Making Decisions at Data Plane Speeds

Srinivas Narayana
Jun 19, 2023
Decision Making in Self-Driving Networks

• **Good data** leads to good decisions

• But good data is hard to find
  • Raw signals can be hard to measure
  • Big data and needle in a haystack
  • Existing algorithms ineffective to collect data

• This talk:
  • Lessons from three personal stories
  • A call to arms
#1: Network Performance Diagnosis

- Problem: Root-causing microbursts
- Queue size just made visible on new RMT-based Tofino switches
  - But a firehose of per-packet data

Lesson #1: Pushdown
Filter and aggregate at data plane speed

Lesson #2: Reduce signal loss
Design primitives that maintain accuracy

Marple, sigcomm’17
#1: Network Performance Diagnosis

- Problem: Root causing microbursts
- Queue size just made visible on new RMT-based Tofino switches
- But a firehose of per-packet data

Lesson #1: Pushdown Filter and aggregate at data plane speed

Lesson #2: Reduce info loss
Design primitives that maintain accuracy

Marple, sigcomm'17
#2: Congestion Control

- Problem: Implementing complex congestion control algorithms
  - Existing datapaths hard to use
  - Re-implementing congestion control on emerging software datapaths

Lesson #3: Low-level software isn’t as fungible as “regular” software

- Limited to fold functions in the datapath

Lesson #4: Apportion flexibility by timescale

CC, sigcomm’18
#3. Server Load Balancing

• Problem: balance server load based on server performance
• State of the art: use a server agent to introspect server performance

Lesson #5: Prefer in-band control
Avoid staleness and eliminate big data

• Use triggered packets to estimate server response latency

feedbackLB, hotnets’22
#3. Server Load Balancing

- Problem: balance server load based on server performance
- State of the art: use a server agent to introspect on performance
- Use triggered packets to estimate server response latency

Lesson #4: Prefer in-band control
Avoid staleness and eliminate big data

95th percentile

Introduced a delay on one of the server

Maglev tail latency increases and stays there

Latency aware feedback loop reduces tail latency in ~1 ms
A Call To Arms

- Emerging new substrates for telemetry data and feedback control
  - Kernel extensions
  - Service meshes in container networking
A Call To Arms

- Emerging new substrates for telemetry data and feedback control
  - Kernel extensions
  - Service meshes in container networking
- **Observability** at all levels of the stack
- Significant barriers exist
  - Safe extensibility: eBPF verifier
  - Poor performance

Sigcomm’21, cgo’22, nsdi’23, cav’23, ongoing…

Thanks to my collaborators! Q?

Alexei Staravoitov, bpfconf’23