PIQ: Persistent Interactive Queries for Network Security Analytics

ACM SDN-NFV Security '19, March 27th 2019

Oliver Michel
John Sonchack
Eric Keller
Jonathan M. Smith

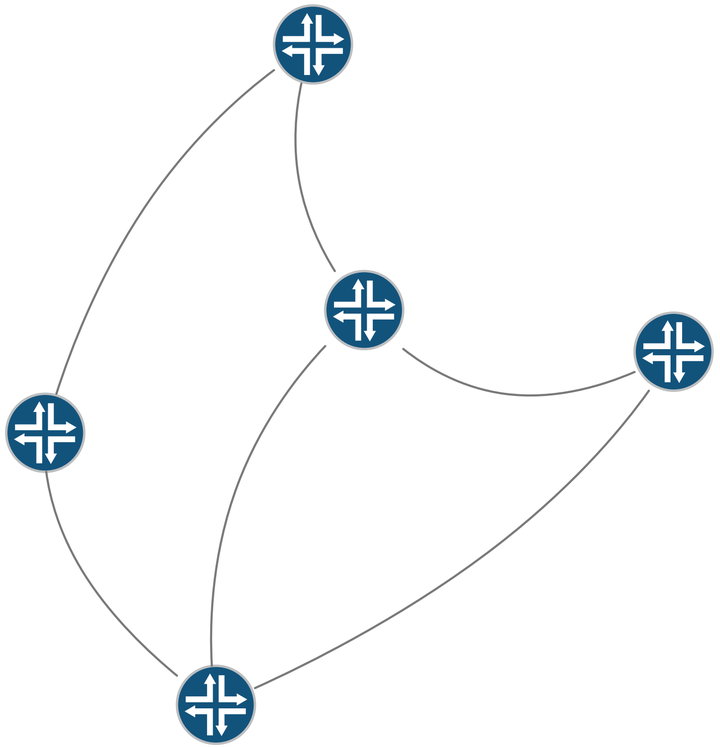
University of Colorado Boulder

Penn

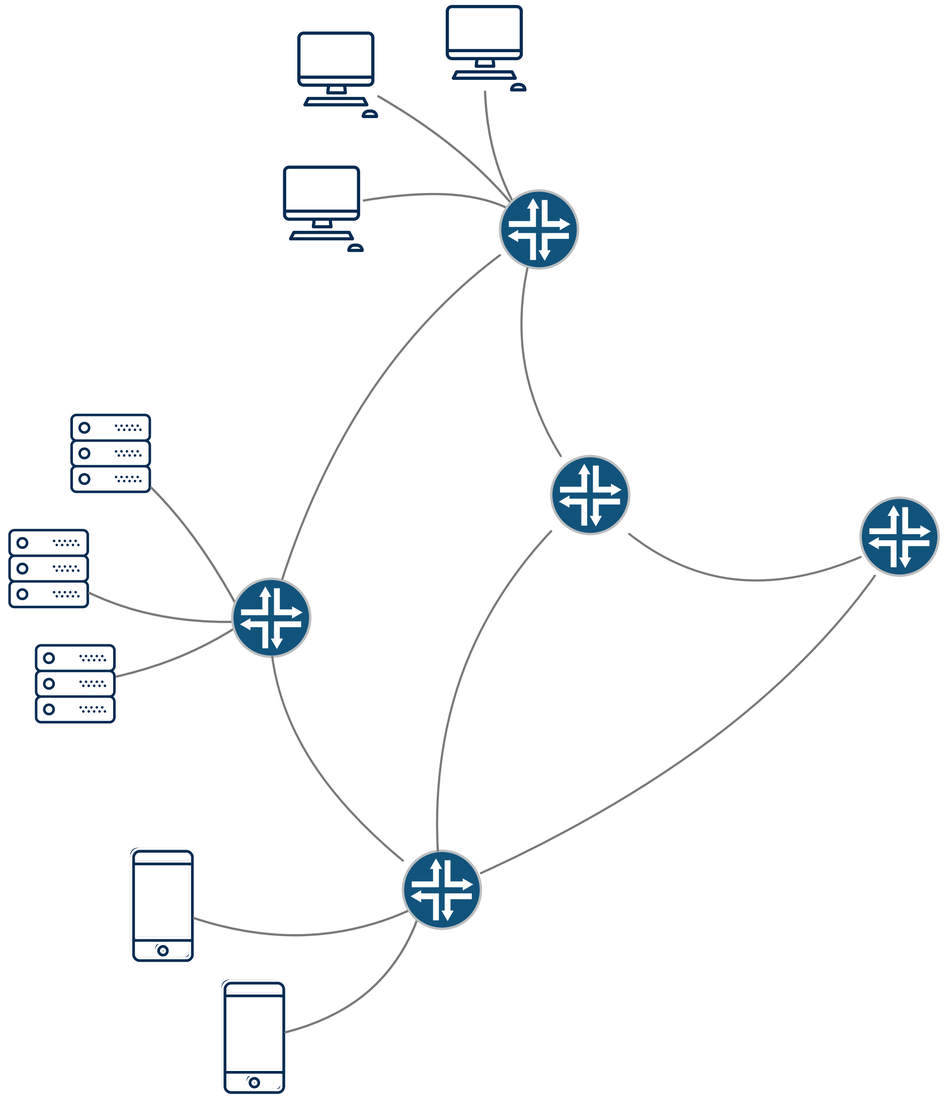
Network Monitoring is important

- Security issues
- Performance issues
- Equipment failure
- Misconfiguration

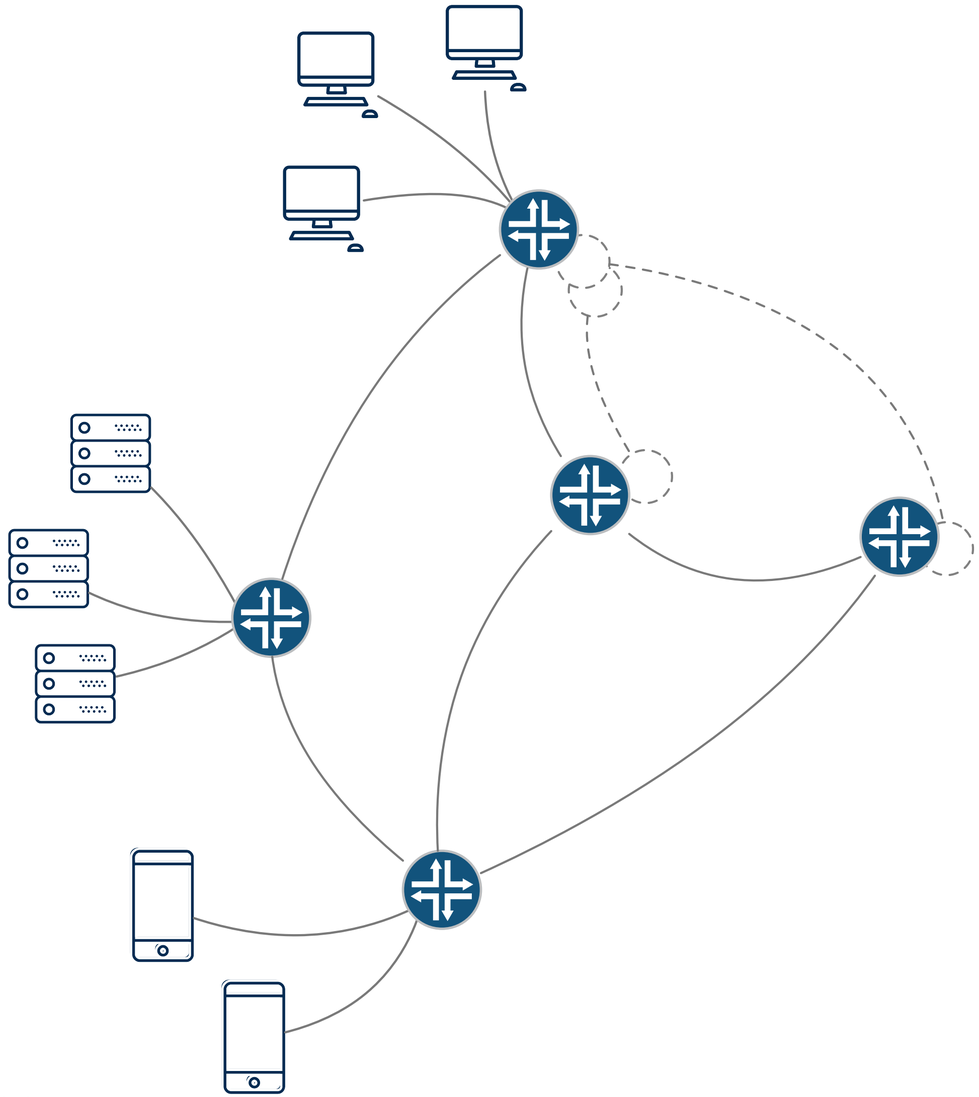
Today’s networks are larger and more complex than ever before



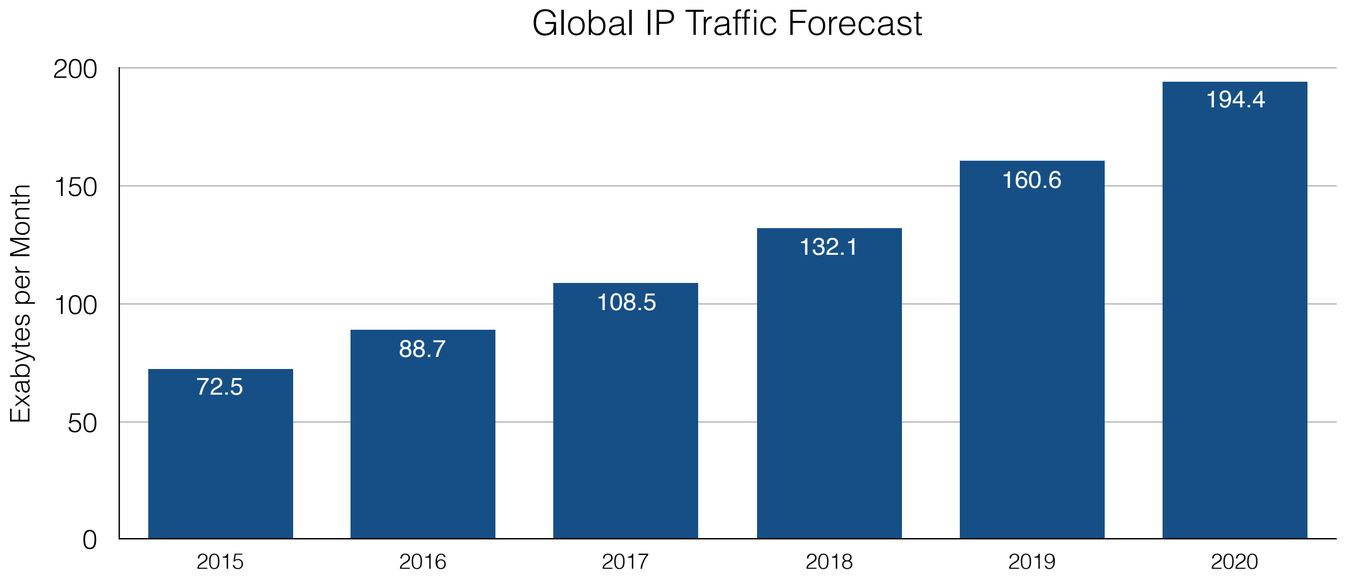
Today’s networks are larger and more complex than ever before



Today’s networks are larger and more complex than ever before

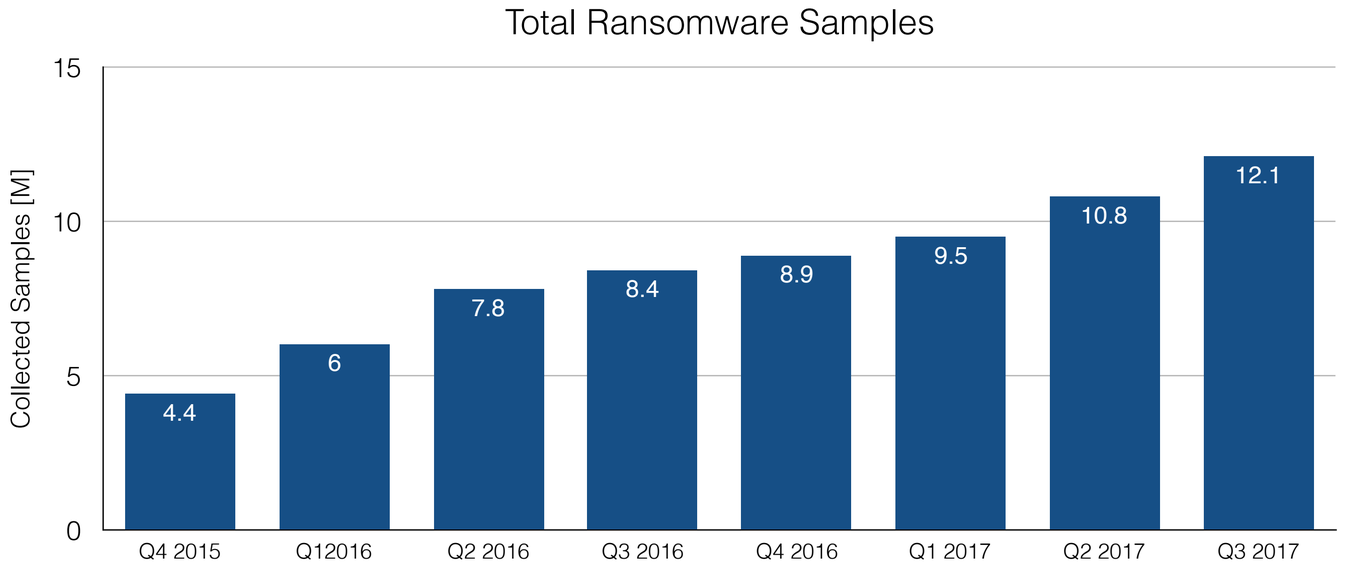


Today’s networks are larger and more complex than ever before



[Cisco Visual Networking Index 2017]

Network Threats grow rapidly



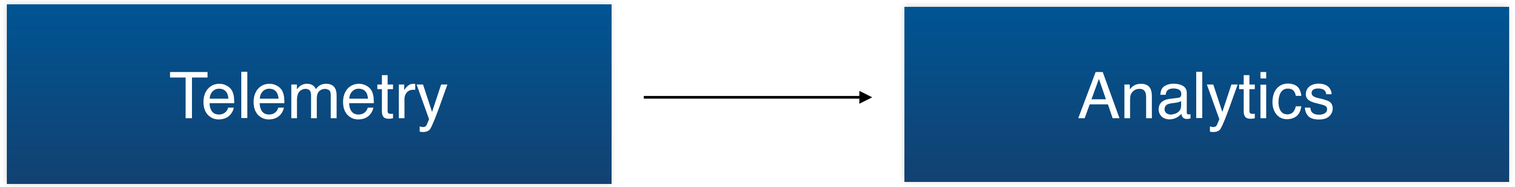
[McAfee Labs Thread Report Dec. 2017]

Passive Monitoring

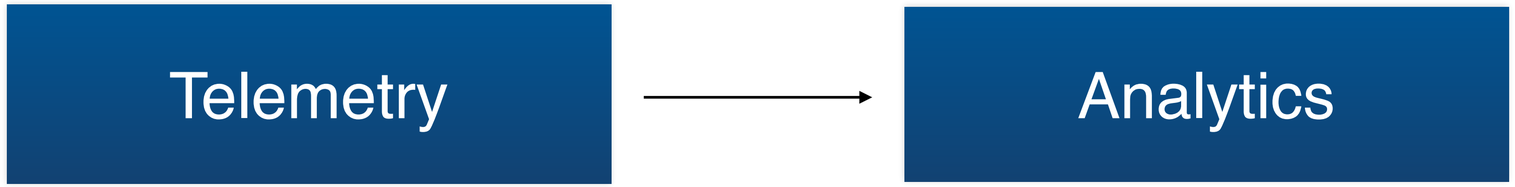
Passive Monitoring

Telemetry

Passive Monitoring

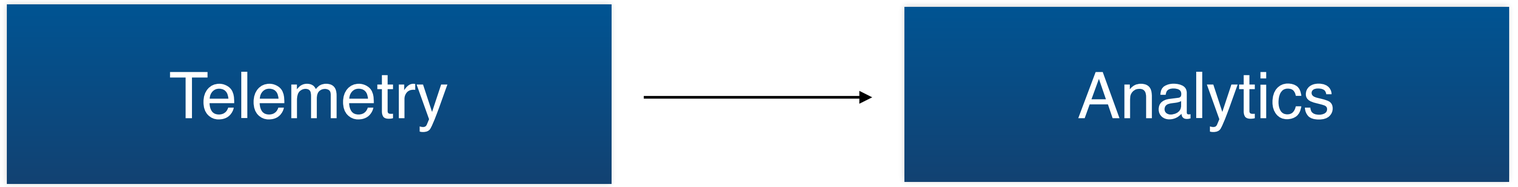


Passive Monitoring



[UnivMon, SIGCOMM ’16]
[Marple, SIGCOMM ’17]
[*Flow, ATC ’18]

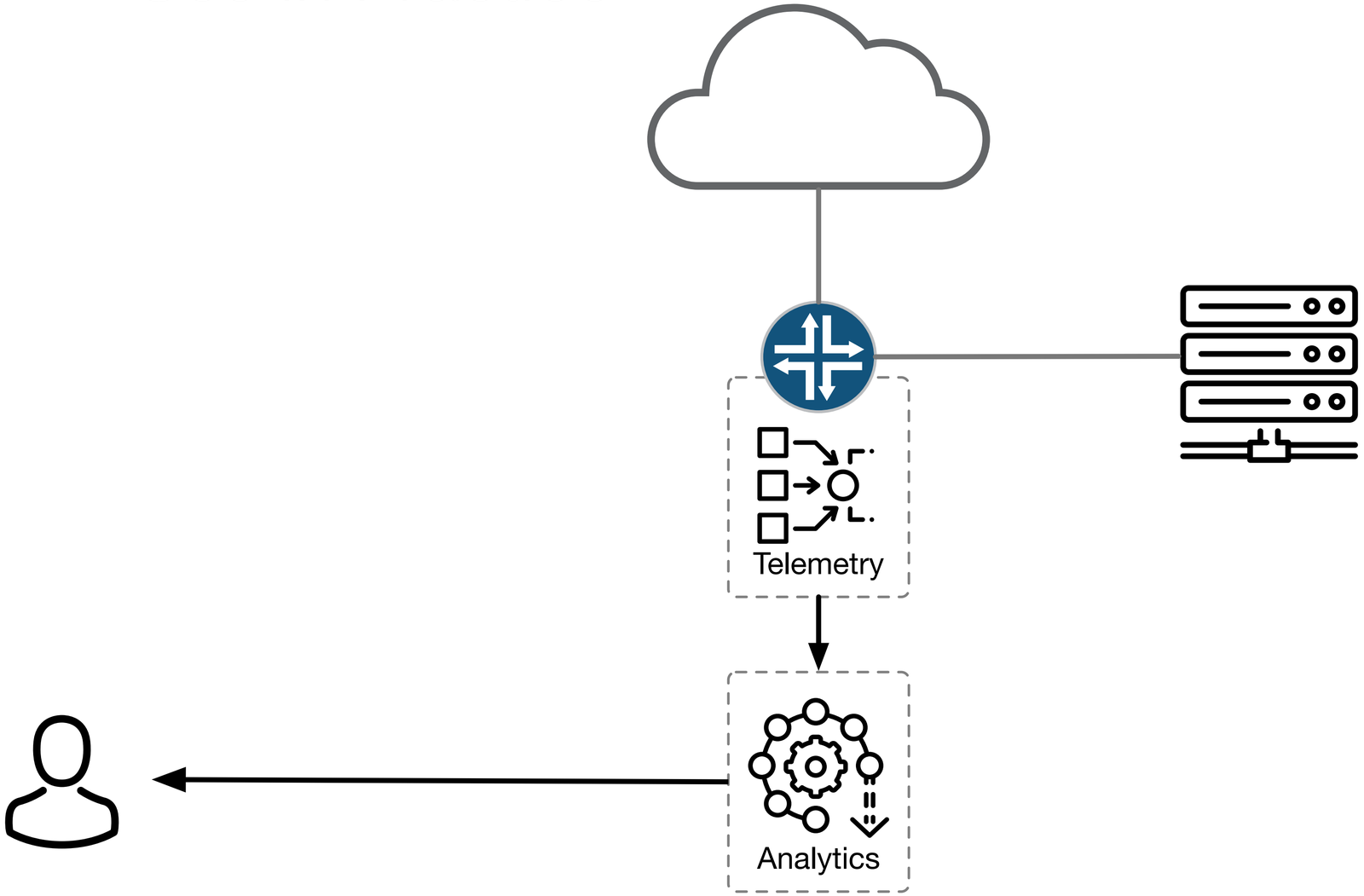
Passive Monitoring



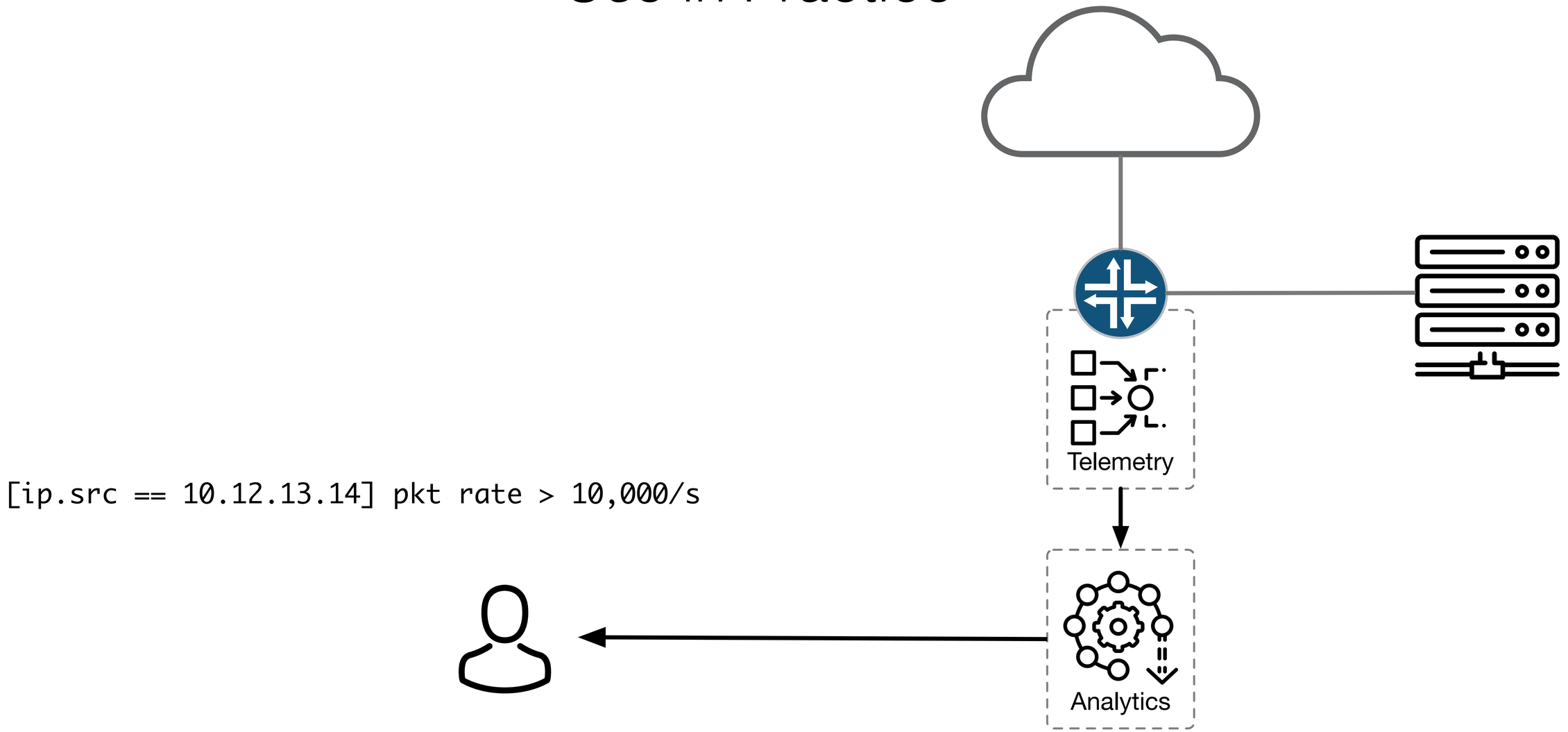
[UnivMon, SIGCOMM ’16]
[Marple, SIGCOMM ’17]
[*Flow, ATC ’18]

[NetQRE, SIGCOMM ’17]
[Jetstream, HotCloud '18]
[Sonata, SIGCOMM ’18]

Use in Practice

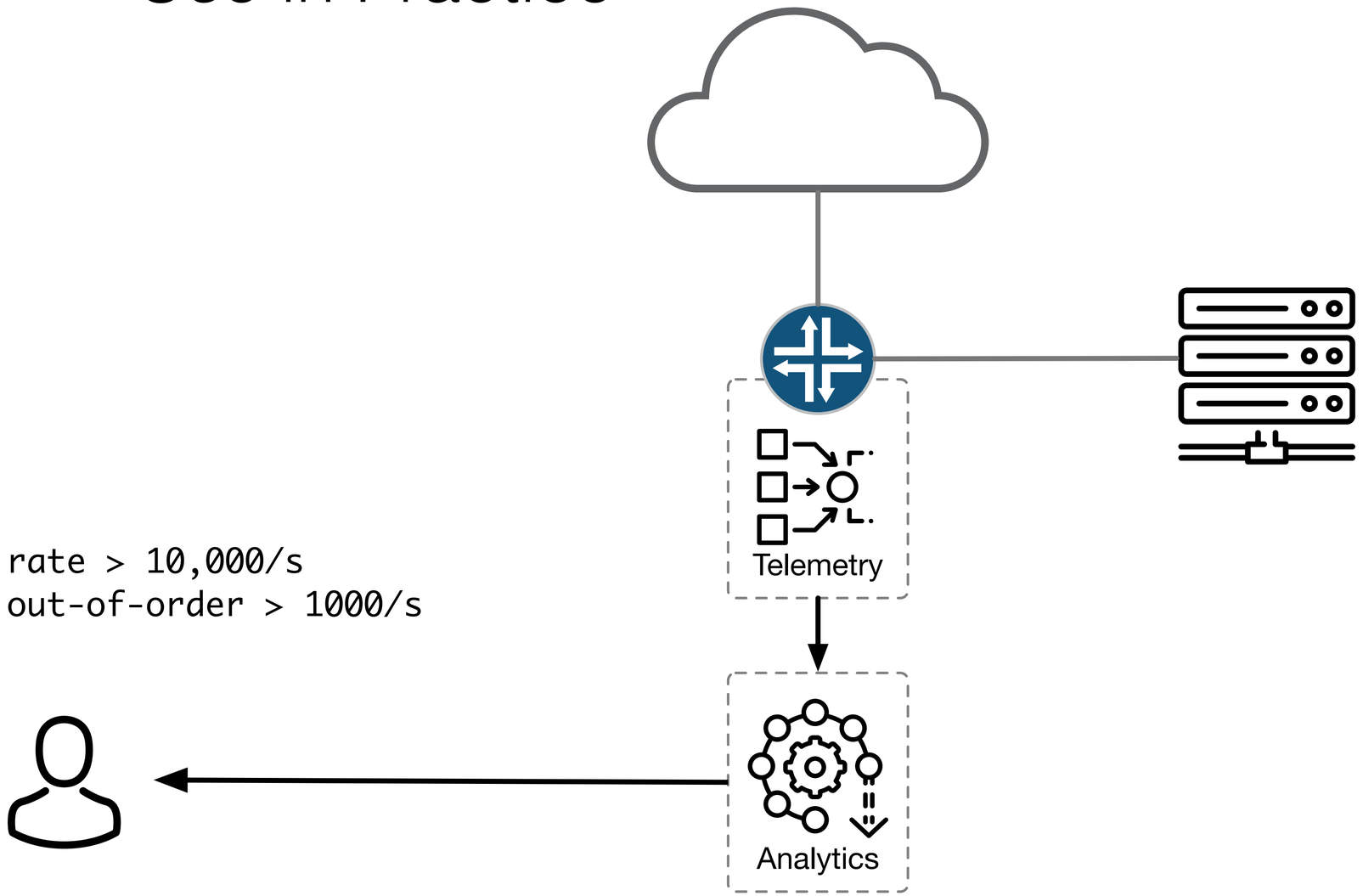


Use in Practice



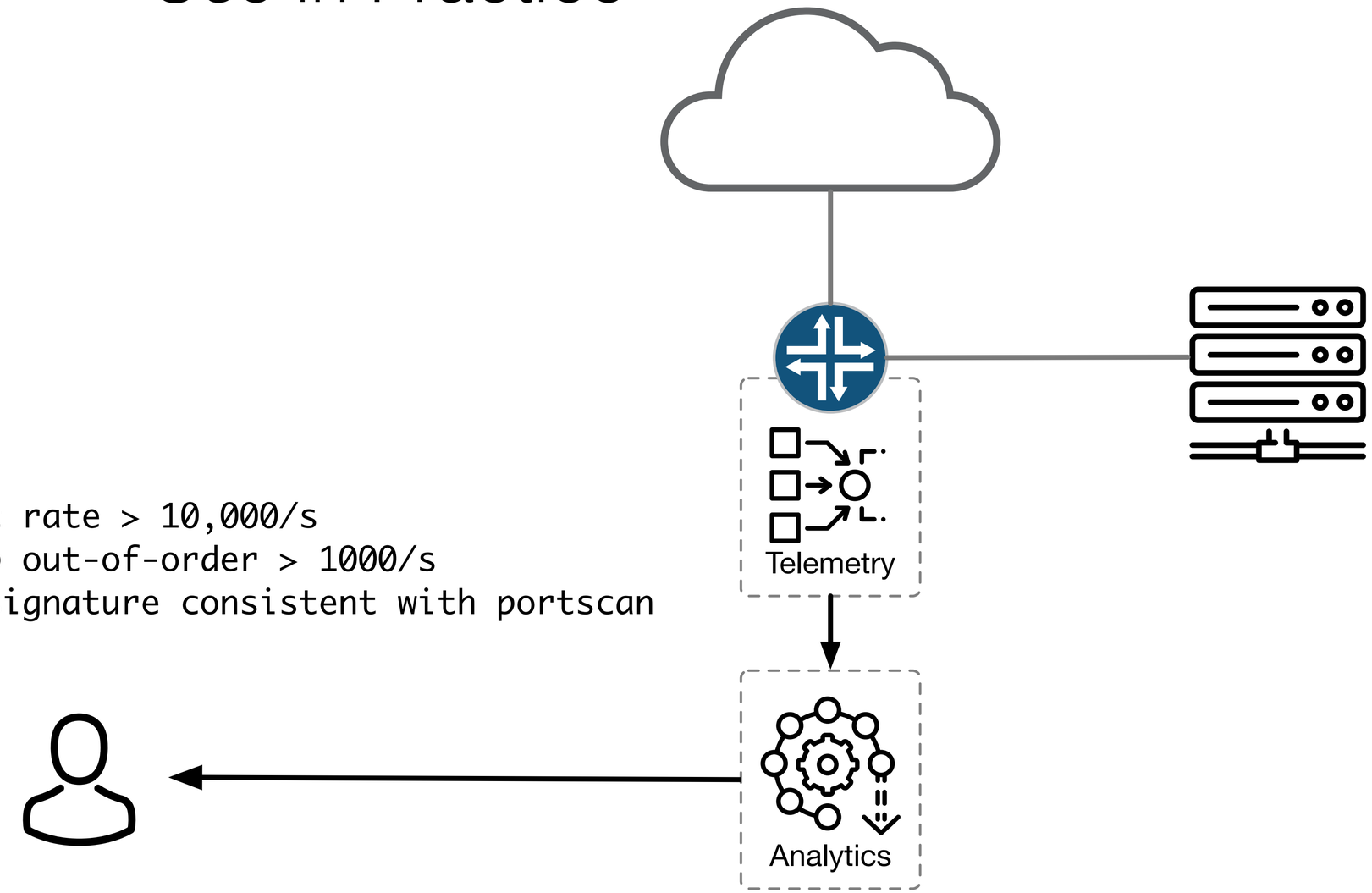
Use in Practice

```
[ip.src == 10.12.13.14] pkt rate > 10,000/s
[ip.dst == 10.12.13.14] tcp out-of-order > 1000/s
```



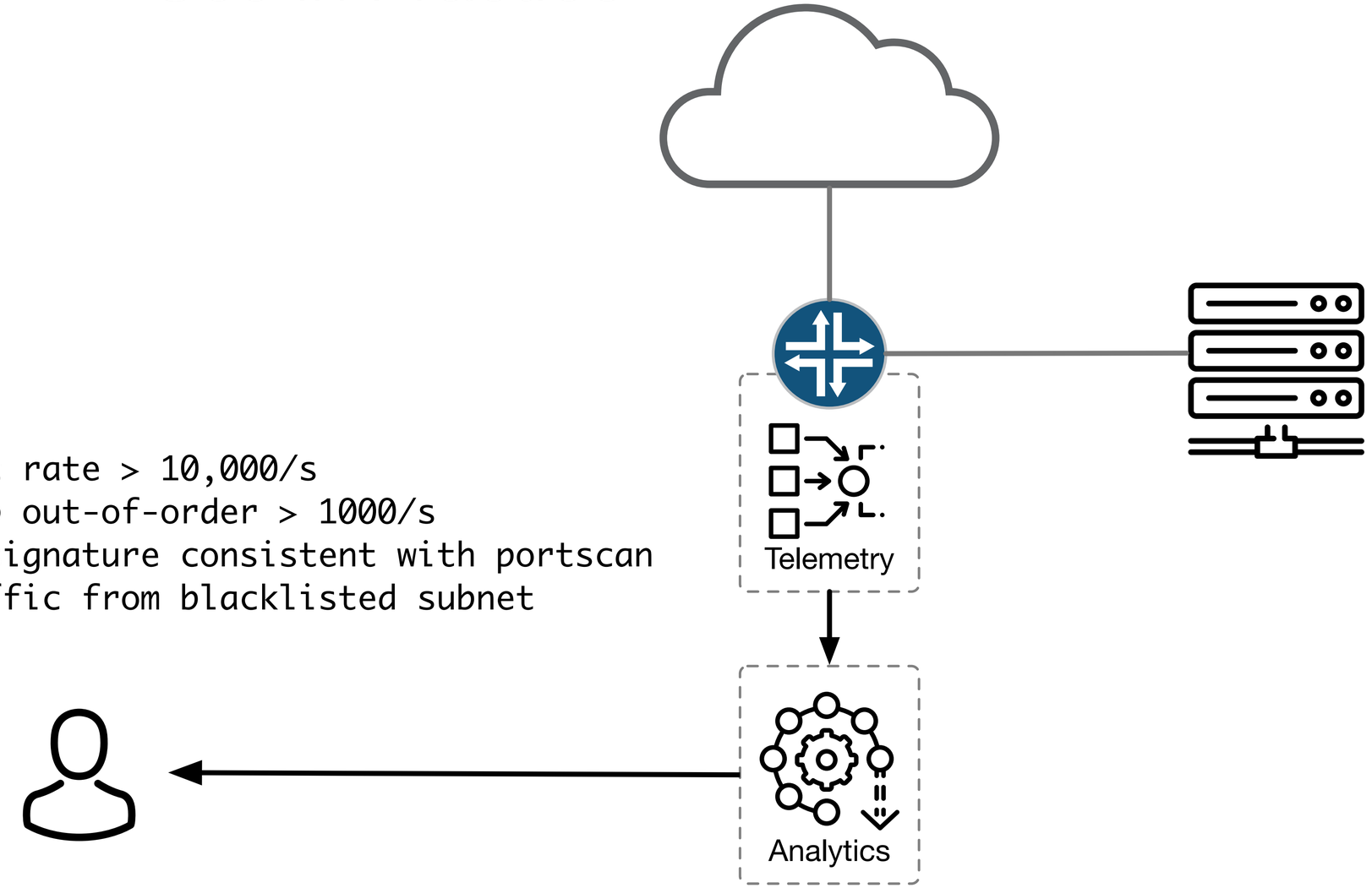
Use in Practice

```
[ip.src == 10.12.13.14] pkt rate > 10,000/s
[ip.dst == 10.12.13.14] tcp out-of-order > 1000/s
[ip.src == 239.232.230.2] signature consistent with portscan
```



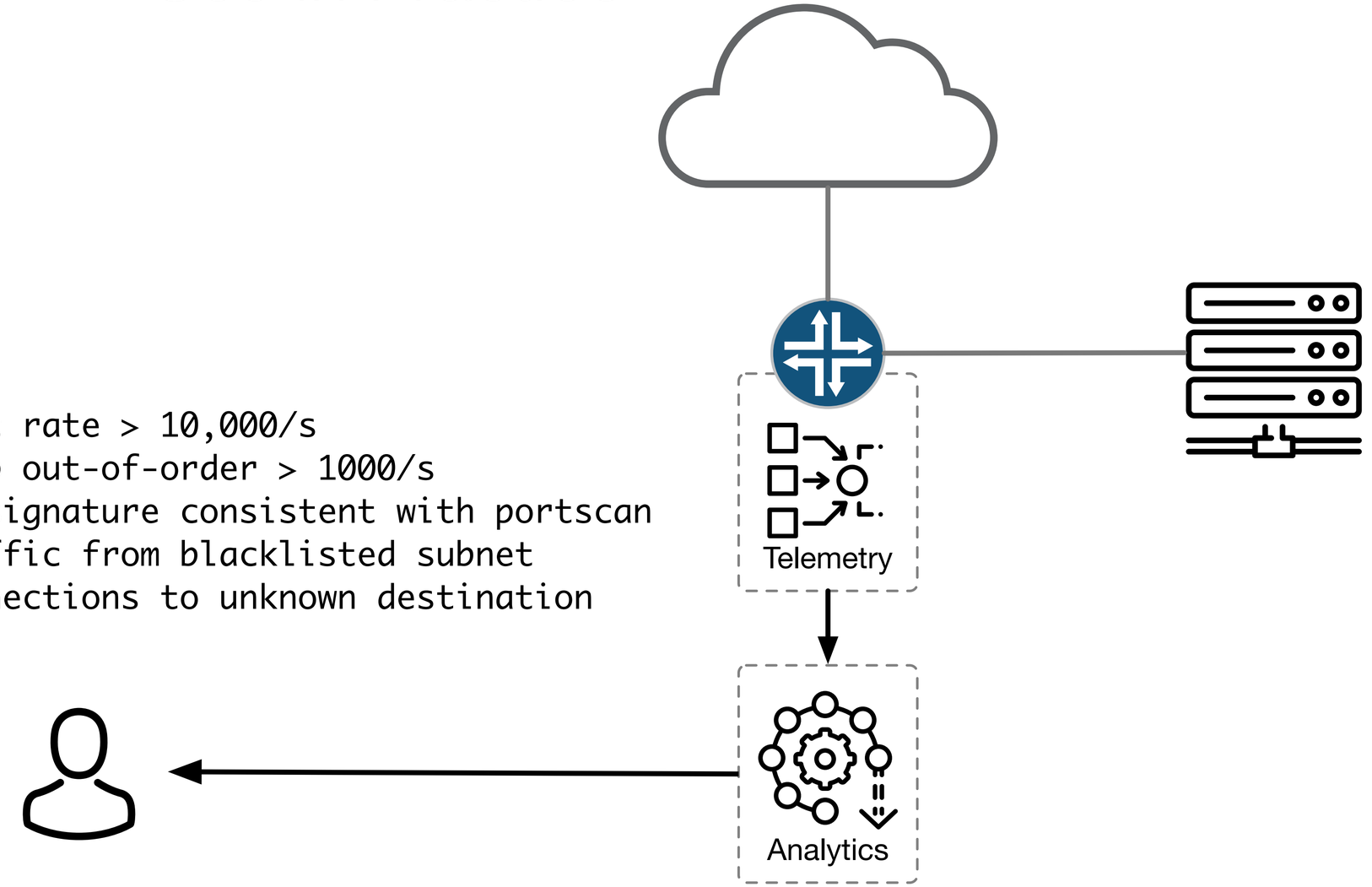
Use in Practice

```
[ip.src == 10.12.13.14] pkt rate > 10,000/s
[ip.dst == 10.12.13.14] tcp out-of-order > 1000/s
[ip.src == 239.232.230.2] signature consistent with portscan
[ip.src == 123.123/24] traffic from blacklisted subnet
```



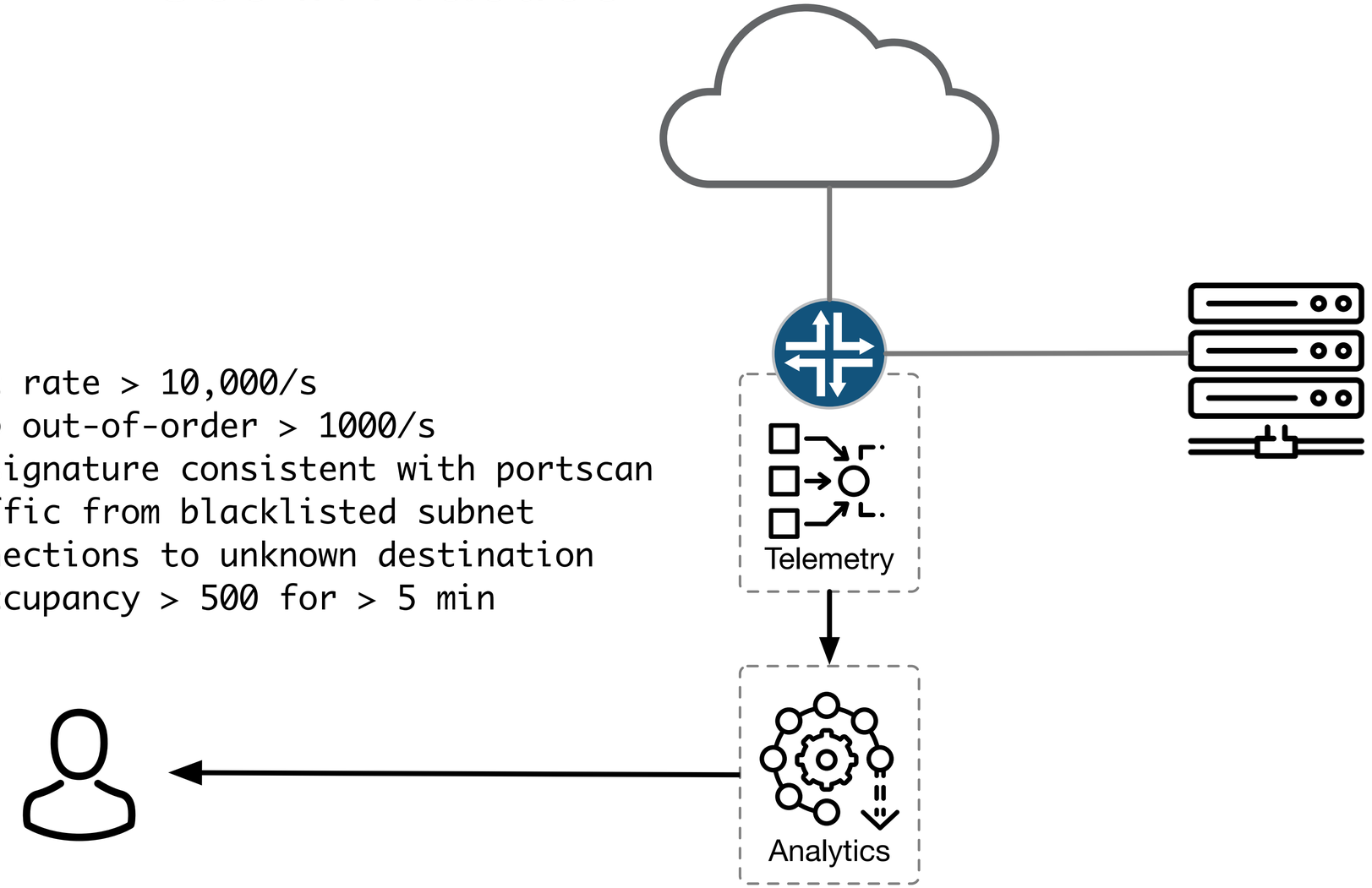
Use in Practice

```
[ip.src == 10.12.13.14] pkt rate > 10,000/s
[ip.dst == 10.12.13.14] tcp out-of-order > 1000/s
[ip.src == 239.232.230.2] signature consistent with portscan
[ip.src == 123.123/24] traffic from blacklisted subnet
[ip.dst == 234.2.3.99] connections to unknown destination
```



Use in Practice

```
[ip.src == 10.12.13.14] pkt rate > 10,000/s
[ip.dst == 10.12.13.14] tcp out-of-order > 1000/s
[ip.src == 239.232.230.2] signature consistent with portscan
[ip.src == 123.123/24] traffic from blacklisted subnet
[ip.dst == 234.2.3.99] connections to unknown destination
[switch1 queue 23] queue occupancy > 500 for > 5 min
```



Use in Practice

```
[switch1 queue 23] queue occupancy > 500 for > 5 min
```

Use in Practice

```
[switch1 queue 23] queue occupancy > 500 for > 5 min
 ↳ avg. occupancy 627
```

Use in Practice

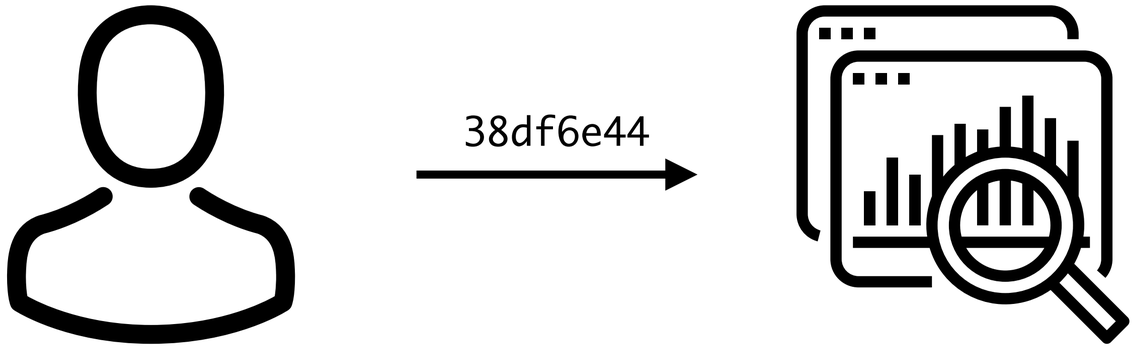
```
[switch1 queue 23] queue occupancy > 500 for > 5 min
 ↳ avg. occupancy 627
 ↳ 2018-11-28 15:23:12.023112
```

Use in Practice

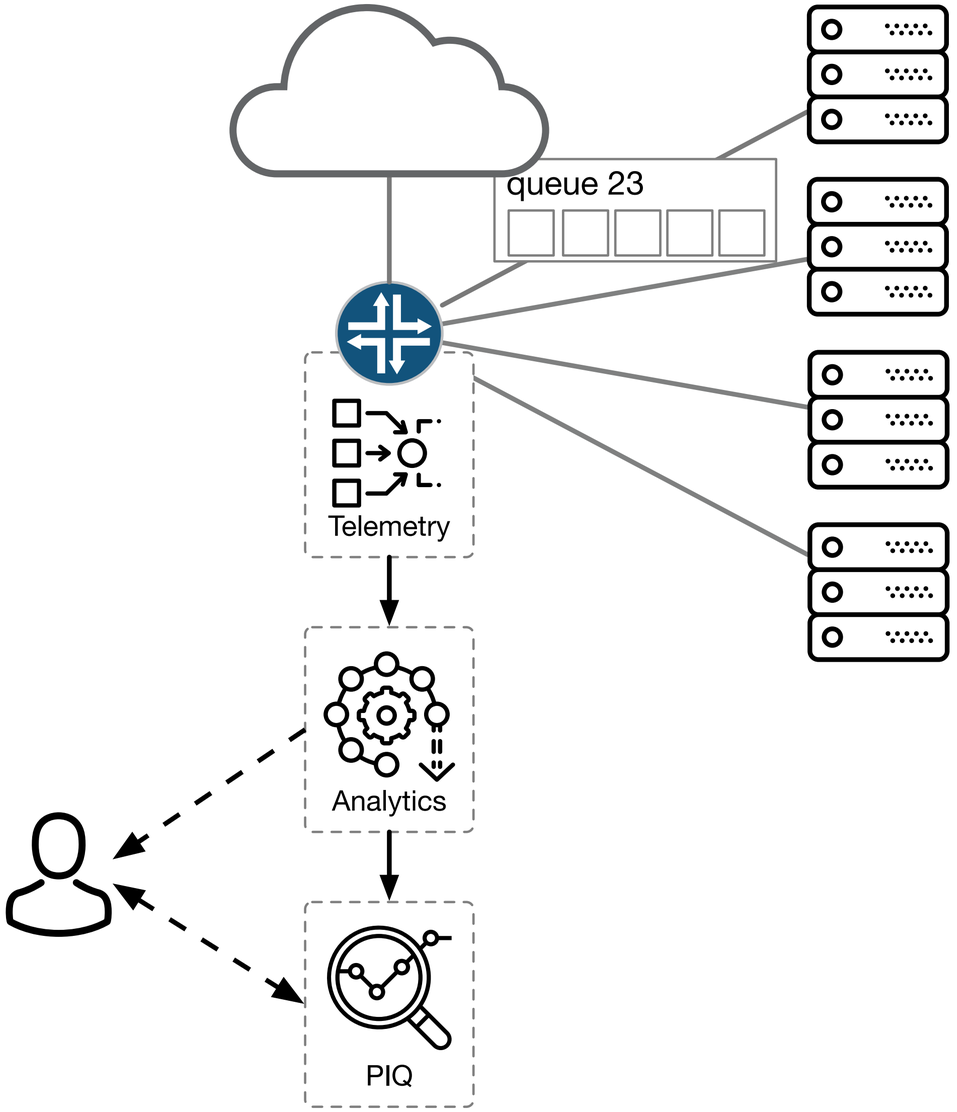
```
[switch1 queue 23] queue occupancy > 500 for > 5 min
 ↳ avg. occupancy 627
 ↳ 2018-11-28 15:23:12.023112
 ↳ incident id: 38df6e44
```

Use in Practice

```
[switch1 queue 23] queue occupancy > 500 for > 5 min
 ↳ avg. occupancy 627
 ↳ 2018-11-28 15:23:12.023112
 ↳ incident id: 38df6e44
```

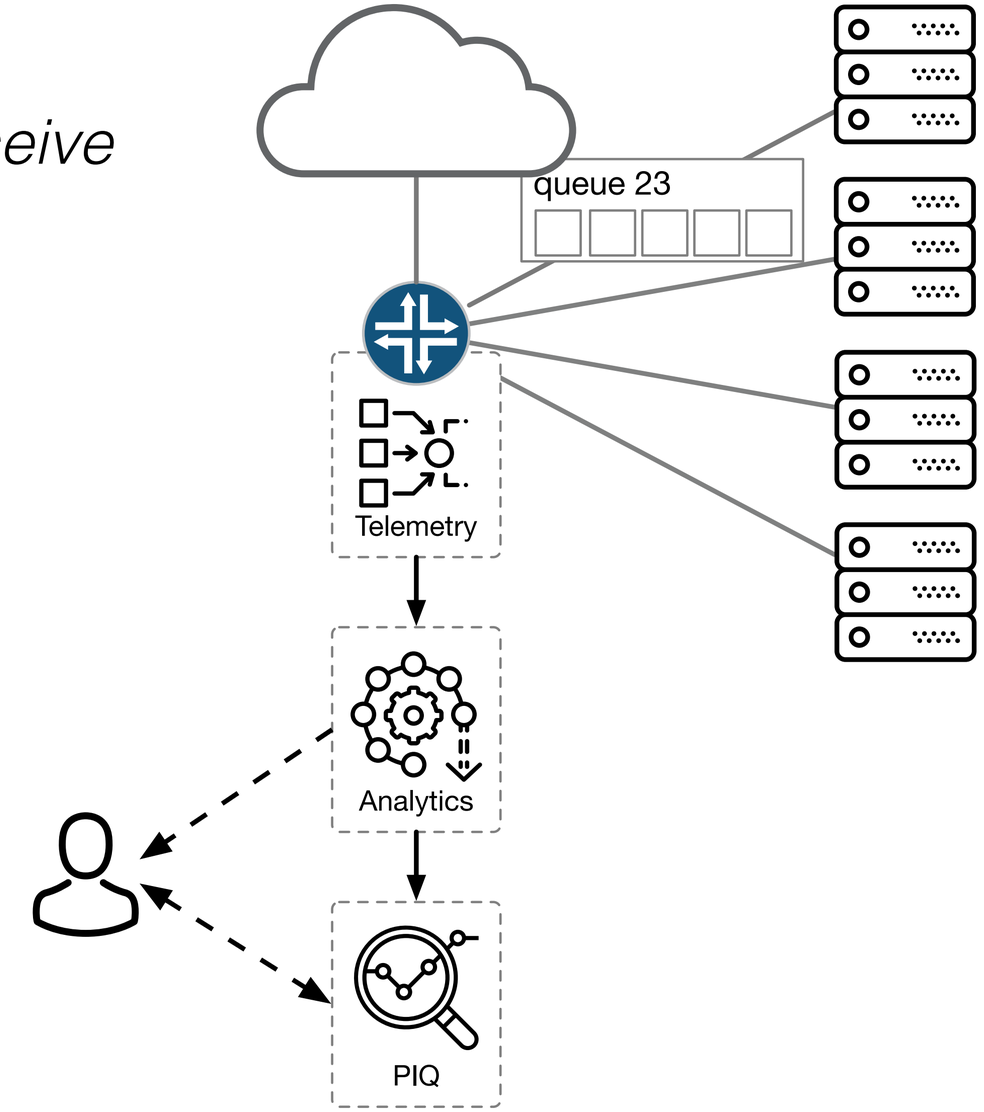


Retrospective Network Queries



Retrospective Network Queries

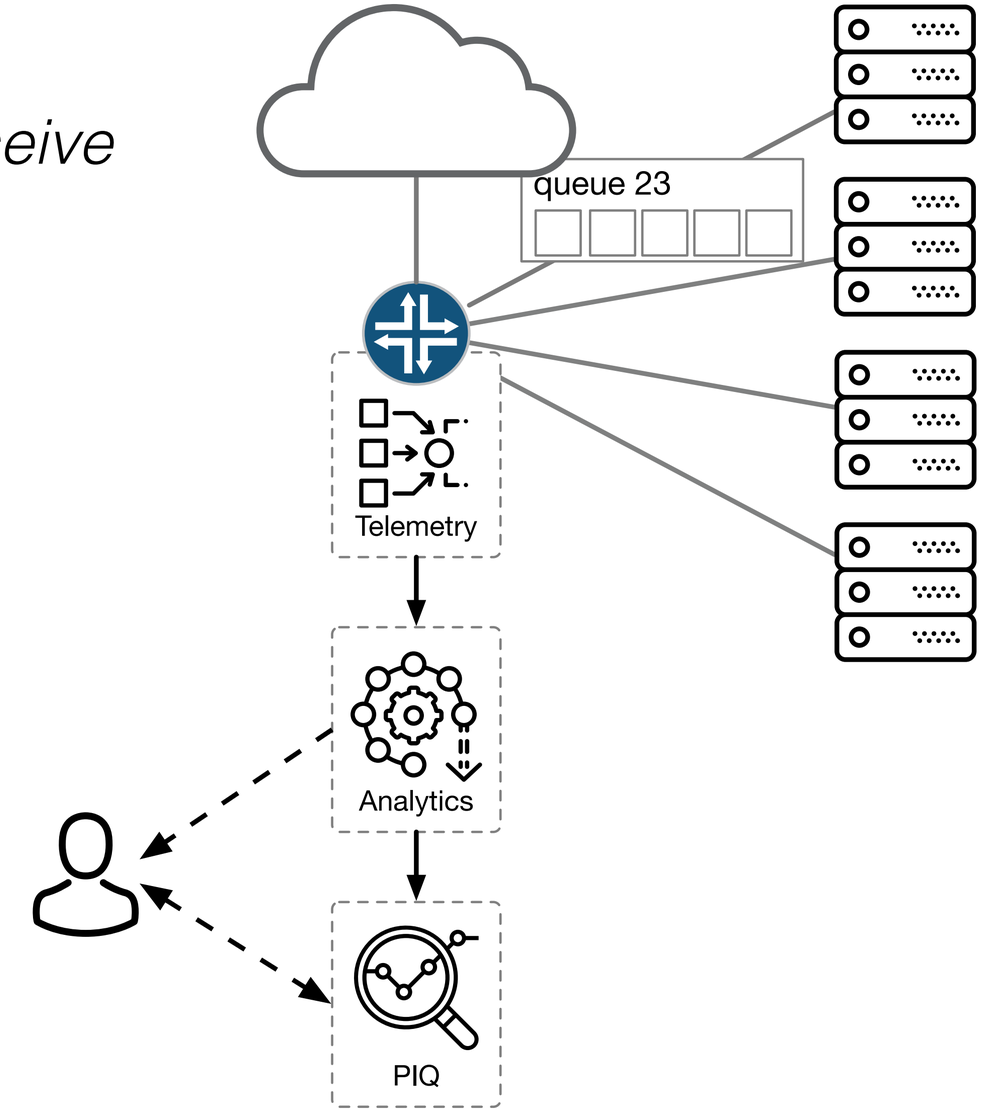
Is there a configuration problem or do we ‘just’ receive more than usual traffic?



Retrospective Network Queries

Is there a configuration problem or do we ‘just’ receive more than usual traffic?

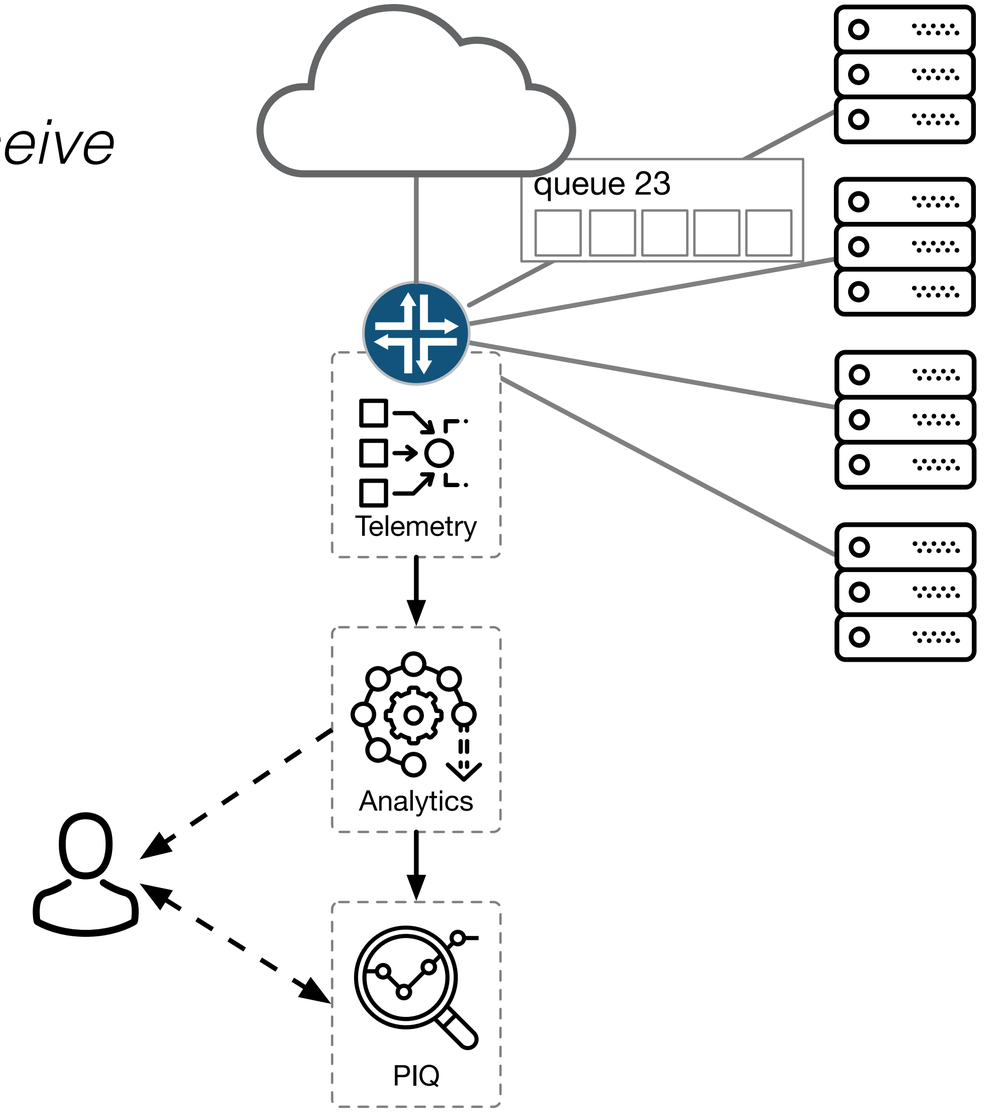
- “show me sample 38df6e44”
  - traffic looks normal



Retrospective Network Queries

Is there a configuration problem or do we ‘just’ receive more than usual traffic?

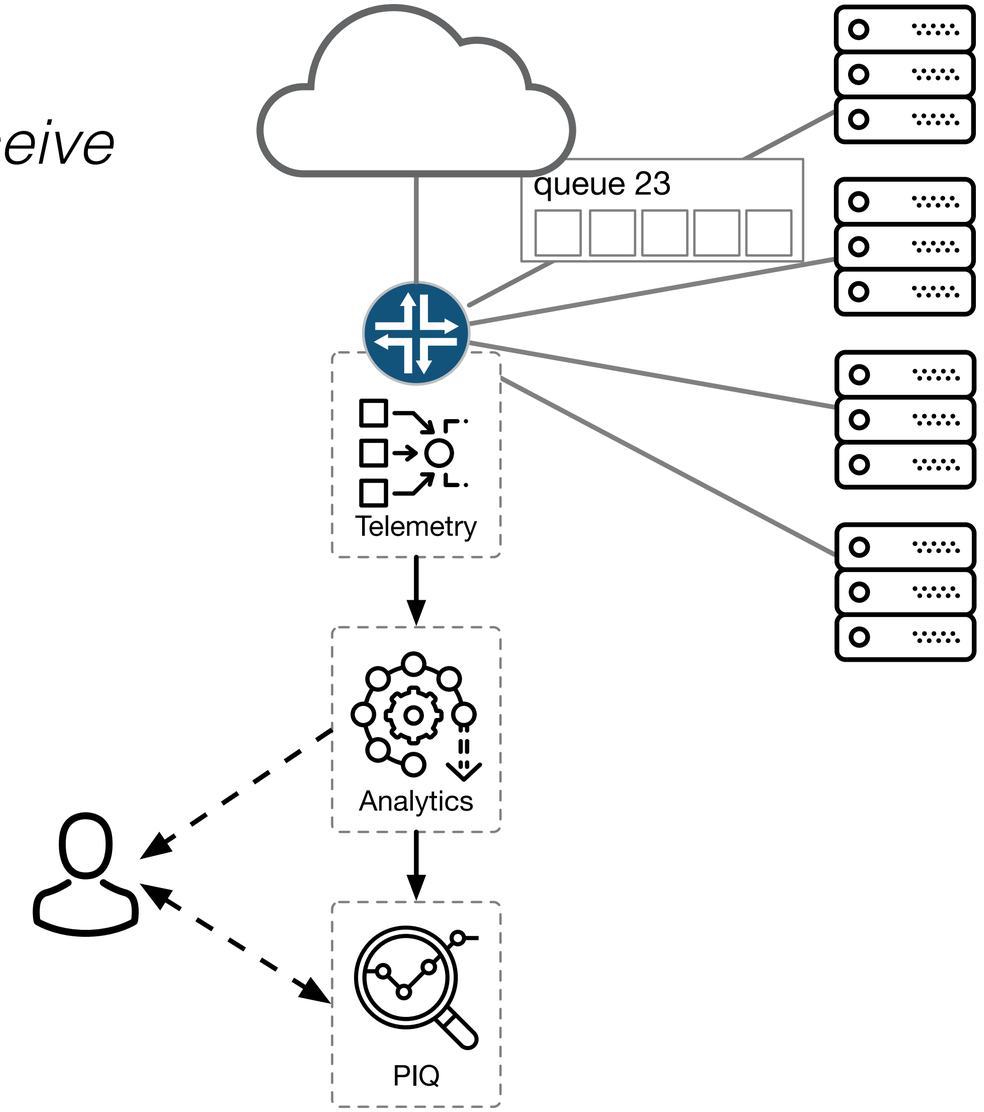
- “show me sample 38df6e44”
  - traffic looks normal
- “show me information about queue 23”
  - queue is part of load balancing group



Retrospective Network Queries

Is there a configuration problem or do we 'just' receive more than usual traffic?

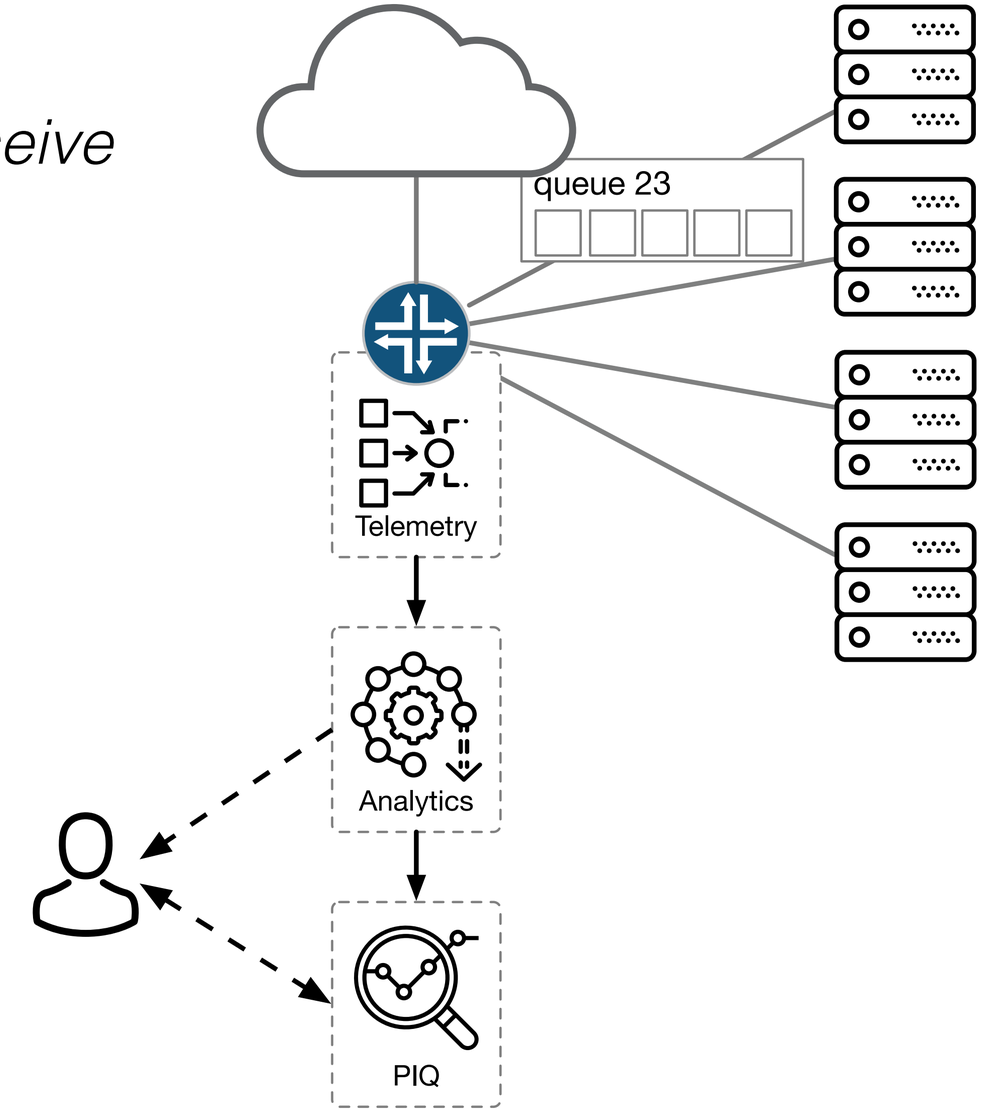
- "show me sample 38df6e44"
  - traffic looks normal
- "show me information about queue 23"
  - queue is part of load balancing group
- "what other queues are in that load balancing group?"
  - queues 24, 25, and 26



Retrospective Network Queries

Is there a configuration problem or do we ‘just’ receive more than usual traffic?

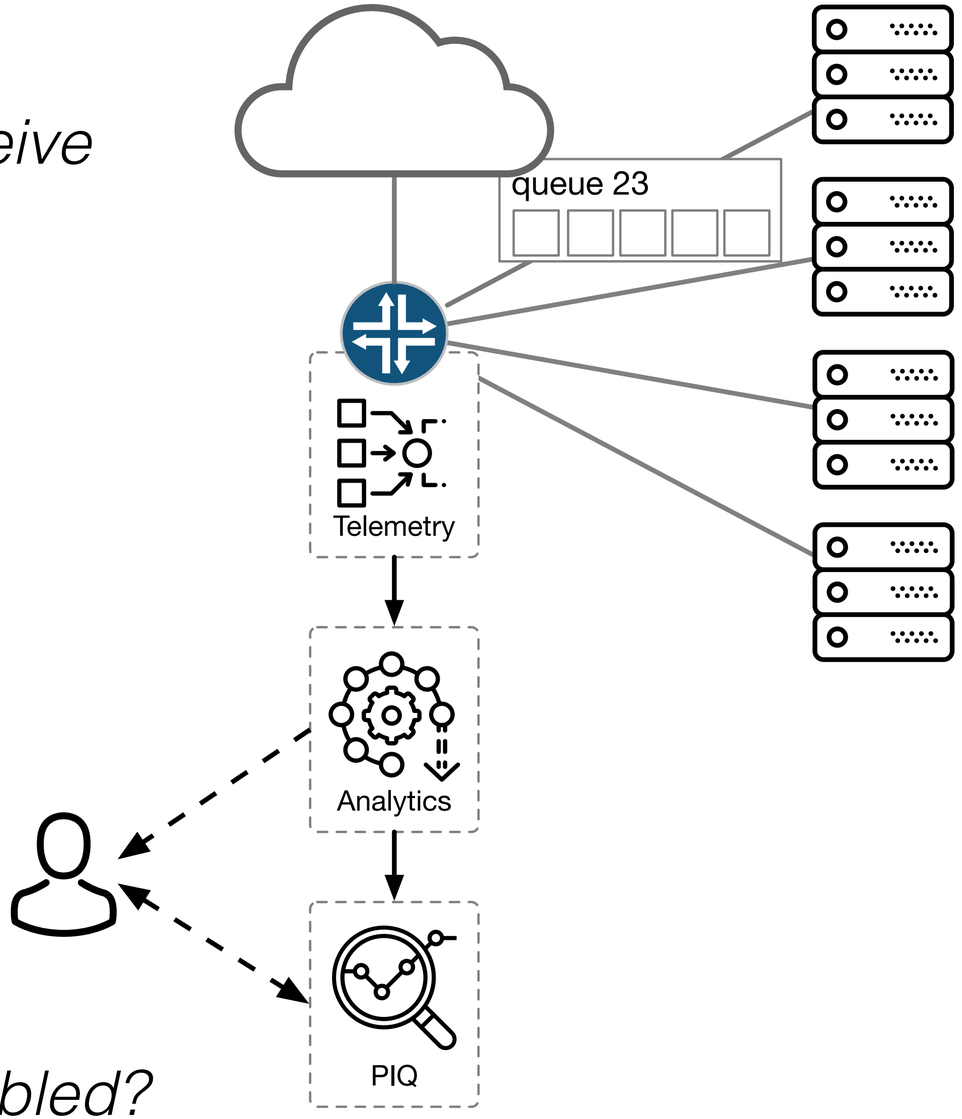
- “show me sample 38df6e44”
  - traffic looks normal
- “show me information about queue 23”
  - queue is part of load balancing group
- “what other queues are in that load balancing group?”
  - queues 24, 25, and 26
- “show me traffic in queues 24, 25, 26 at 2018-11-28 15:23:12.023112”
  - only control traffic



Retrospective Network Queries

Is there a configuration problem or do we ‘just’ receive more than usual traffic?

- “show me sample 38df6e44”
  - traffic looks normal
- “show me information about queue 23”
  - queue is part of load balancing group
- “what other queues are in that load balancing group?”
  - queues 24, 25, and 26
- “show me traffic in queues 24, 25, 26 at 2018-11-28 15:23:12.023112”
  - only control traffic



Problem with Hashing? Other queues in group disabled?

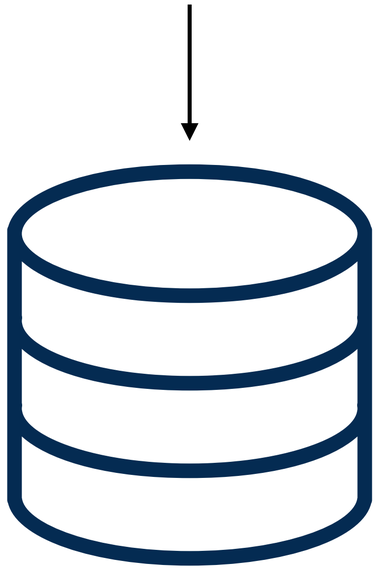
Retrospective Network Queries

Persistent

Interactive Queries

Retrospective Network Queries

Persistent

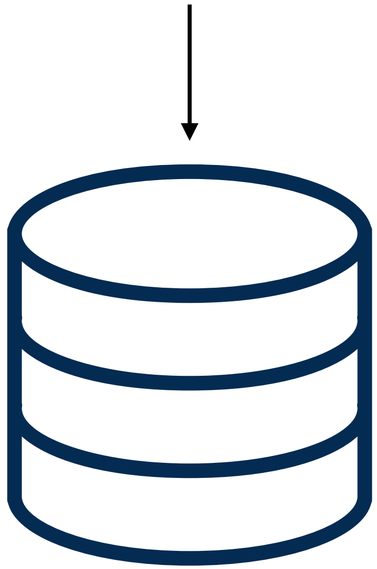


Database

Interactive Queries

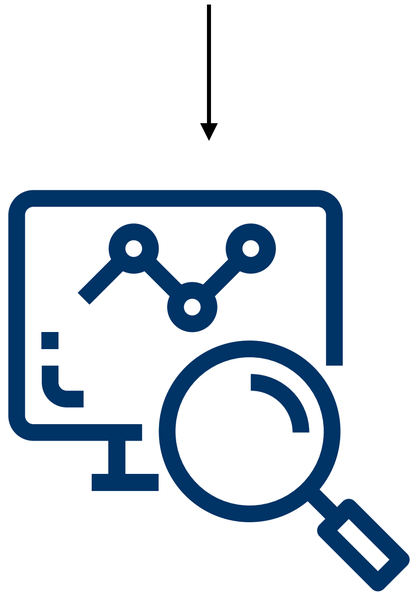
Retrospective Network Queries

Persistent



Database

Interactive Queries



Query Interface

Retrospective Network Queries

1. Storing Network Records
2. Injecting Records into a Database
3. Querying the Database

Grouped Packet Vectors

[Sonchack et.al. *Flow, USENIX ATC 2019]

- per-packet header fields
- meta data: *e.g.*, queue depth, ingress/egress timestamps

aggregation keys

packet features

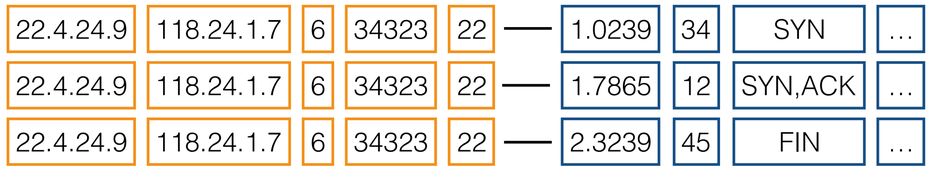
aggregated features

Grouped Packet Vectors

[Sonchack et.al. *Flow, USENIX ATC 2019]

- per-packet header fields
- meta data: *e.g.*, queue depth, ingress/egress timestamps

Packet Records



aggregation keys

packet features

aggregated features

Grouped Packet Vectors

[Sonchack et.al. *Flow, USENIX ATC 2019]

- per-packet header fields
- meta data: *e.g.*, queue depth, ingress/egress timestamps

Packet Records

22.4.24.9	118.24.1.7	6	34323	22	1.0239	34	SYN	…
22.4.24.9	118.24.1.7	6	34323	22	1.7865	12	SYN,ACK	…
22.4.24.9	118.24.1.7	6	34323	22	2.3239	45	FIN	…

aggregation keys

packet features

aggregated features

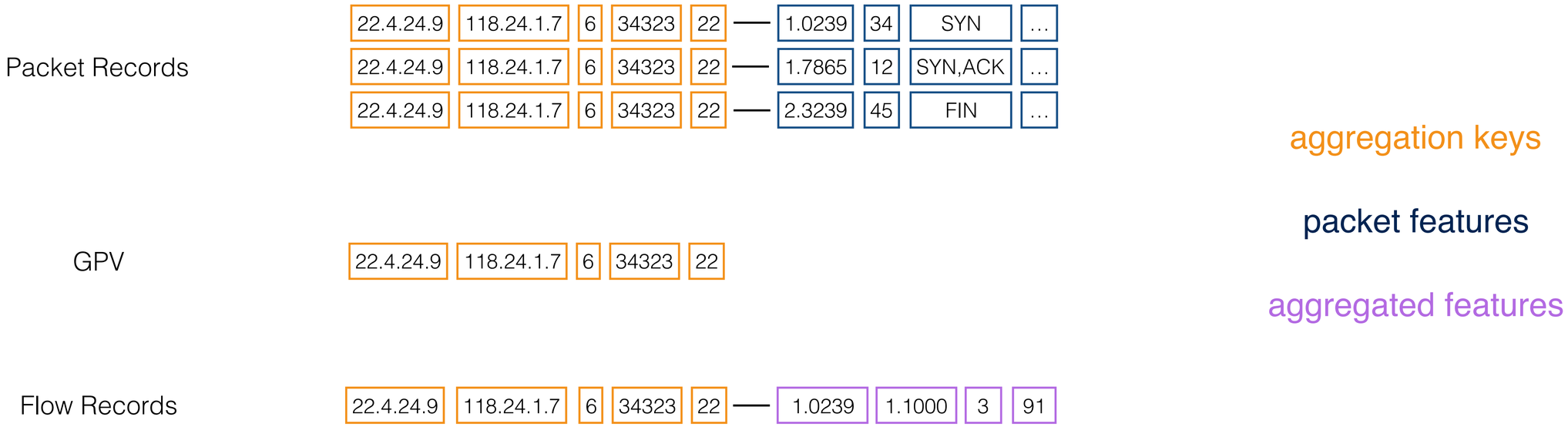
Flow Records

22.4.24.9	118.24.1.7	6	34323	22	1.0239	1.1000	3	91

Grouped Packet Vectors

[Sonchack et.al. *Flow, USENIX ATC 2019]

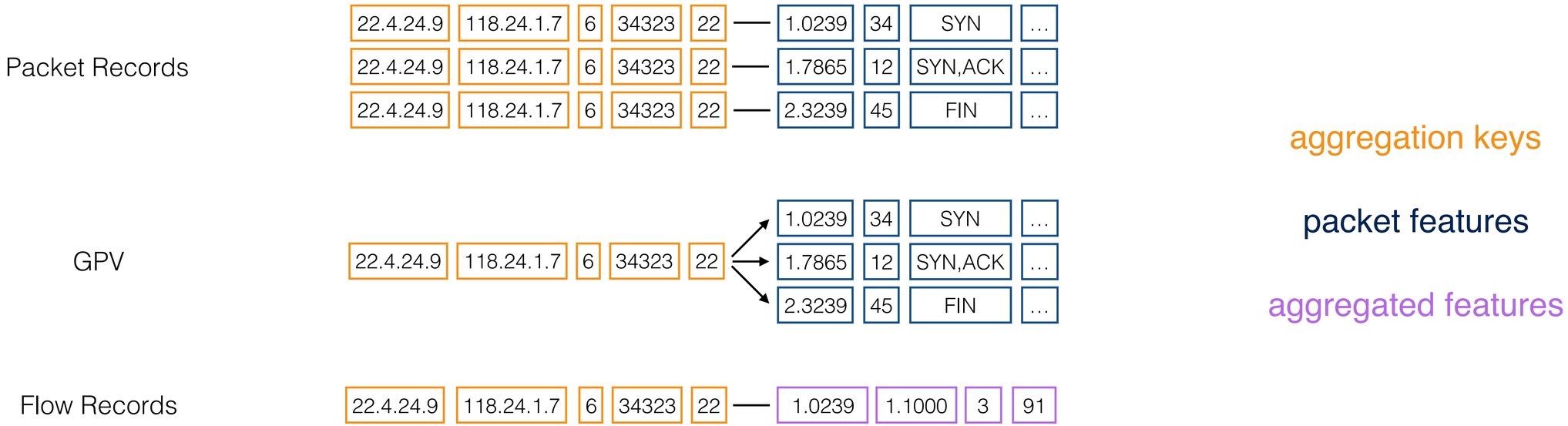
- per-packet header fields
- meta data: *e.g.*, queue depth, ingress/egress timestamps



Grouped Packet Vectors

[Sonchack et.al. *Flow, USENIX ATC 2019]

- per-packet header fields
- meta data: *e.g.*, queue depth, ingress/egress timestamps

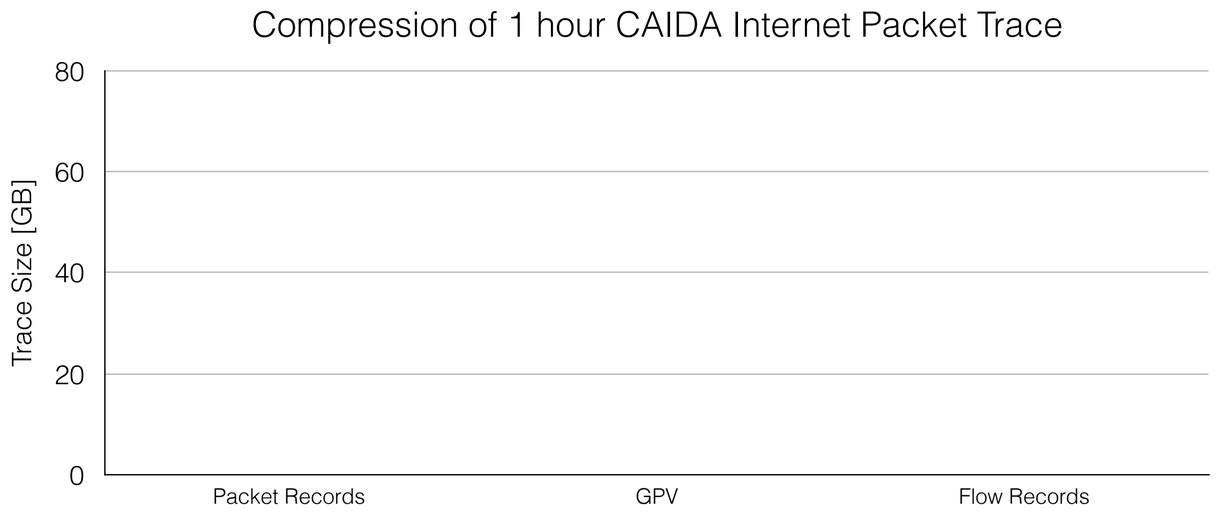


Grouped Packet Vectors

- GPVs provide high compression while maintaining information richness

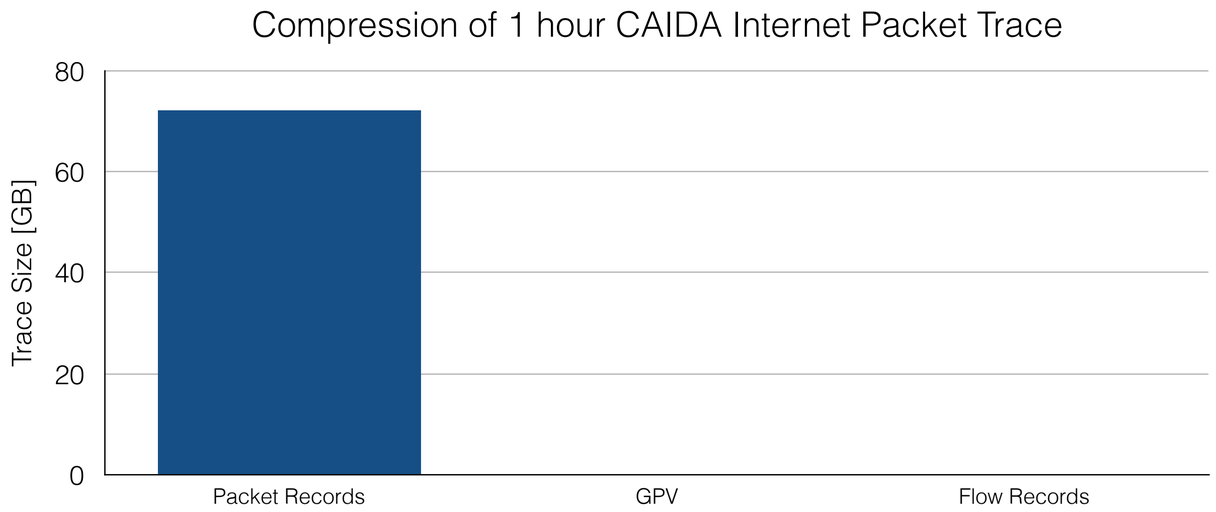
Grouped Packet Vectors

- GPVs provide high compression while maintaining information richness



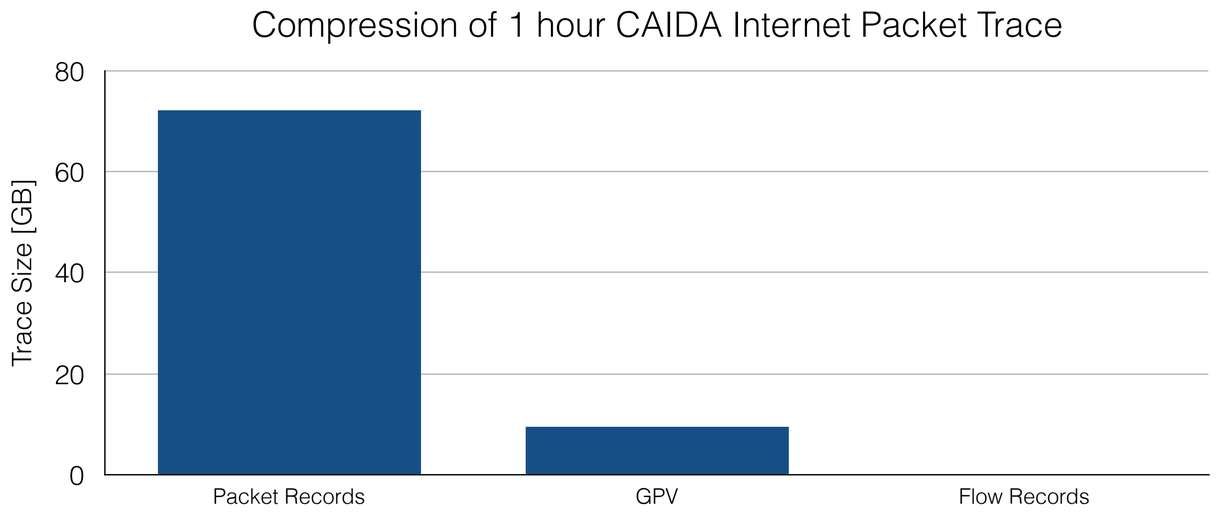
Grouped Packet Vectors

- GPVs provide high compression while maintaining information richness



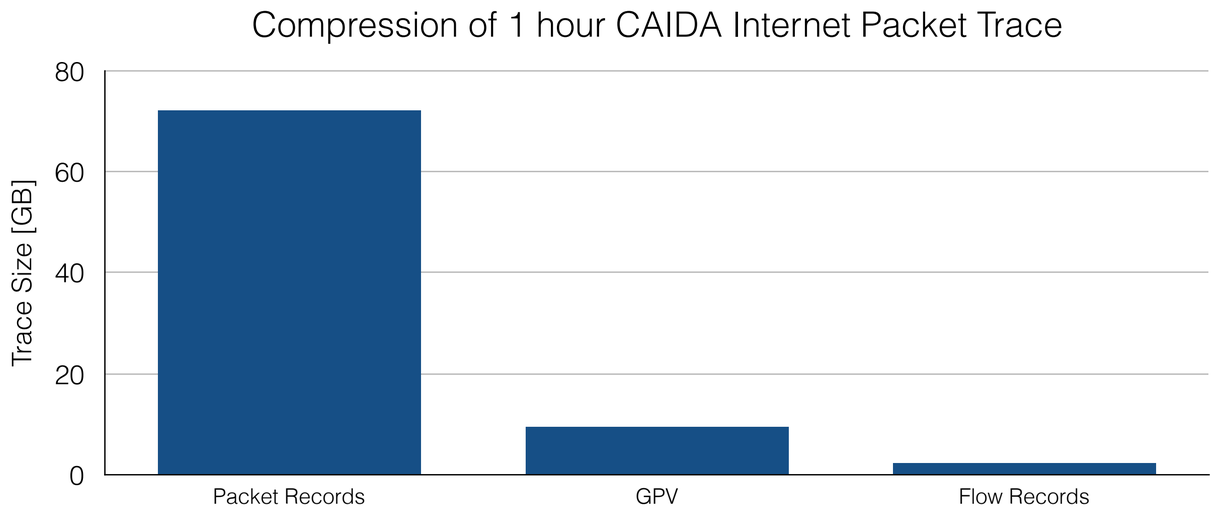
Grouped Packet Vectors

- GPVs provide high compression while maintaining information richness



Grouped Packet Vectors

- GPVs provide high compression while maintaining information richness



Database Models

Database Models

Relational

Database Models

Relational

NoSQL

Database Models

Relational

NoSQL

Time-Series

Database Models

Relational

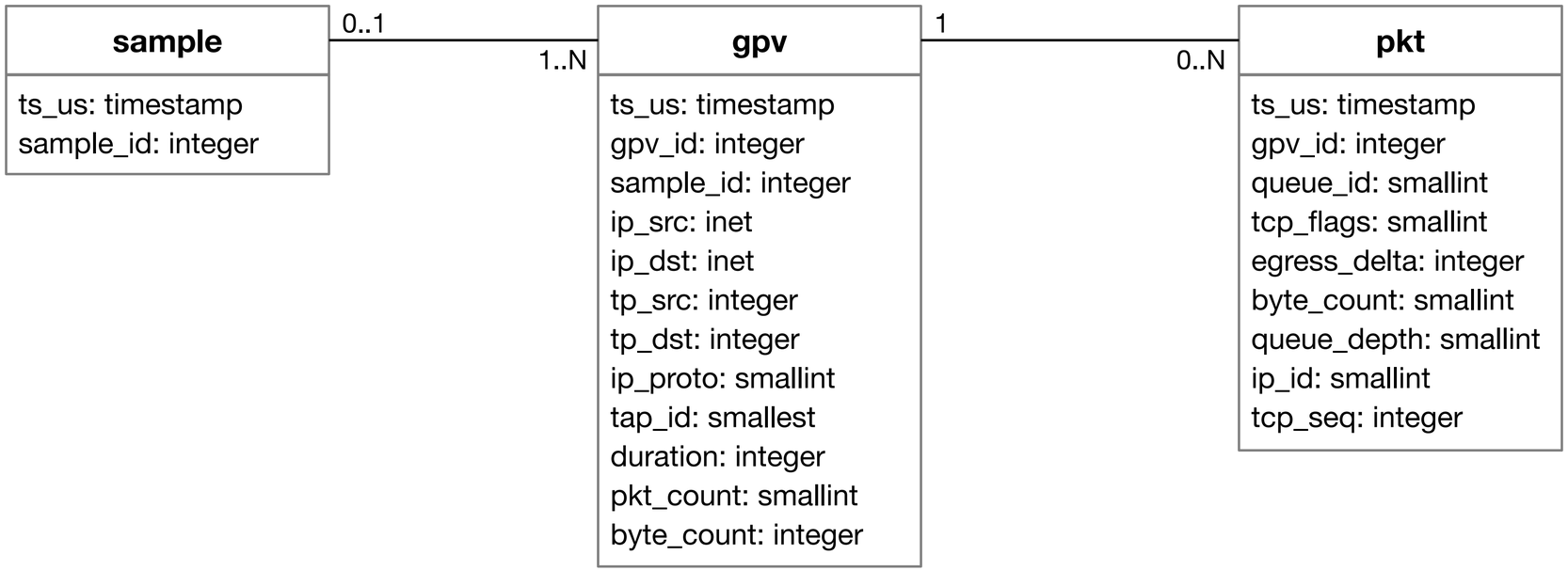
NoSQL

Time-Series

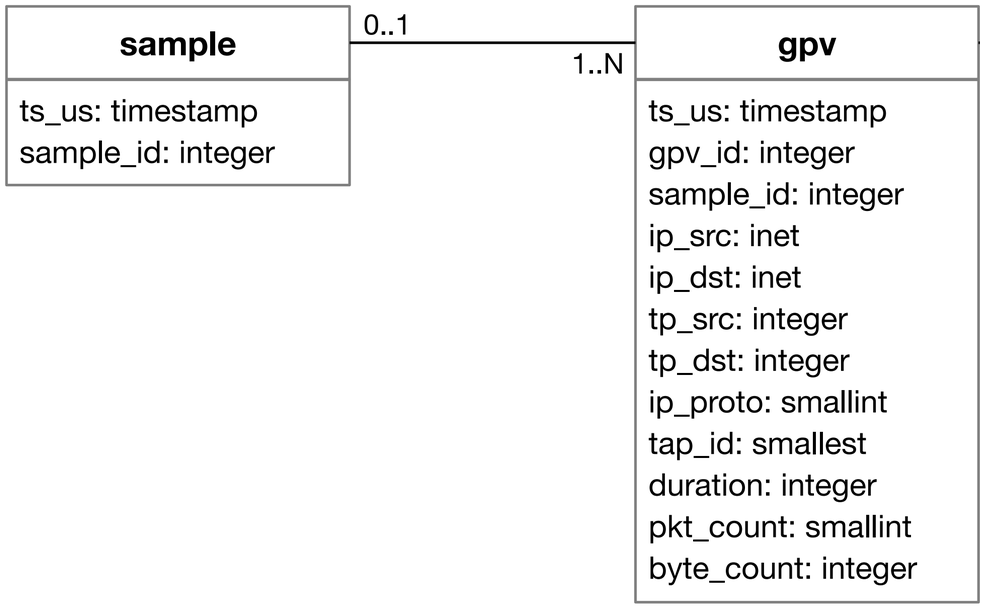
- optimized for time-series data
  - append-heavy workloads
  - grouping functionality based on time
  - timestamps as primary keys for data
  - relational or NoSQL

Storing Network Records

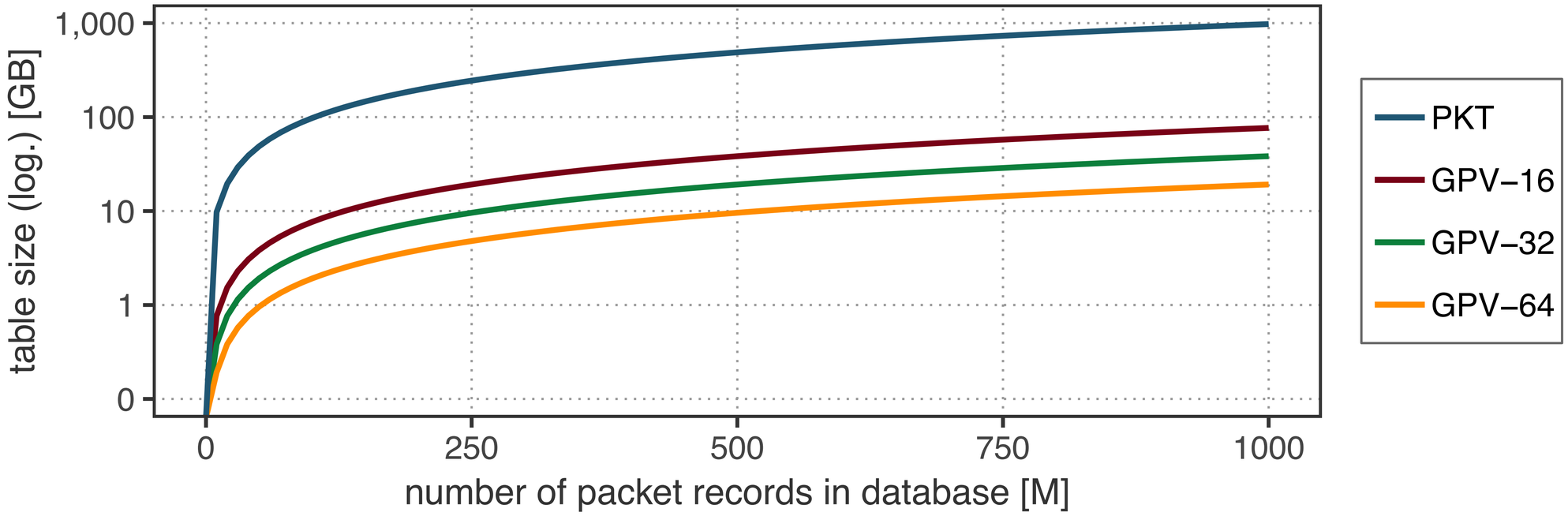
Storing Network Records



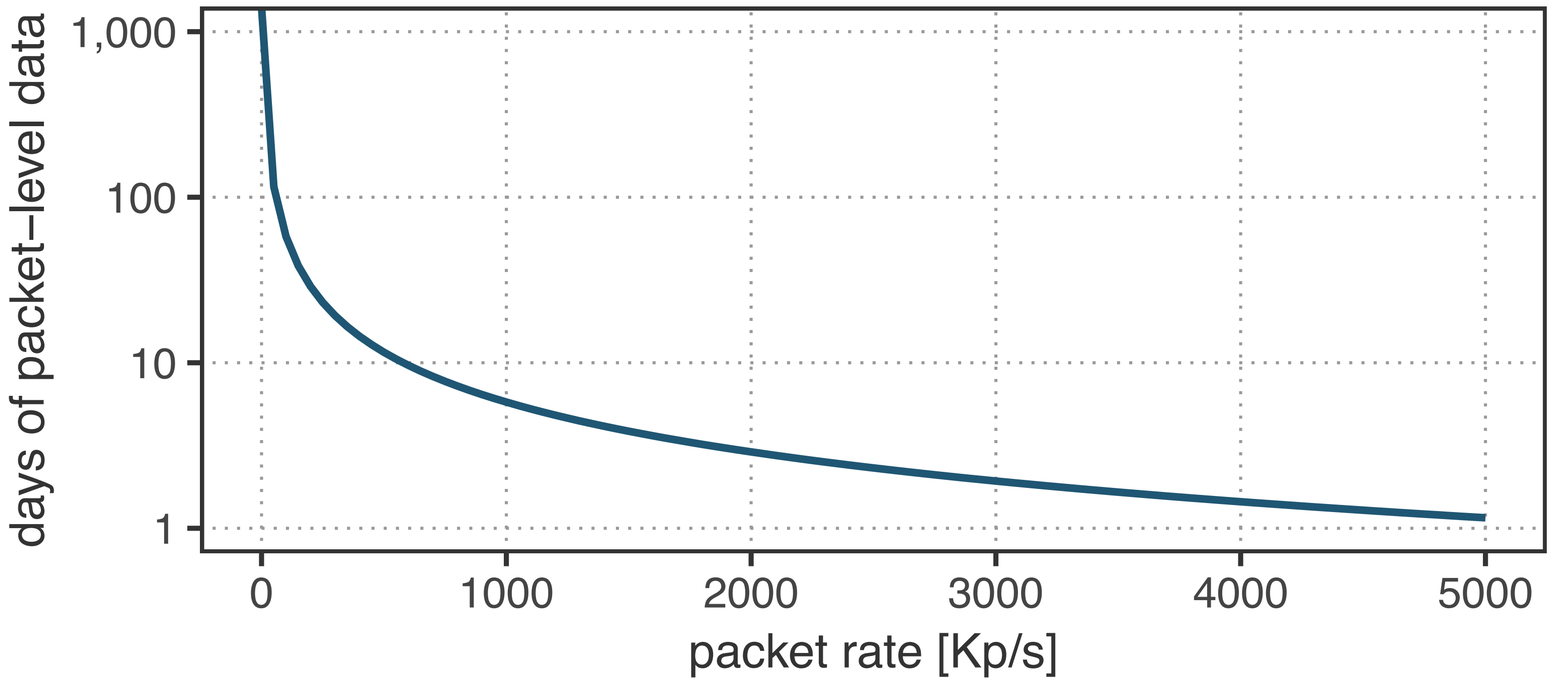
Storing Network Records



Storing Network Records



Storing Network Records



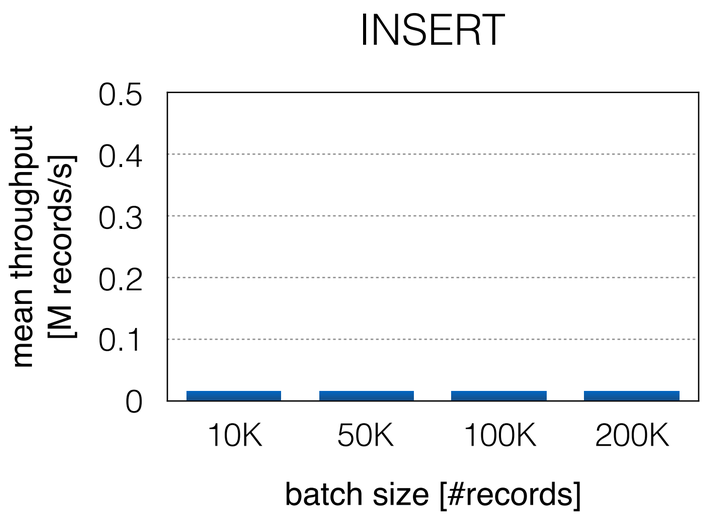
Injecting Network Records

Injecting Network Records

```
INSERT INTO packets VALUES
 (‘2018-11-25 19:02:23.023219’,
 ‘123.213.123.4’, ‘213.11.23.23’, …)
```

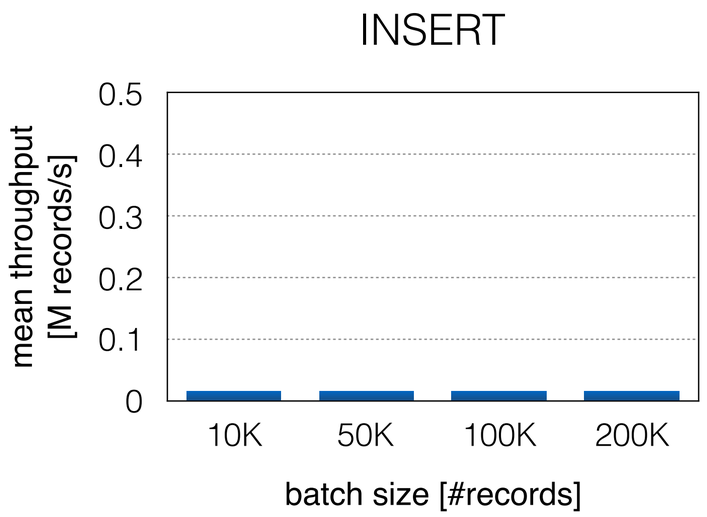
Injecting Network Records

```
INSERT INTO packets VALUES
 (‘2018-11-25 19:02:23.023219’,
 ‘123.213.123.4’, ‘213.11.23.23’, …)
```



Injecting Network Records

```
INSERT INTO packets VALUES
 (‘2018-11-25 19:02:23.023219’,
 ‘123.213.123.4’, ‘213.11.23.23’, …)
```

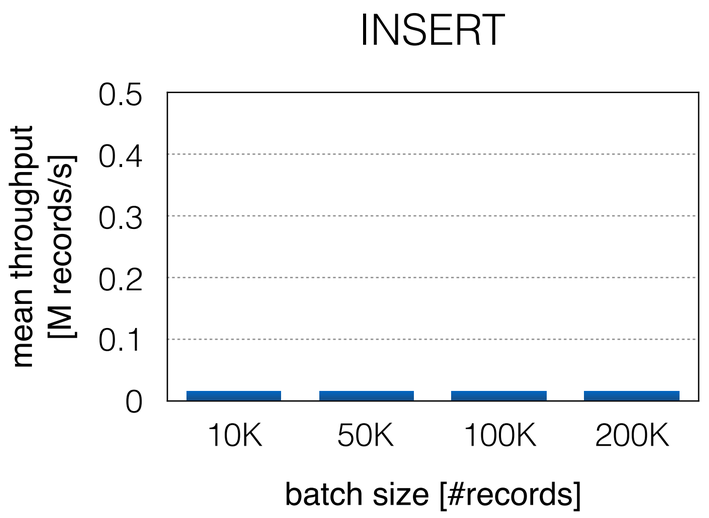


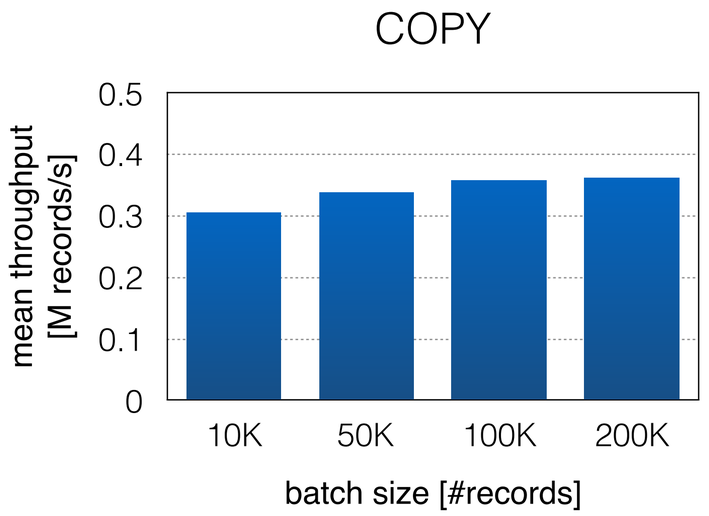

```
1  auto buf = new char[h_len + 10000
2     * rec_len + t_len];
3
4  // 10000 times
5  pkt.write_to_pg_buf(buf, offset);
6
7  pg_conn.copy(“pkt”, buf, sizeof(buf));
```

Injecting Network Records

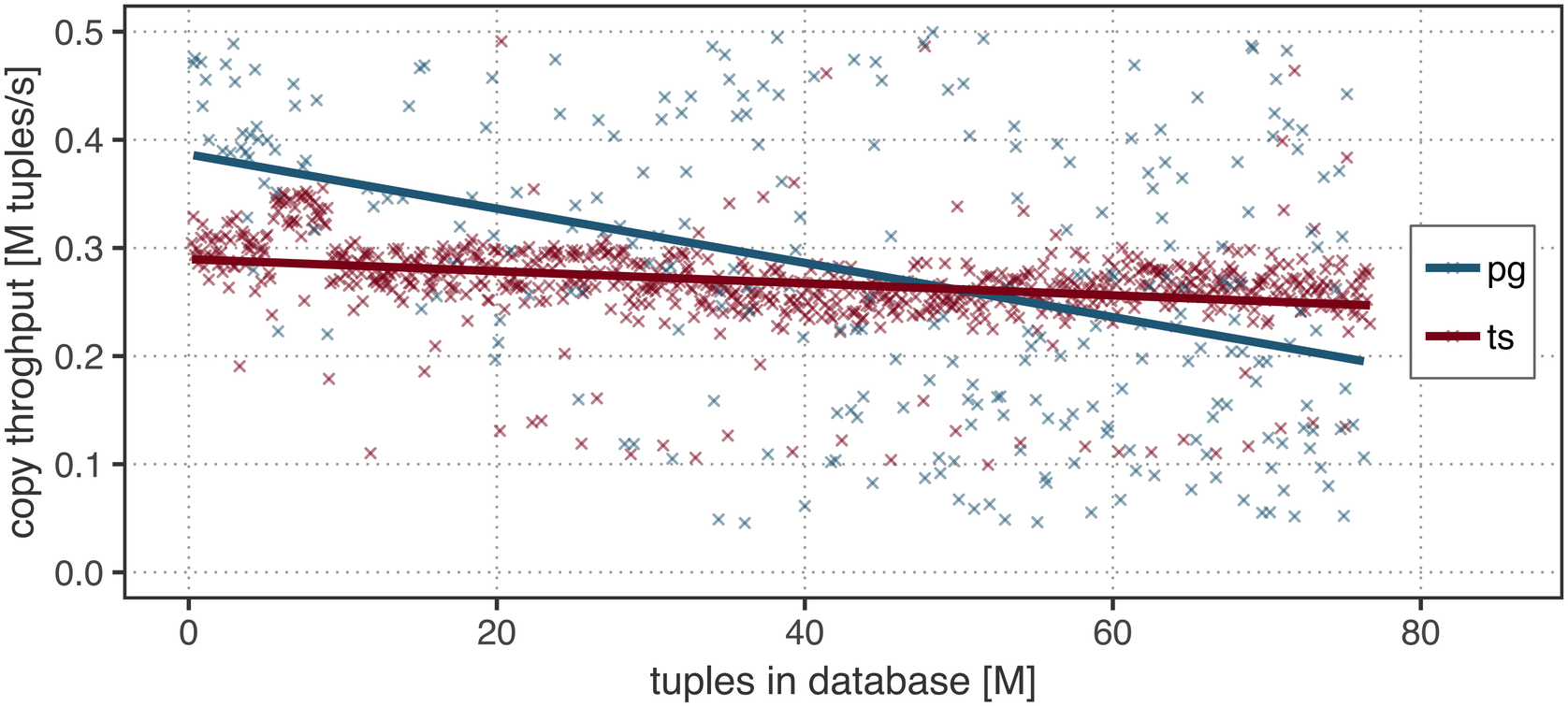
```
INSERT INTO packets VALUES
 (‘2018-11-25 19:02:23.023219’,
 ‘123.213.123.4’, ‘213.11.23.23’, …)
```

```
1  auto buf = new char[h_len + 10000
2     * rec_len + t_len];
3
4  // 10000 times
5  pkt.write_to_pg_buf(buf, offset);
6
7  pg_conn.copy(“pkt”, buf, sizeof(buf));
```





Injecting Network Records



Querying Network Records

Querying Network Records

```
SELECT time_bucket('1 seconds', ts_us) AS interval, SUM(pkt_count), SUM(byte_count)
  FROM gpv GROUP BY interval ORDER BY interval ASC
```

Querying Network Records

```
SELECT time_bucket('1 seconds', ts_us) AS interval, SUM(pkt_count), SUM(byte_count)
  FROM gpv GROUP BY interval ORDER BY interval ASC
```

```
SELECT * FROM (gpv RIGHT JOIN pkt ON gpv.gpv_id = pkt.gpv_id)
  WHERE gpv.ip_src << inet '60.70.0.0/16'
```

Querying Network Records

Querying Network Records

```
SELECT DISTINCT gpv.ip_src, gpv.ip_dst, gpv.ip_proto, gpv.tp_src, gpv.tp_dst
  FROM (gpv RIGHT JOIN pkt ON gpv.gpv_id = pkt.gpv_id)
  WHERE pkt.ingress_ts >= '2018-02-19 12:59:11.595’
    AND pkt.ingress_ts < '2018-02-19 12:59:11.600'
    AND pkt.queue_id = ’23’
```

Querying Network Records

```
SELECT DISTINCT gpv.ip_src, gpv.ip_dst, gpv.ip_proto, gpv.tp_src, gpv.tp_dst
  FROM (gpv RIGHT JOIN pkt ON gpv.gpv_id = pkt.gpv_id)
  WHERE pkt.ingress_ts >= '2018-02-19 12:59:11.595’
    AND pkt.ingress_ts < '2018-02-19 12:59:11.600'
    AND pkt.queue_id = ’23’
```

```
SELECT pkt.queue_id, COUNT(pkt.queue_id)
  FROM (gpv RIGHT JOIN pkt ON gpv.gpv_id = pkt.gpv_id)
  WHERE gpv.ip_dst << inet ’53.231/16'
    AND pkt.queue_id IN (23,24,25,26)
  GROUP BY pkt.queue_id;
```

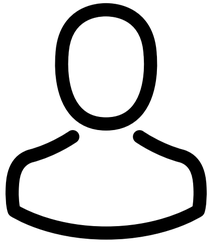
Conclusion

Conclusion

identify the need for retrospective network analytics

Conclusion

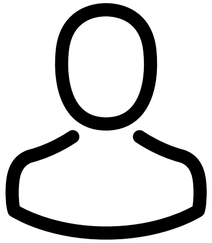
identify the need for retrospective network analytics



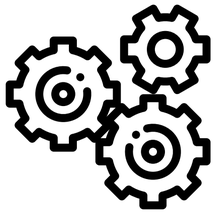
show using example queries how retrospective analytics can be used to get better insight into what happened in the network

Conclusion

identify the need for retrospective network analytics



show using example queries how retrospective analytics can be used to get better insight into what happened in the network



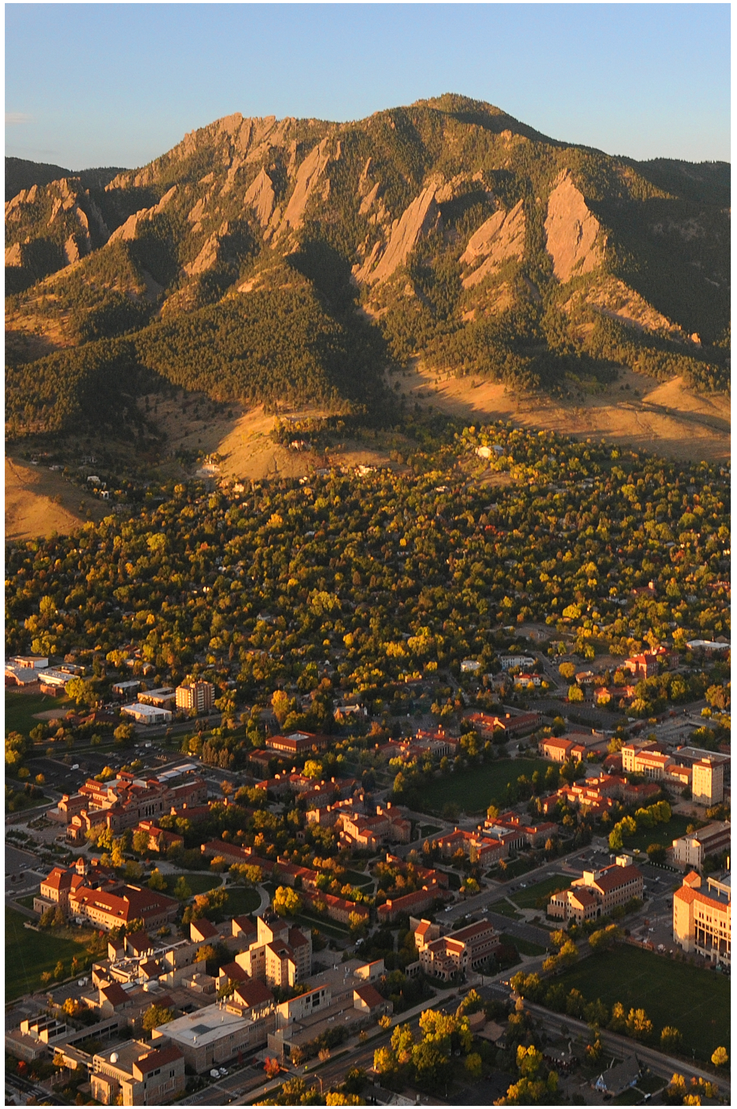
implement a prototype application demonstrating technical feasibility

Q&A / Discussion

Oliver Michel

oliver.michel@colorado.edu
http://nsr.colorado.edu/oliver

University of Colorado **Boulder**



BACKUP SLIDES

Example SQL Queries

Example SQL Queries

```
SELECT ip_src, SUM(pkt_count)
  FROM gpv WHERE ip_src << inet ’60.70.0.0/16’
    AND gpv.ts_us > now() - interval '1 hours’
  GROUP BY ip_src
```