unavailable in hardware

Feasibility study: CoDel



Synthesis numbers on the Xilinx Kintex-7

Resource	Usage	Fraction
Slice logic	1,256	1%
Slice logic dist.	1,975	2%
IO/GTX ports	27	2%
DSP slices	0	0%
Maximum speed	12.9 million pkts/s ~10 Gbps	

- Small fraction of the FPGA's resources.
- Can be improved by pipelining or parallelizing.

Limitations and Practical Considerations:

- Cannot express several network functions that need payloads.
- Mechanism to signal application objectives.
- Feasibility at 10G on high port-density switches.
- Energy, area, and performance costs of flexibility.

Related Work

- Active Networking, e.g., ANTS
- Software Routers, e.g., Click
- Software-Defined Networking, e.g., OpenFlow

Conclusion

- No silver bullet to in-network resource allocation.
- Algorithms will evolve: the data plane should help
- Reproduce our results: http://web.mit.edu/anirudh/www/sdn-data-plane.html