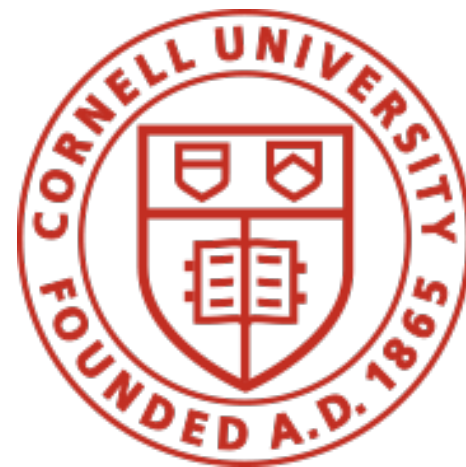


Causal Network Telemetry

Yunhe Liu

Nate Foster

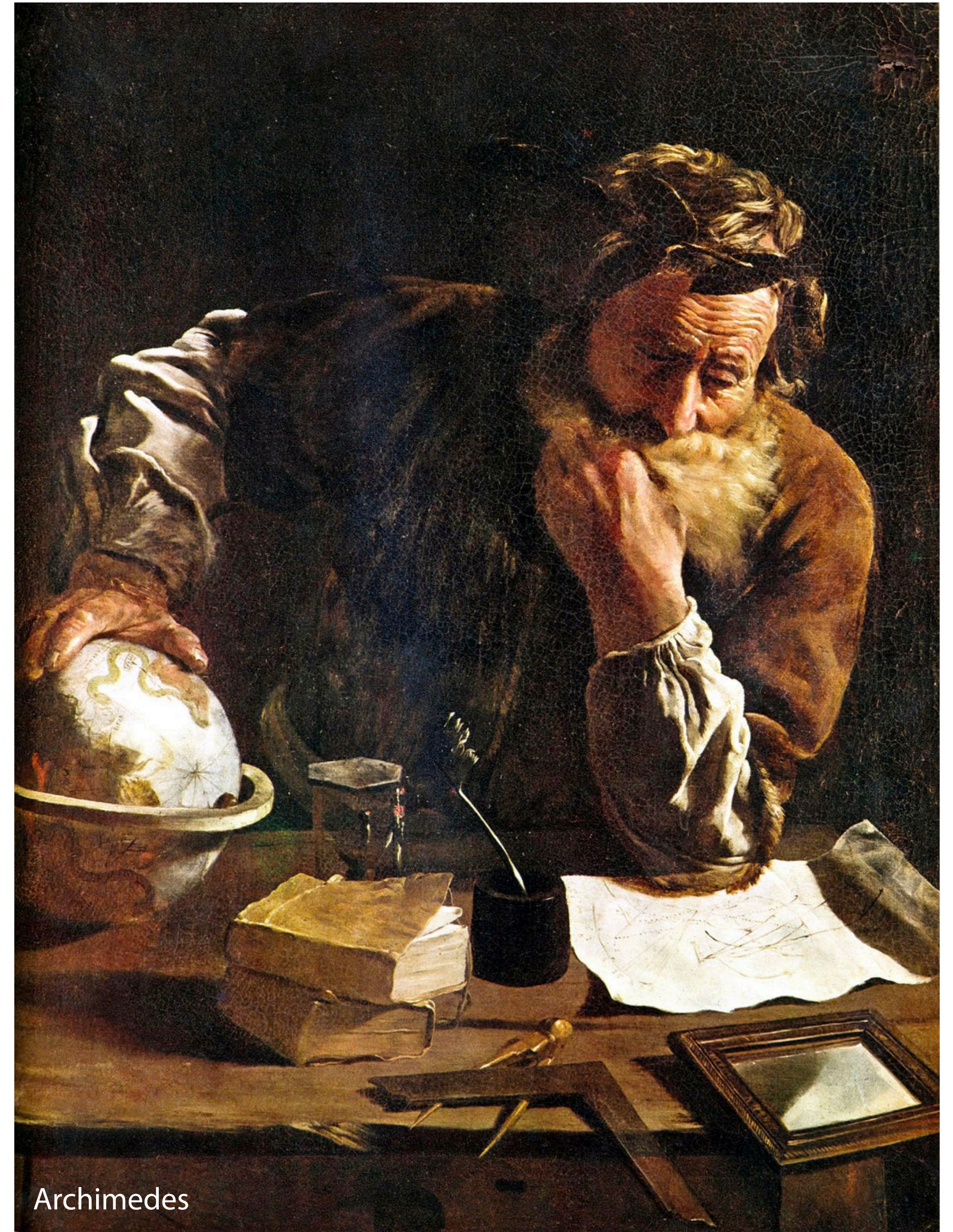
Fred B. Schneider



Ancient Methods

Operators have long relied on rudimentary tools and feats of engineering to understand network behavior...

- **ping** and **traceroute**
- Packet and byte counters
- NetFlow and sFlow



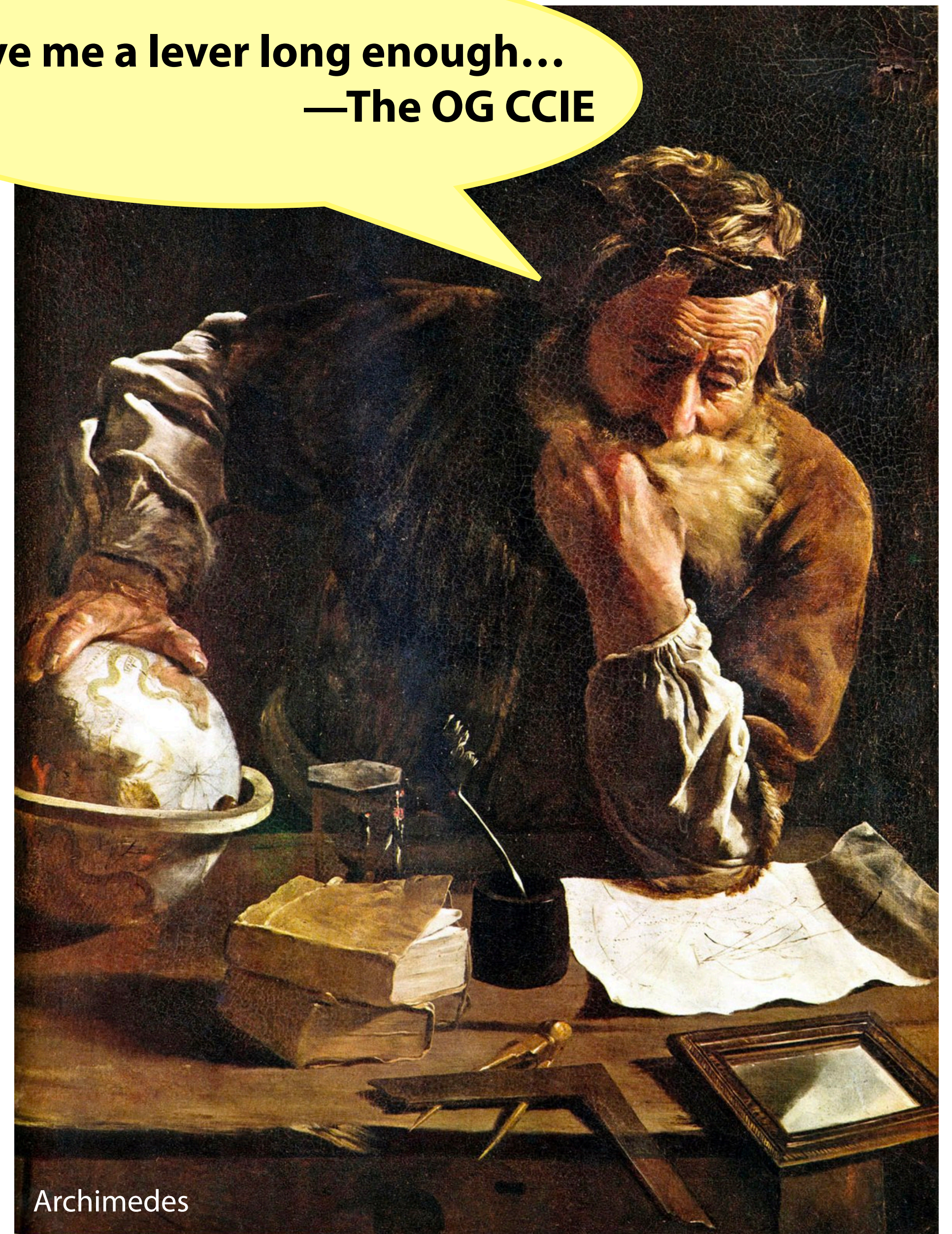
Archimedes

Ancient Methods

Operators have long relied on rudimentary tools and feats of engineering to understand network behavior...

- **ping** and **traceroute**
- Packet and byte counters
- NetFlow and sFlow

Give me a lever long enough...
—The OG CCIE



Archimedes

Software-Centric Model

Then, about a decade ago, we discovered that software, not standards and vendors, could be at the center of the networking universe.



Copernicus

Software-Centric Model

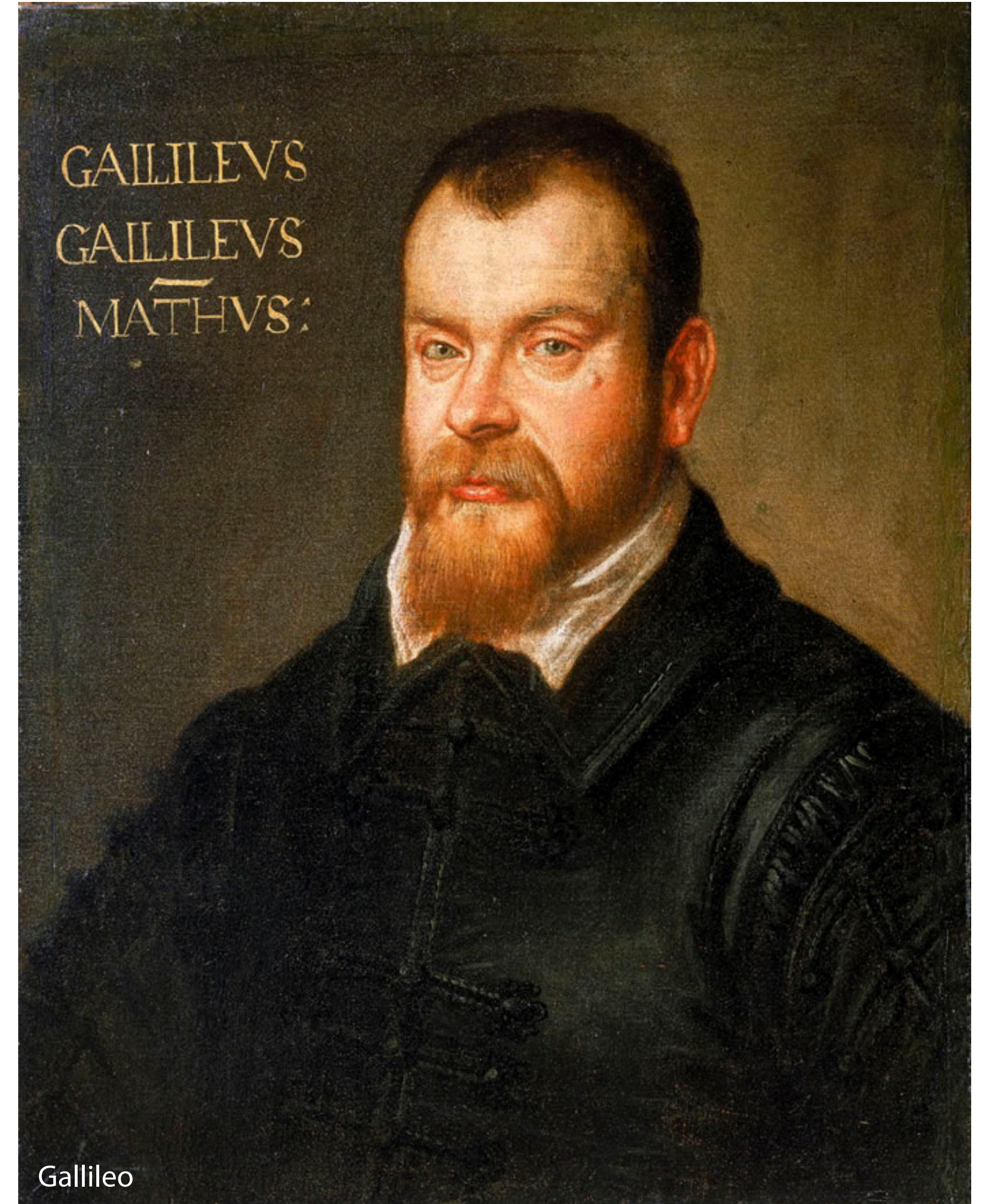
Then, about a decade ago, we discovered that software, not standards and vendors, could be at the center of the networking universe.



Copernicus

The Renaissance

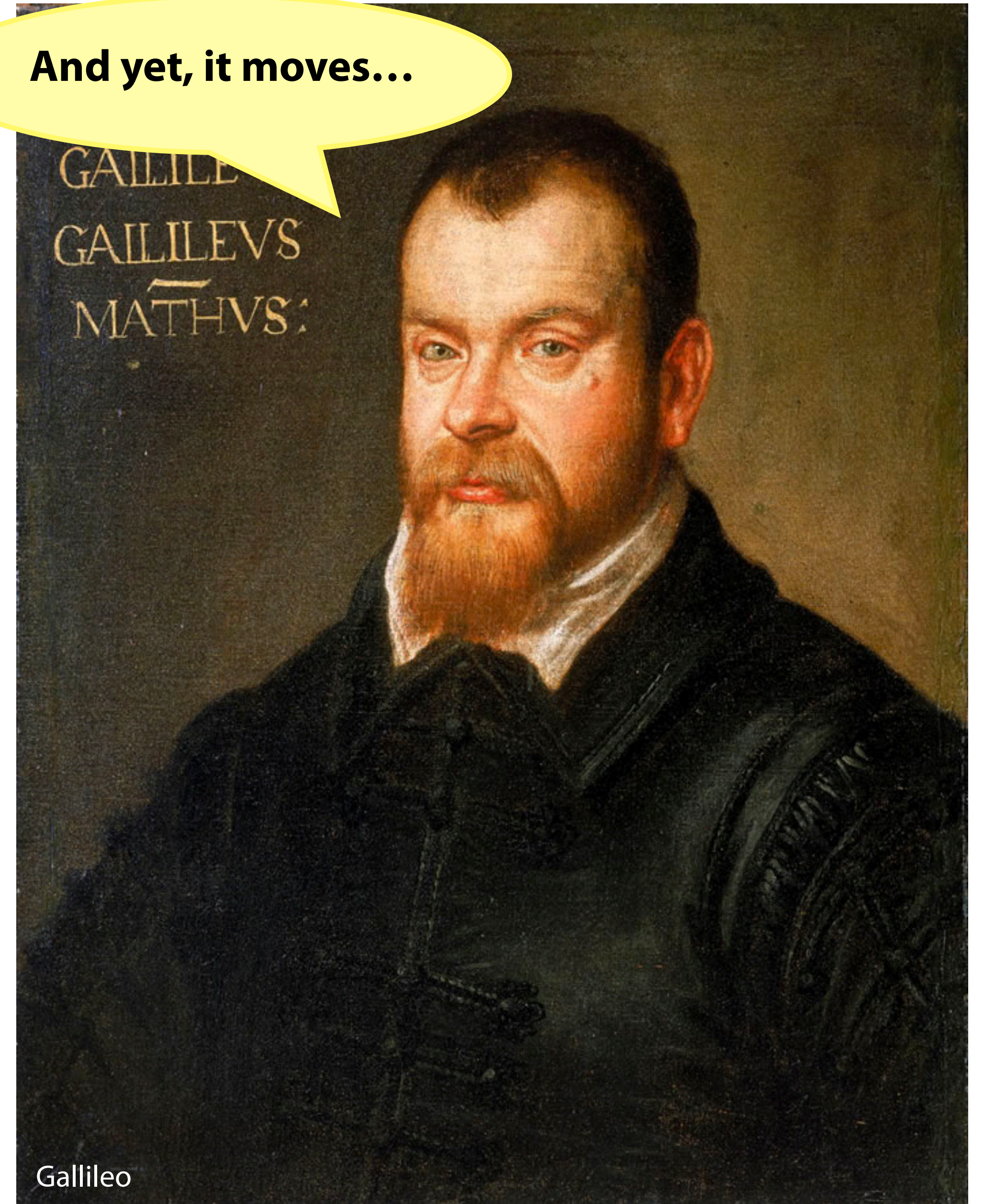
Programmability has led to a renaissance with novel systems such as in-band network telemetry (INT), which keeps a log for every packet traversing the networks and enables fine-grained debugging and analysis.



The Renaissance

Programmability has led to a renaissance with novel systems such as in-band network telemetry (INT), which keeps a log for every packet traversing the networks and enables fine-grained debugging and analysis.

And yet, it moves...



Galileo

Enlightenment Thinking

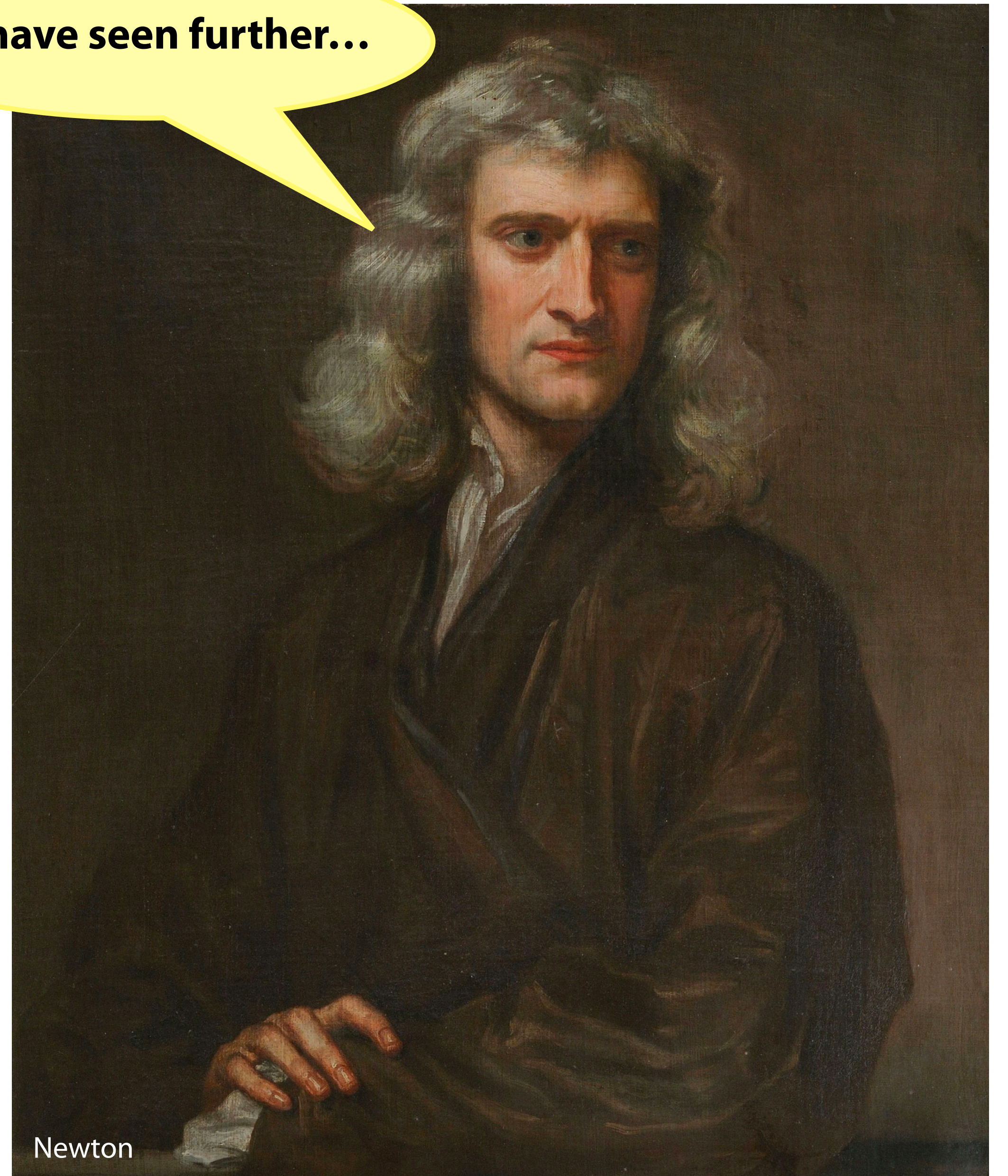
And now, the shift from art to science has begun. Researchers have been designing clean models for telemetry with elegant schemas, high-level query languages, hardware compilers, etc.



Enlightenment Thinking

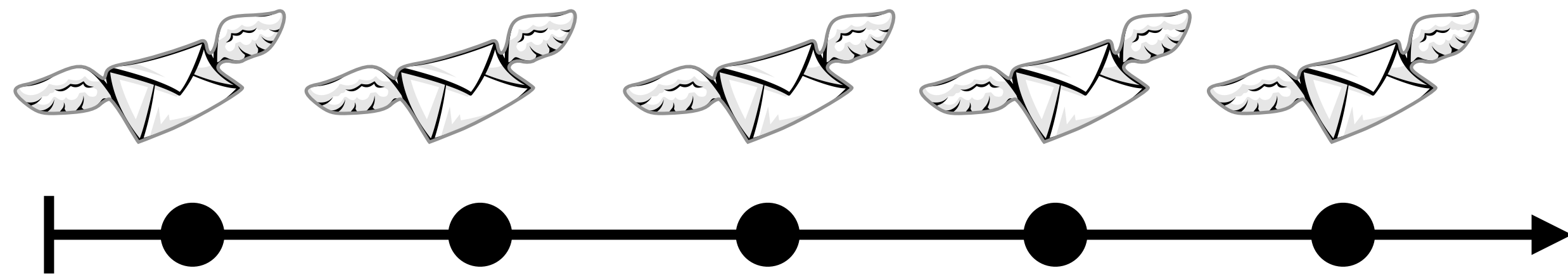
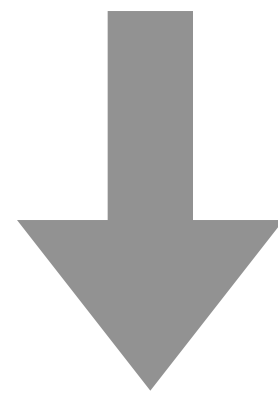
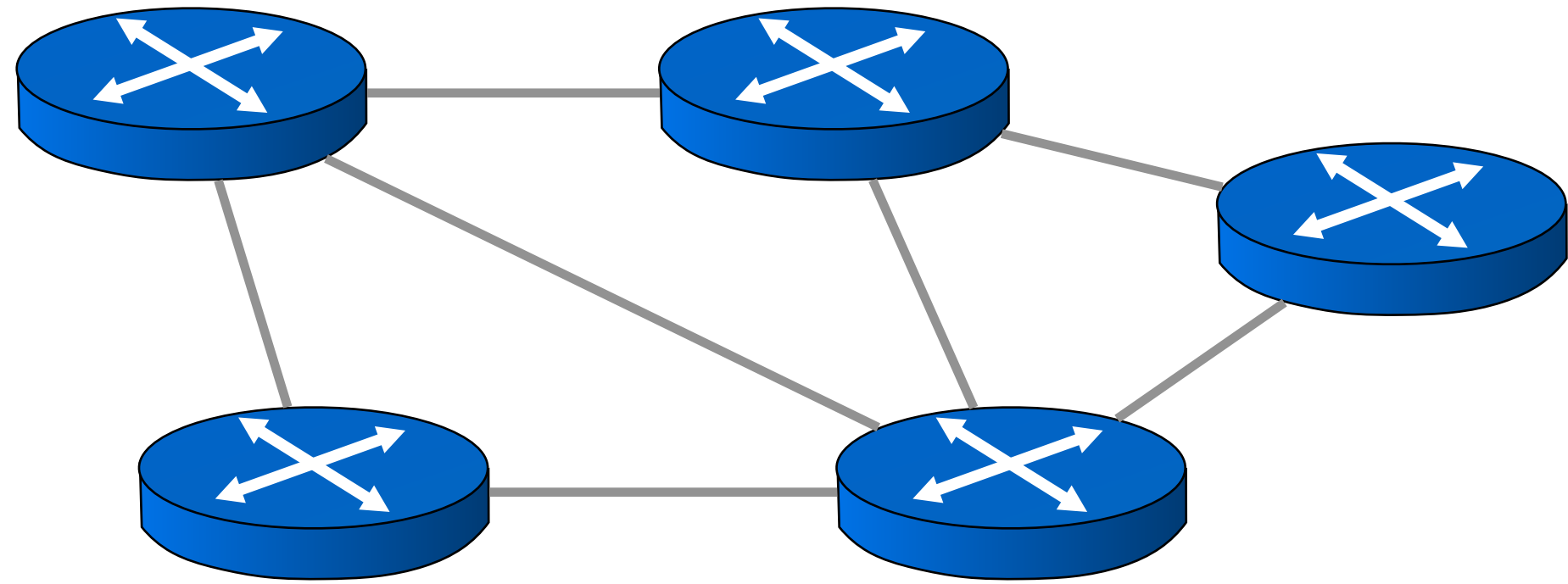
And now, the shift from art to science has begun. Researchers have been designing clean models for telemetry with elegant schemas, high-level query languages, hardware compilers, etc.

If I have seen further...



Newton

Classical Approach: Packet Histories



Packet performance stream. As part of the base input stream, which we call `pktstream`, Marple provides one tuple for each packet at each queue with the following fields.

`(switch, qid, hdrs, uid, tin, tout, qsize)`

`switch` and `qid` denote the switch and queue at which the packet was observed. A packet may traverse multiple queues even within a single switch, so we provide distinct fields. The regular packet headers (Ethernet, IP, TCP, *etc.*) are available in the `hdrs` set of fields, with a `uid` that uniquely determines a packet.¹

Marple [SIGCOMM '17, Best Paper]

Definition A packet history is the route a packet takes through a network plus the switch state and header modifications it encounters at each hop.

A single packet history can be the “smoking gun” that reveals *why*, *how*, and *where* a network failed, evidence that would otherwise remain hidden in gigabytes of message logs, flow records [8, 34], and packet dumps [15, 18, 32].

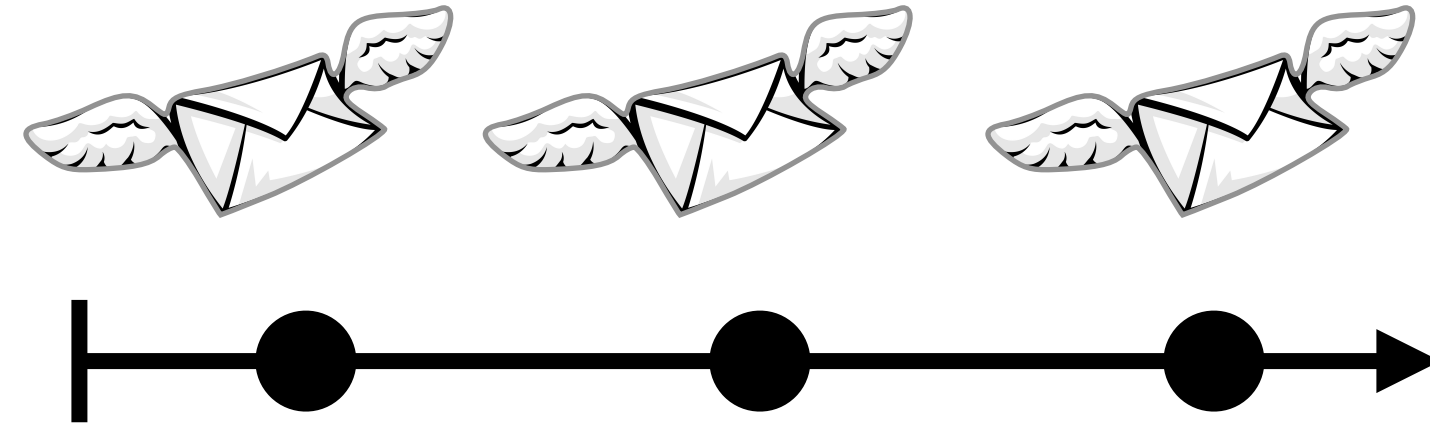
NetSight [NSDI '14]

Limitations of Status Quo

Packet histories work well enough in networks based on mostly-stateless forwarding...

But they don't adequately model complex behaviors such as:

- Stateful processing
- Control-plane reconfiguration
- Concurrent interacting packets
- End-host protocols

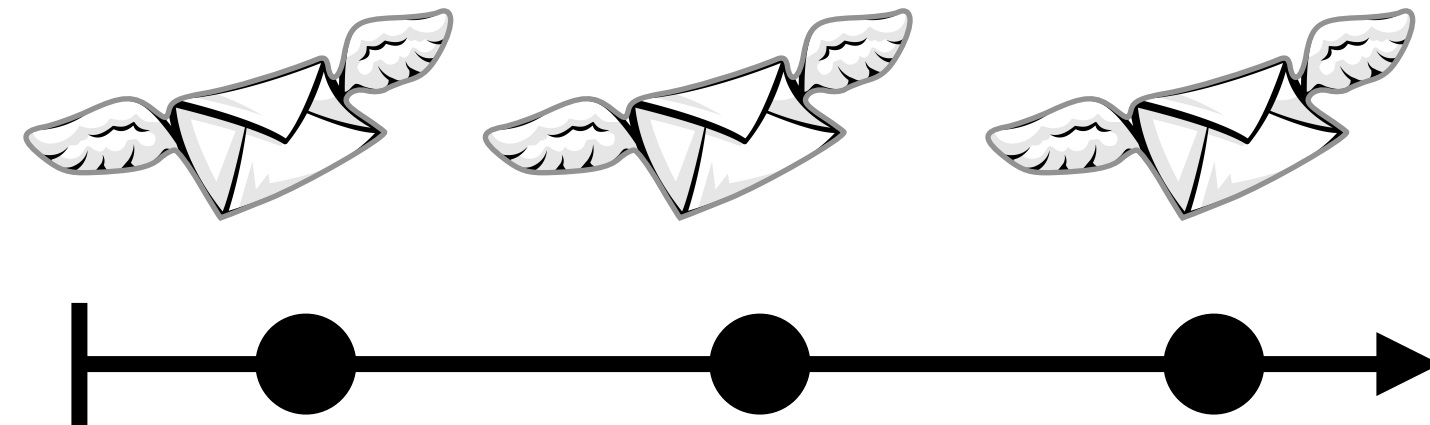


Limitations of Status Quo

Packet histories work well enough in networks based on mostly-stateless forwarding...

But they don't adequately model complex behaviors such as:

- Stateful processing
- Control-plane reconfiguration
- Concurrent interacting packets
- End-host protocols



Example: was a change to a forwarding rule at switch A responsible for the delays at switch B, or was it the delays themselves that prompted the rule change?

Causal Telemetry

A new model for network telemetry based a *relativistic* account of events in the underlying distributed system. It captures relationships involving multiple packets, across the data plane and control plane, etc.



Causal Telemetry

A new model for network telemetry based a *relativistic* account of events in the underlying distributed system. It captures relationships involving multiple packets, across the data plane and control plane, etc.



Why Causality?

Packet Histories

Fail to capture interesting relations between events

Global Snapshots

Packets often “see” more than one global state

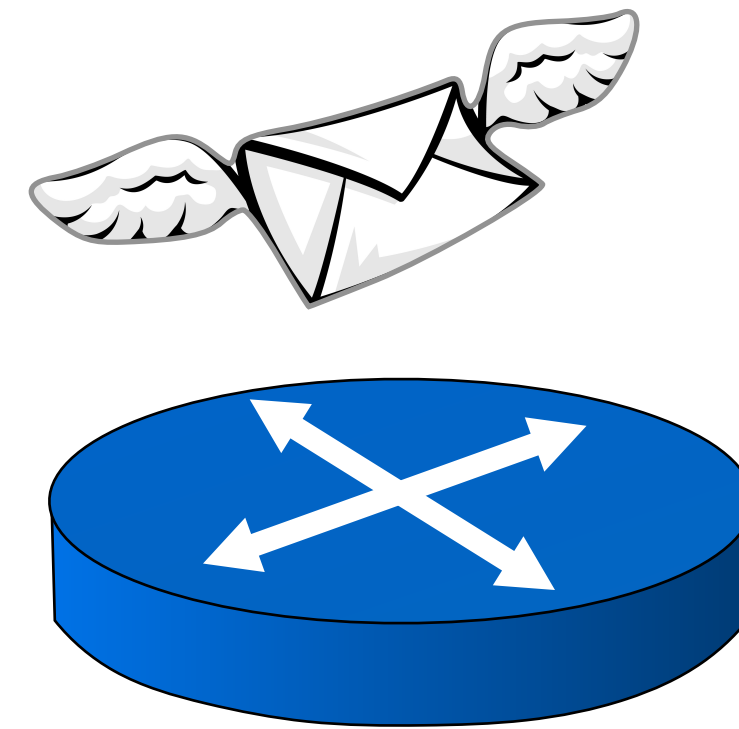
Synchronized Clocks

- Impossible to fully achieve
- Total ordering: correlation not causality
- Still need to stitch together multiple events

Causal Telemetry: Semantic Model

Events:

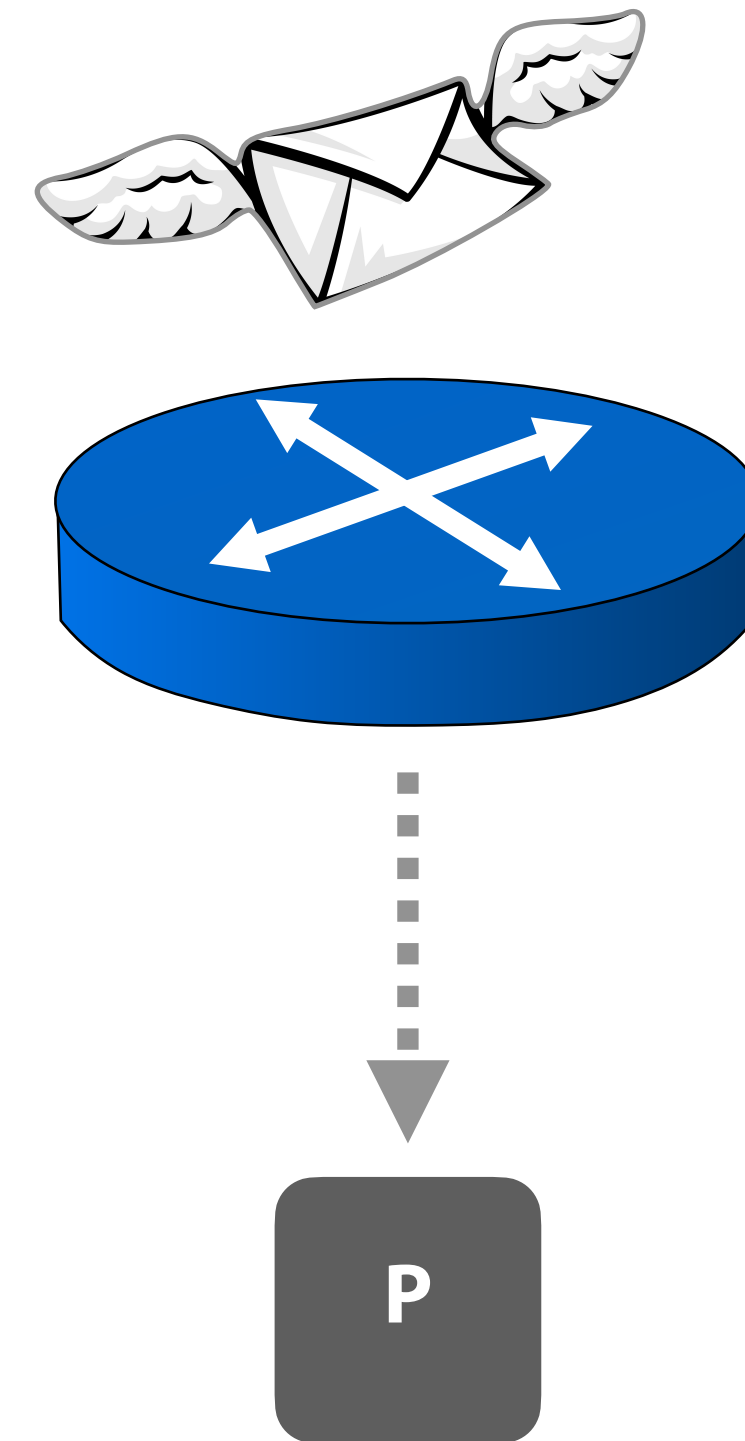
- For each packet processed



Causal Telemetry: Semantic Model

Events:

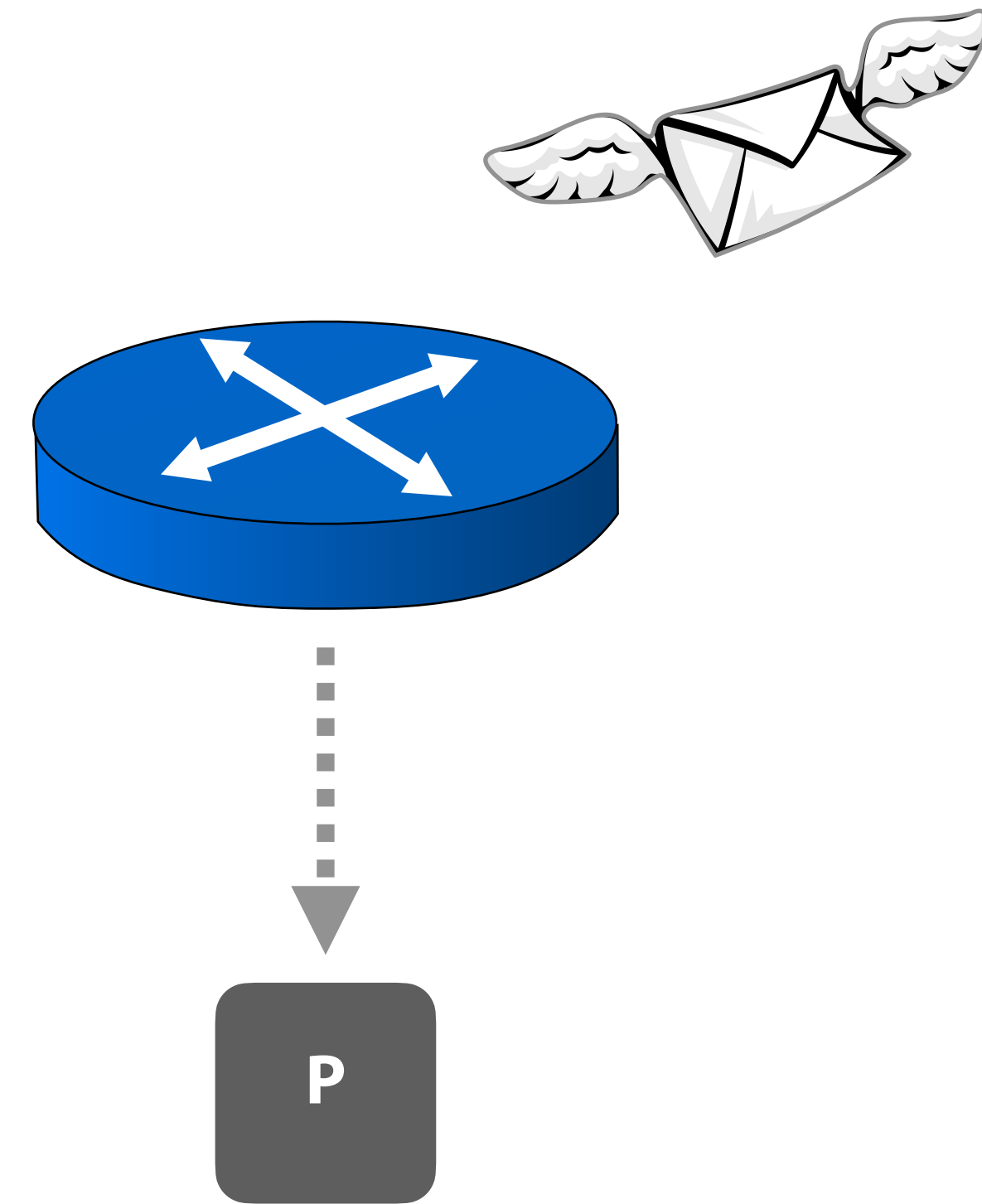
- For each packet processed



Causal Telemetry: Semantic Model

Events:

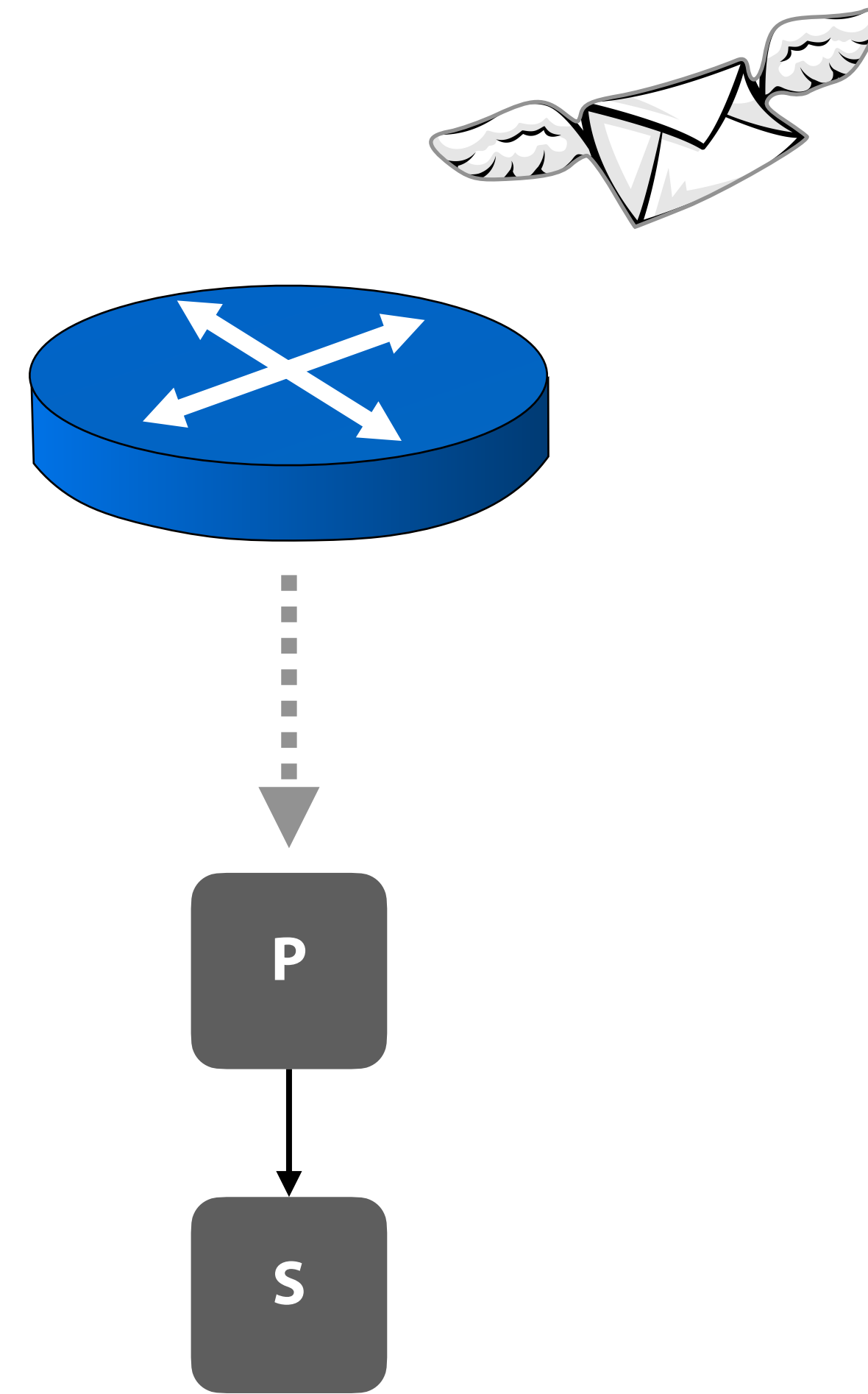
- For each packet processed
- For each packet sent



Causal Telemetry: Semantic Model

Events:

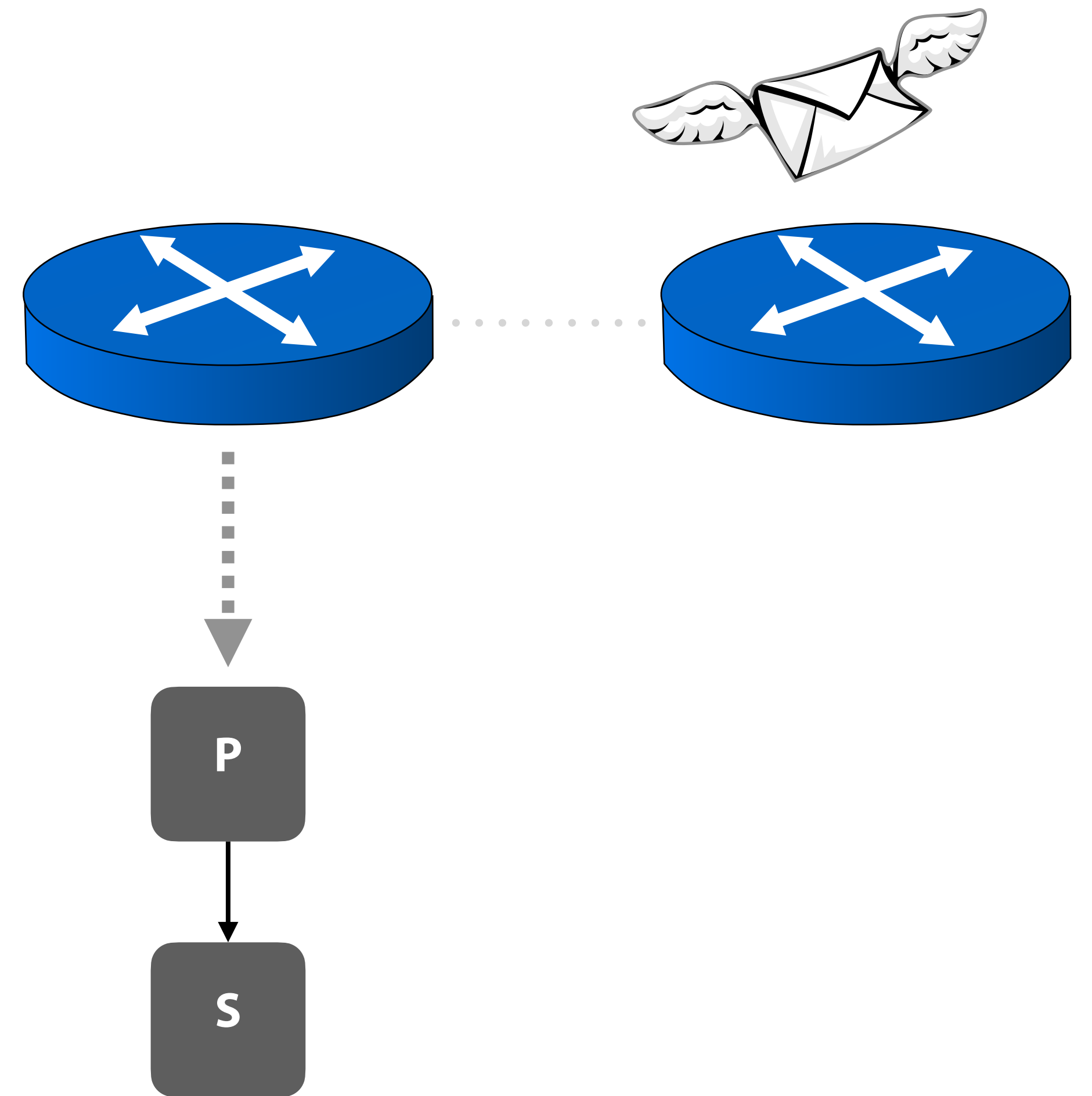
- For each packet processed
- For each packet sent



Causal Telemetry: Semantic Model

Events:

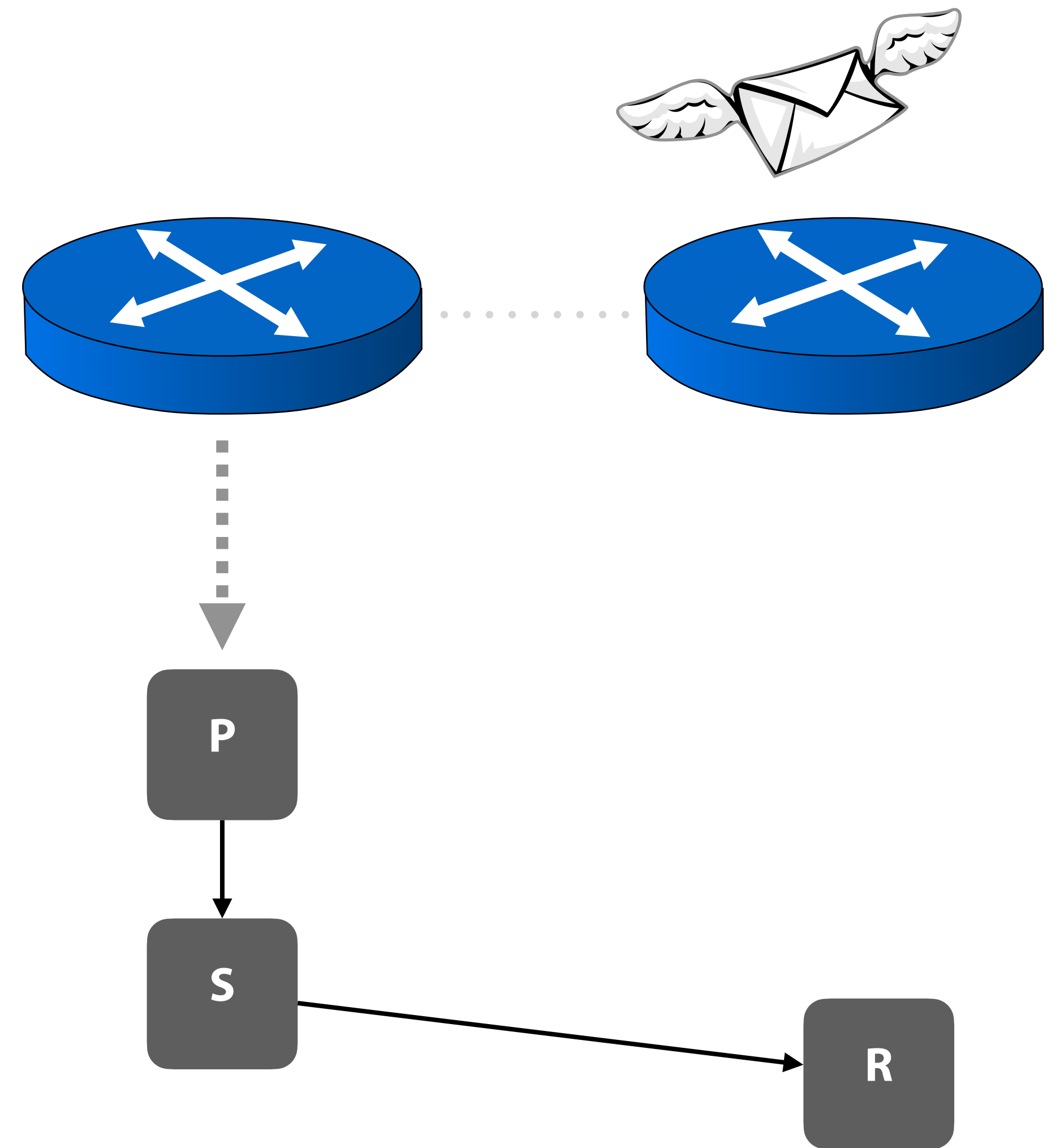
- For each packet processed
- For each packet sent
- For each packet received



Causal Telemetry: Semantic Model

Events:

- For each packet processed
- For each packet sent
- For each packet received



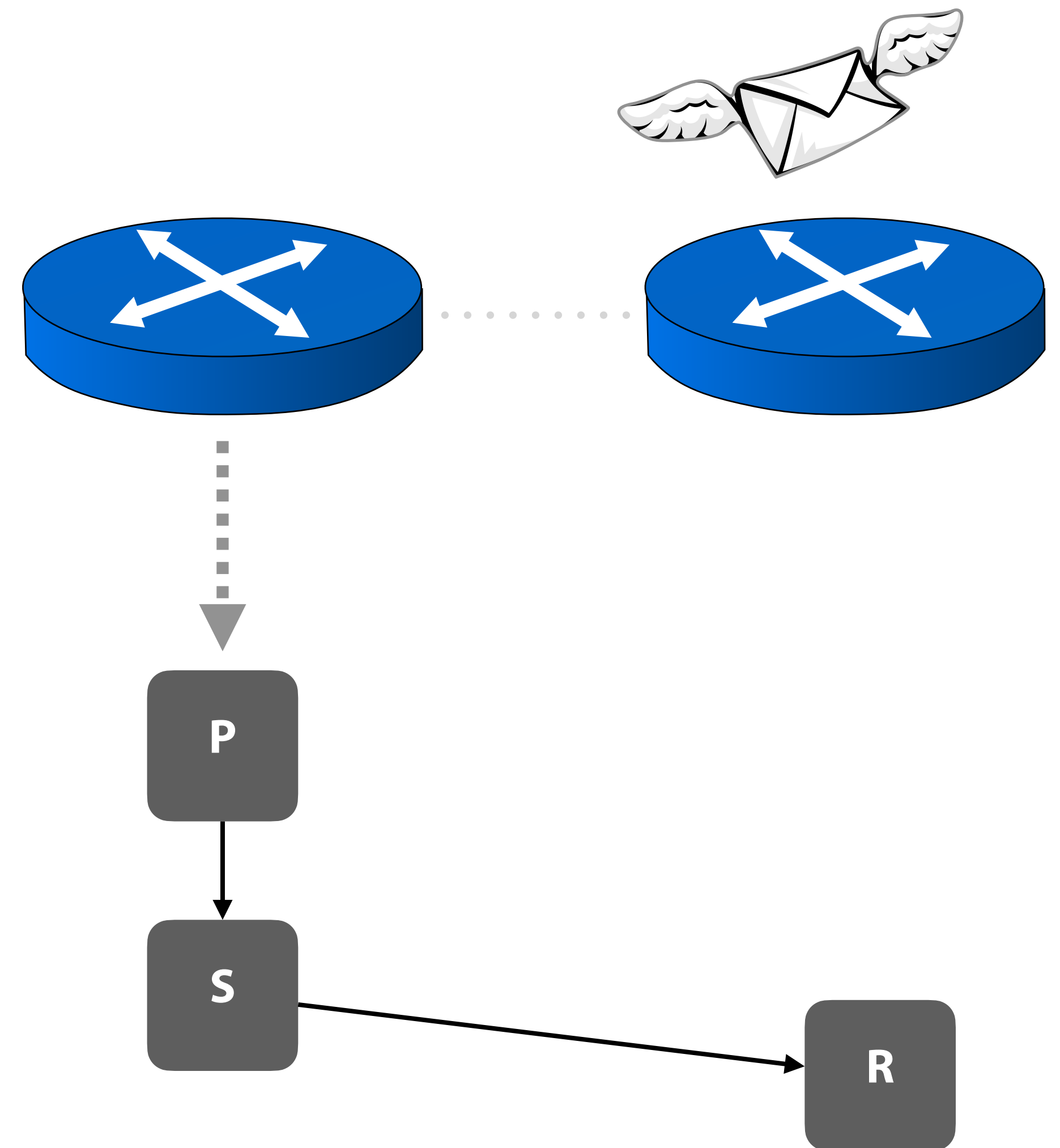
Causal Telemetry: Semantic Model

Events:

- For each packet processed
- For each packet sent
- For each packet received

Edges:

- Adjacent events at each node
- From send to corresponding receive



Causal Telemetry: Semantic Model

Events:

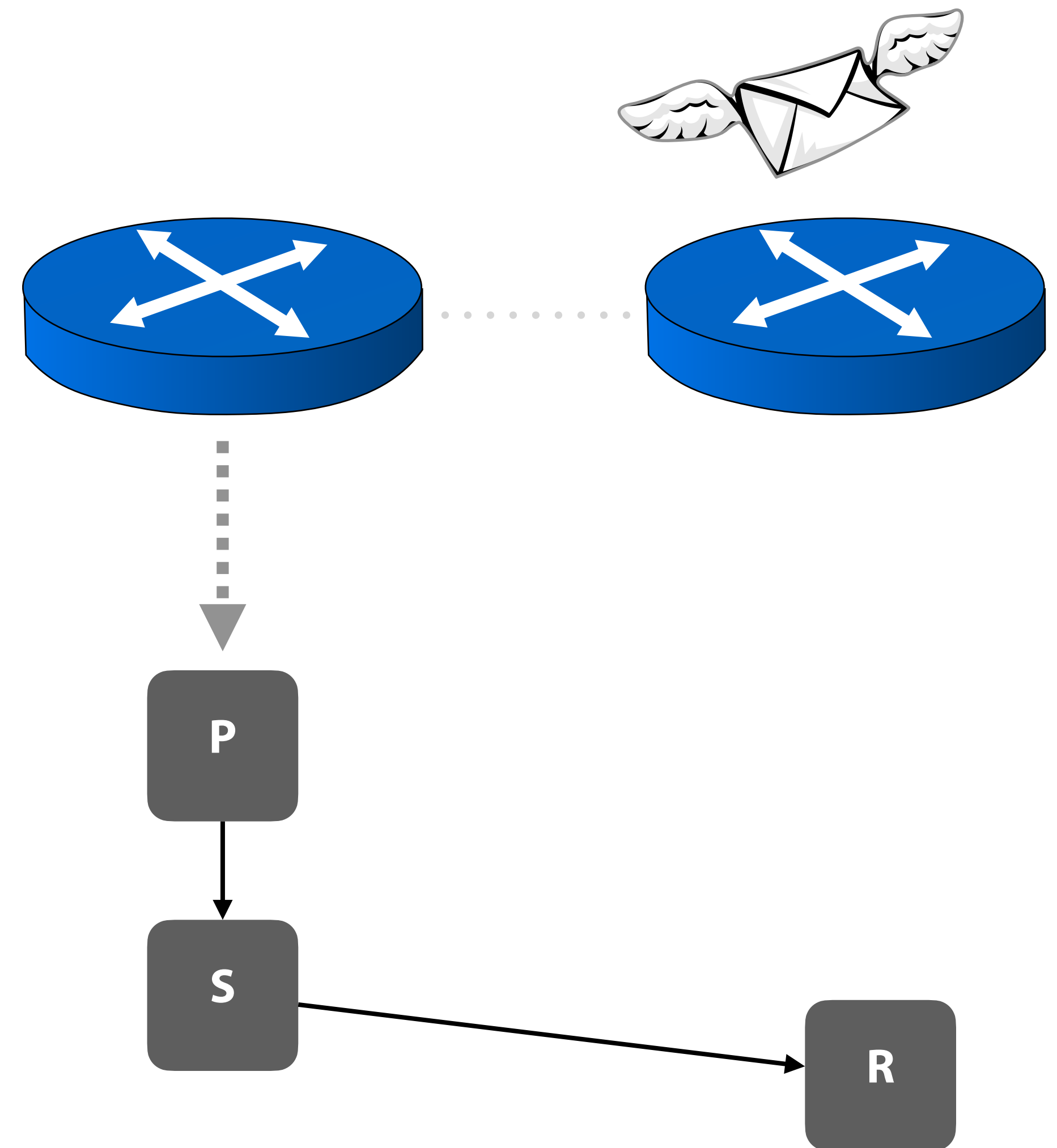
- For each packet processed
- For each packet sent
- For each packet received

Edges:

- Adjacent events at each node
- From send to corresponding receive

See paper:

- Control plane, end hosts, etc.
- Event data
- Pipeline granularity



Expressiveness & Causal Queries

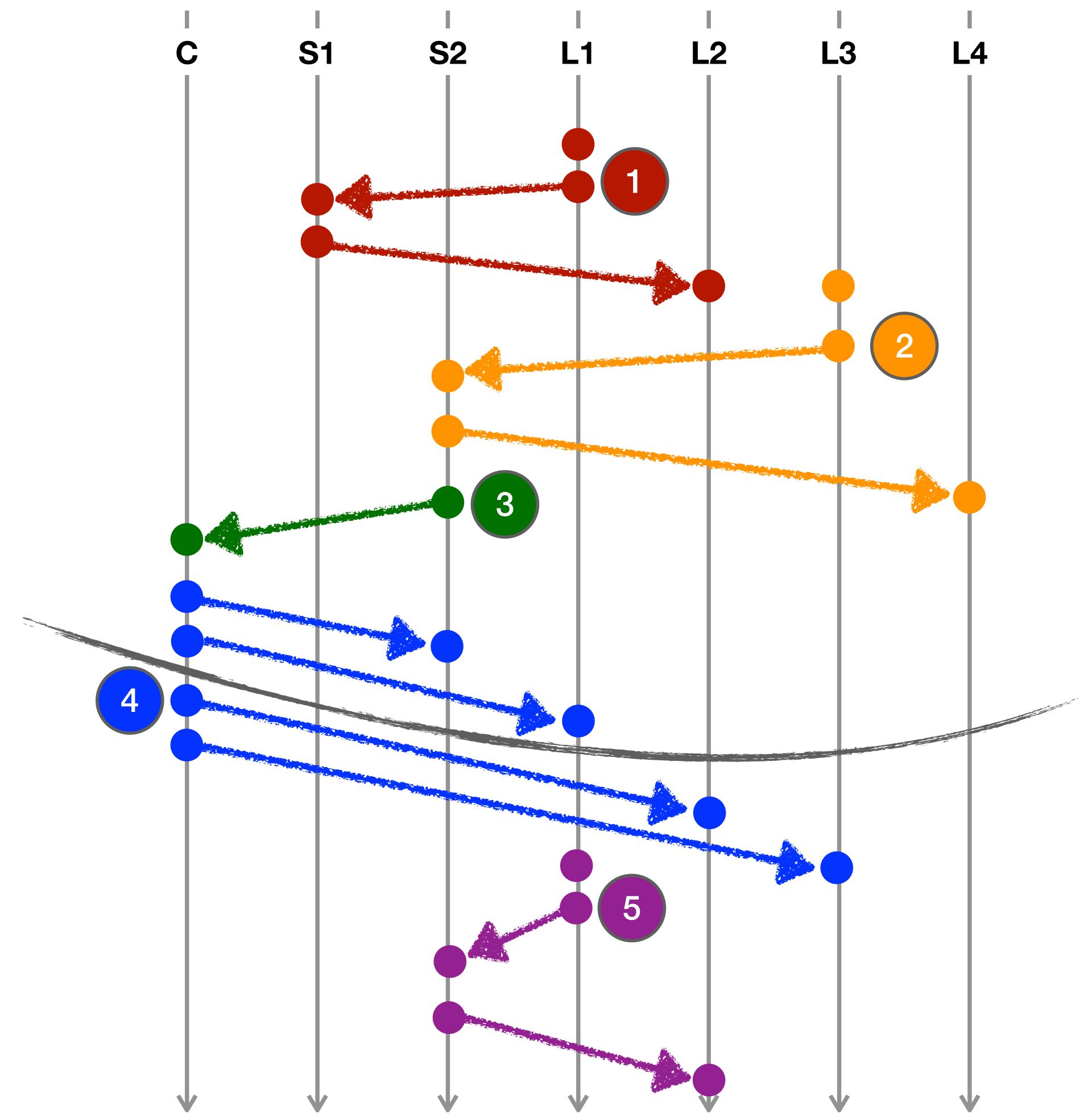
Expressiveness

It's easy to see that space-time diagram is expressive, as data model used by other telemetry systems can be "read off" the graph:

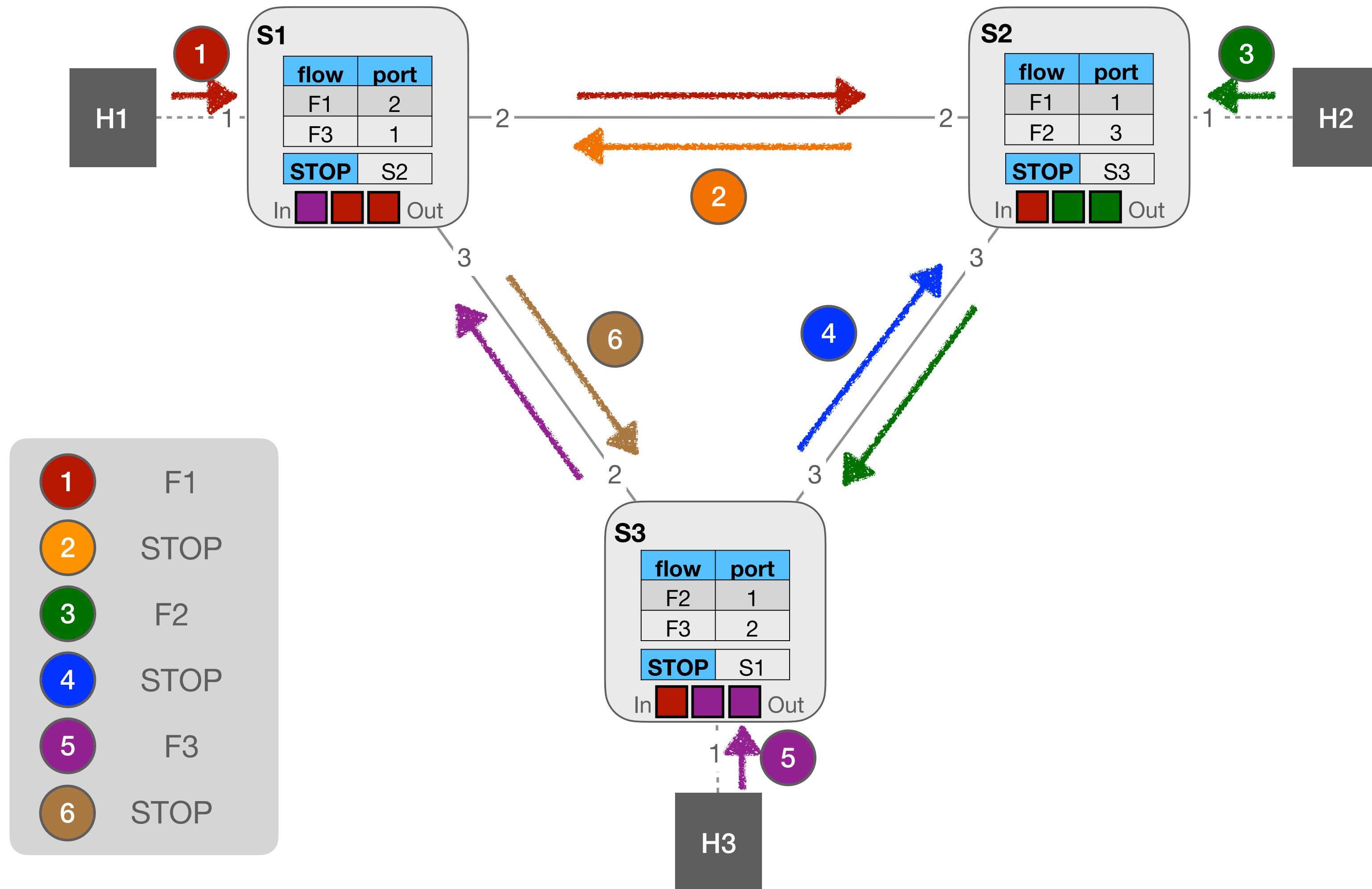
- Stream of packets at each switch
Sonata [SIGCOMM '18]
- Histories for each data-plane packets
NetSight [NSDI '14, Marple [SIGCOMM '17]

Causal Queries

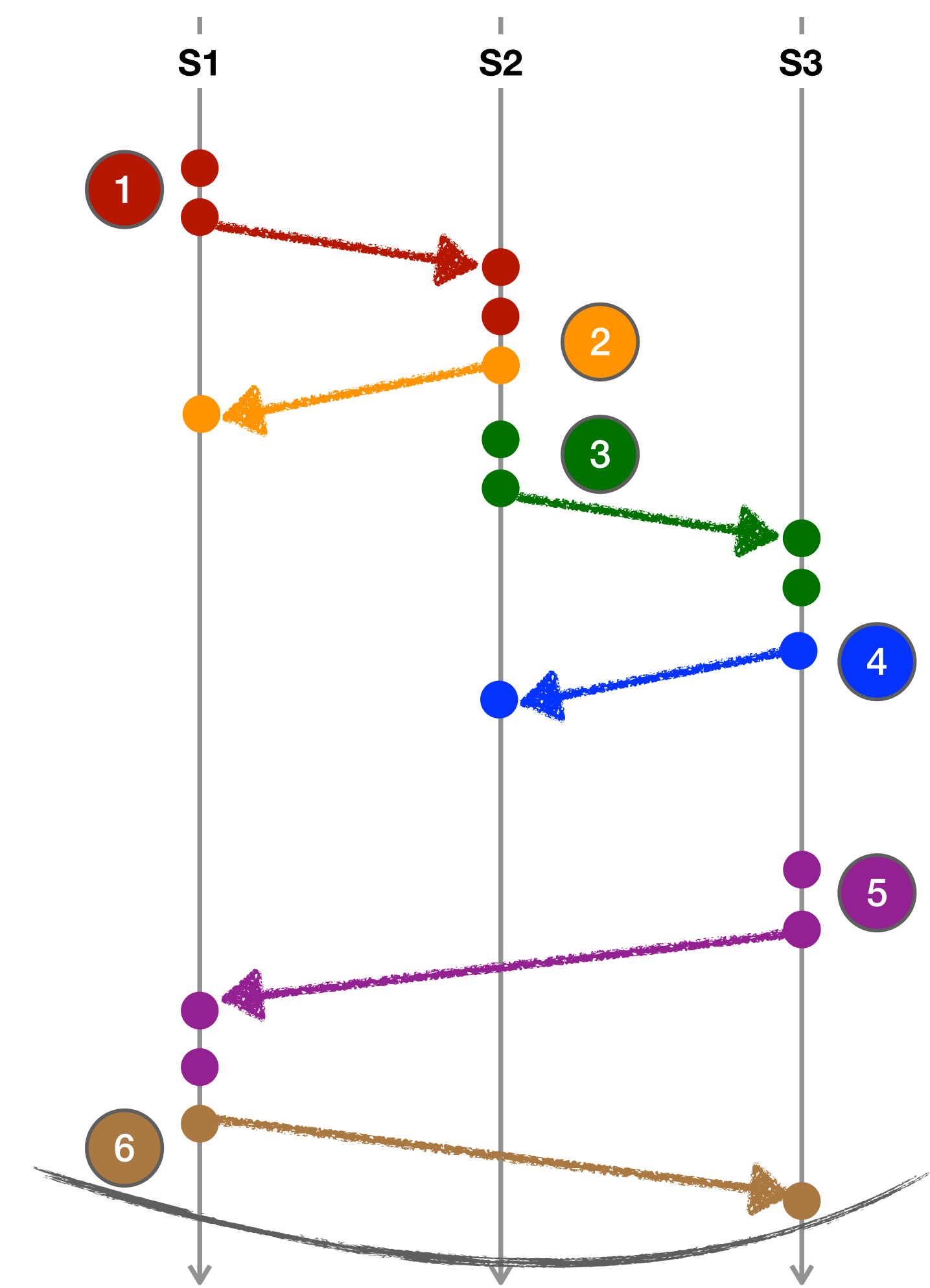
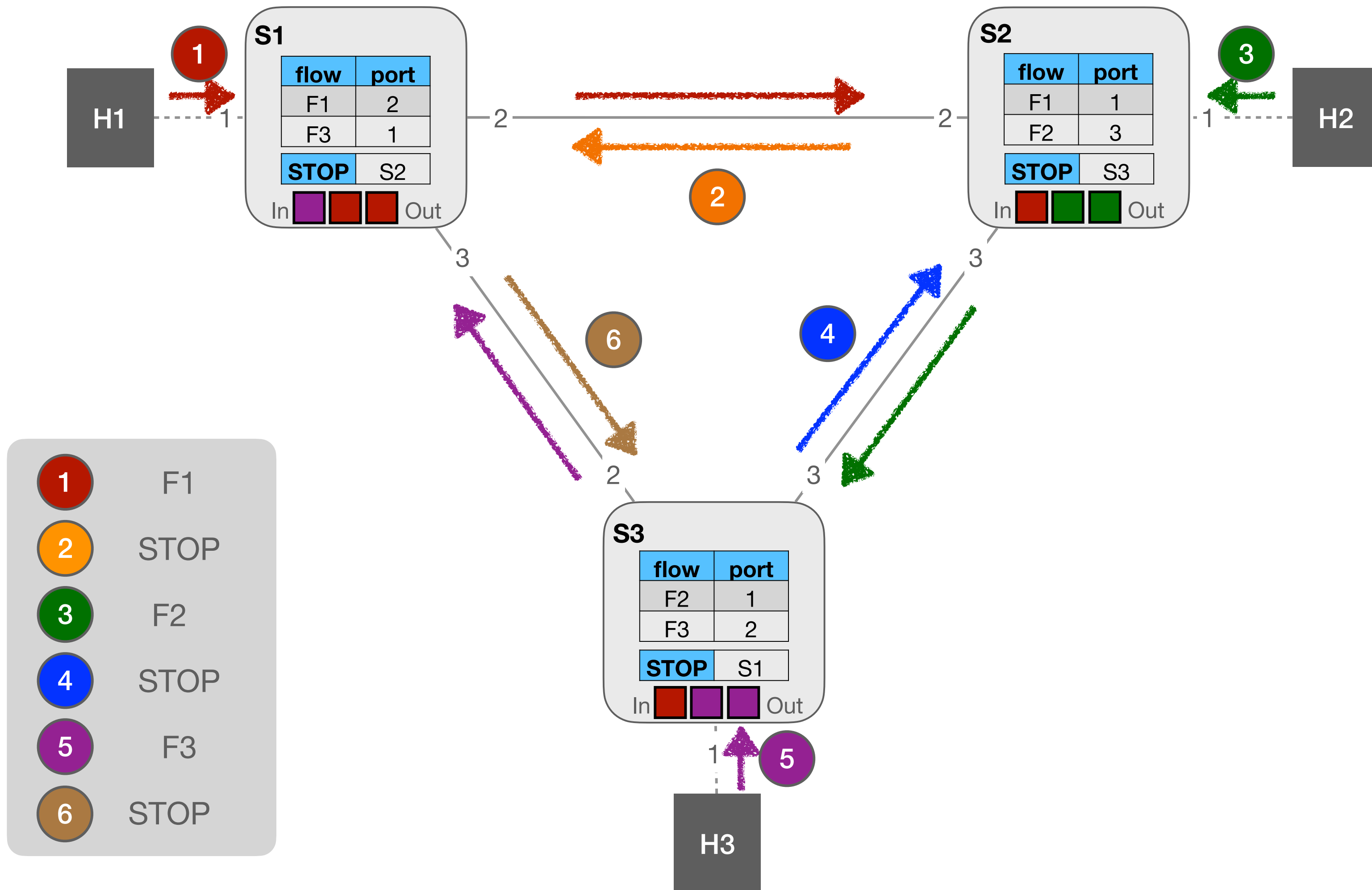
Space-time graphs also supports other queries, like enumerating the consistent cuts



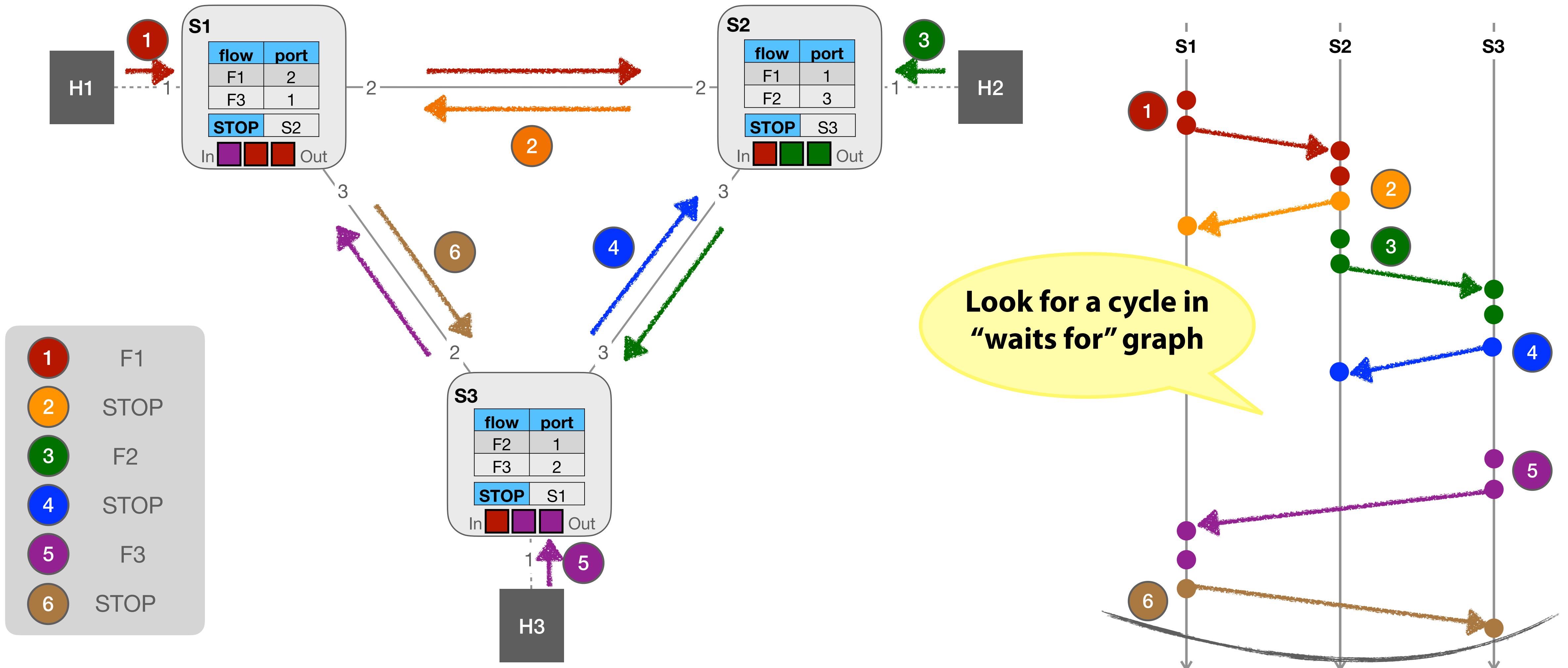
Case Study: PFC Deadlock



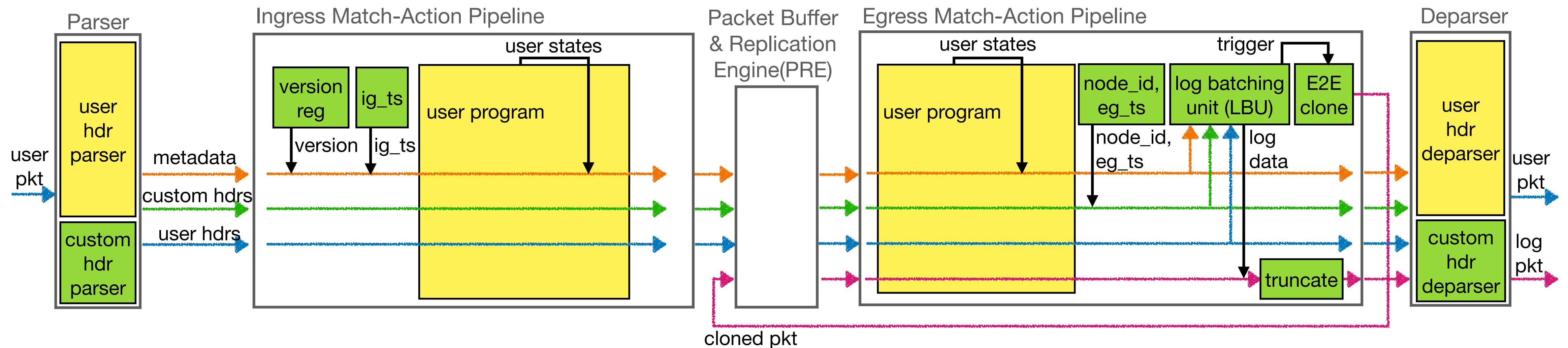
Case Study: PFC Deadlock



Case Study: PFC Deadlock



P4 Implementation (BMv2)



Highlights

- Timestamp for each pipeline (register)
- Version register for control-plane (register)
- Encode event identifier in packets (custom header, parser, deparser)
- Note: this is *not* logical clocks—see paper for details

Ongoing Work

Domain-Specific Language

Customize data collection; support multi-tenancy

In-Network Processing

Batching, filtering, and aggregation of telemetry records

Handling Failures

Enhance model with better support for dropped packets

Security

Specify and enforce policies related to use of telemetry

Relativity teaches us the connection between the different descriptions of one and the same reality.

-Einstein

Thank You!



Lead author Yunhe Liu is graduating this spring and looking for industry positions!

Acknowledgments

- Griffin Berlstein
- Cornell NetLab
- Pronto Project
- EuroP4 Reviewers

I'm hiring! (PhDs, postdocs, P4 h4x0rs)

