



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

StreamWorks: Continuous Pattern Detection on Streaming Data

SUTANAY CHOUDHURY, KHUSHBU AGARWAL, SHERMAN BEUS, DANIEL
DOHNALEK, KSHITEESH HEGDE

Pacific Northwest National Laboratory



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

What is StreamWorks?



The Promise of Patterns

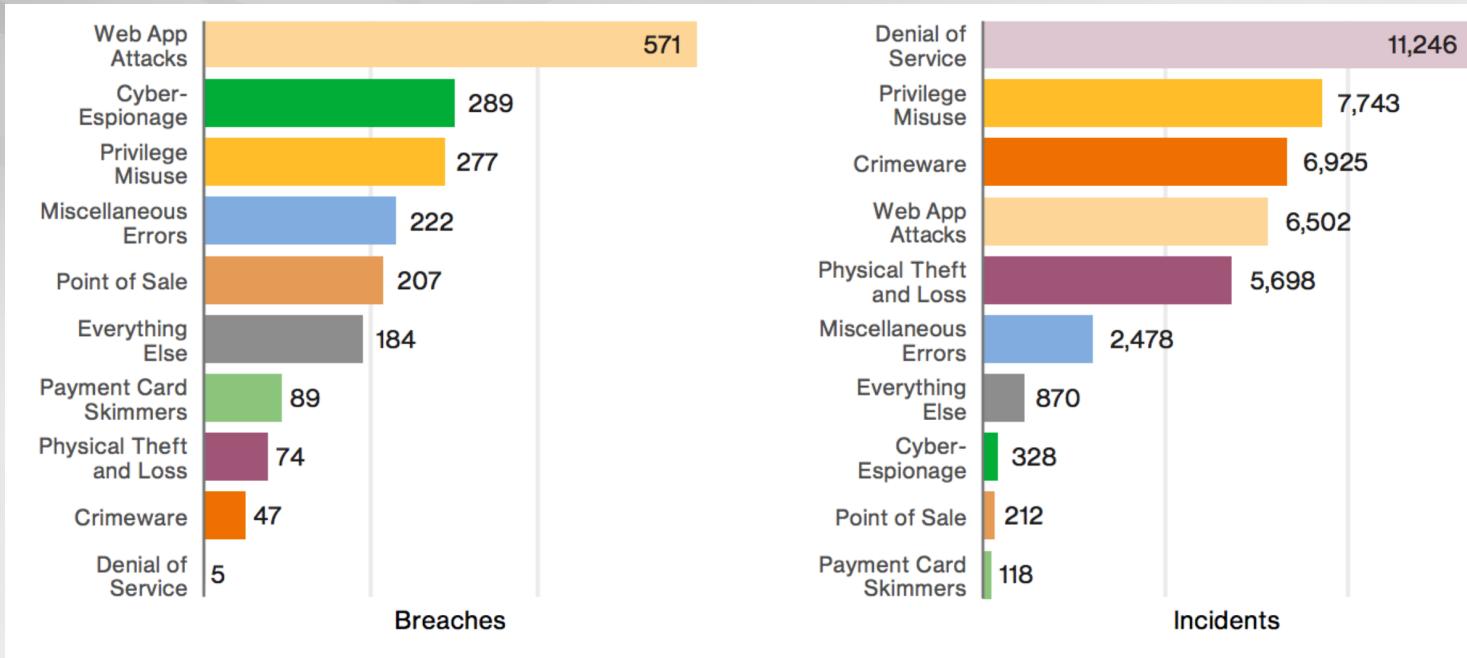


Figure 33: Percentage and count of breaches per pattern (n=1,935)

Figure 34: Percentage and count of incidents per pattern (n=42,068)

Source: Verizon 2017 Data Breach Investigations Report

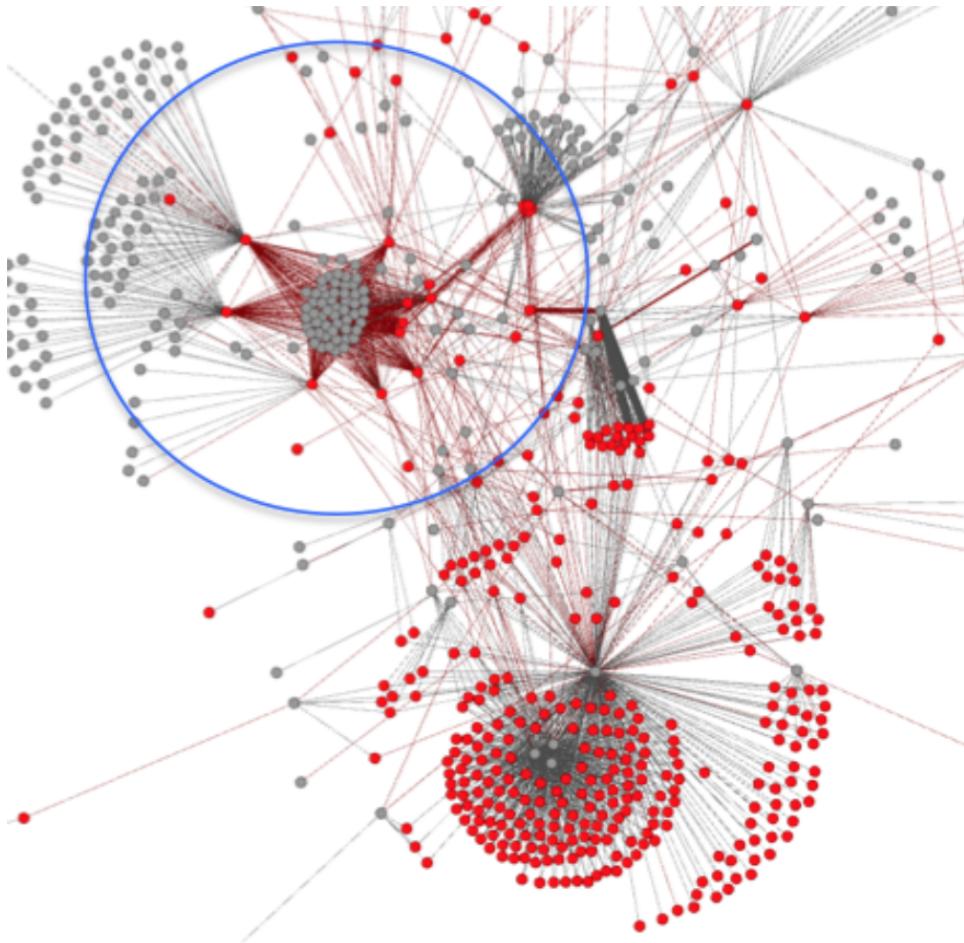
- The median number of days to detect security breaches was 146 days in 2015 – FireEye/Mandiant Report
- In its “Data Breach Investigations Report” in 2014, Verizon analyzed 100,000 security incidents from past decade and concluded 90% attacks fell in 10 attack patterns



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Graphs and Patterns



Red nodes are **Services**

Gray nodes are **Clients**

Users with similar role demonstrate same pattern of service usage



Tell me as soon as it happens!

► How do you read email?

- Read every email as soon as it comes in (**Continuous Processing**)
- Read every 4 hours (**Periodic or Batched Processing**)

► Unfortunately, being late is not better than never in all cases

- **Cyber:** Data leaving your network or a malware spread in action
- **Finance:** Price dips intraday, your late order buys high end of the day ☺

Approach for Continuous Pattern Detection

► Incremental Querying is key to Performance

- We turn streaming data into a graph model

► Guiding our insight

- We interviewed tens of analysts and system defenders, and asked about the top patterns they would like to detect

► Pattern Queries in Action

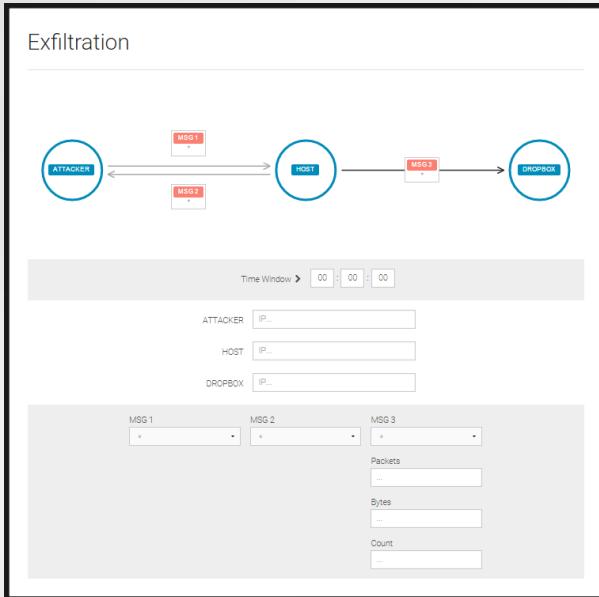
- “Tell me when a chain of 3 logins are detected with increasing privileges?”





And last, but not the least ...

- ▶ One more “Driver”
- ▶ Visual Querying: Real users should not need to learn a new query language to use the system.



```
SELECT ?control ?target ?dropbox ?xfil WHERE {  
# Control Message from C2 to target  
?control ?ctrlmsg ?target .  
?ctrlmsg :FTIME ?ftime1 .  
?ctrlmsg :STIME ?stime1 .  
?ctrlmsg :DPKTS ?pkts1 .  
?ctrlmsg :DOCTETS ?octets1 .  
FILTER (?pkts1 < 3 && ?octets1 < 300)  
  
# xFil occurs within the next hour to ?dropbox  
{ SELECT ?target ?dropbox (SUM(?octets) AS ?xfil)  
WHERE {  
?target ?flow ?dropbox .  
?flow :DOCTETS ?octets .  
?flow :STIME ?stime .  
FILTER (?stime > ?ftime1  
&& ?stime - ?ftime1 < 3600)  
} GROUP BY ?target ?dropbox  
HAVING (SUM(?octets) > 200000)  
}  
  
# xFil did NOT happen from target in previous  
# hour (target usually does not send lots of  
# data to external hosts).  
{ SELECT ?target  
{ SELECT ?target (SUM(?octets) as ?outRate)  
WHERE {  
?target ?flow ?dst .  
?flow :DOCTETS ?octets .  
?flow :STIME ?stime .  
FILTER (?stime < ?stime1  
&& ?stime1 - ?stime < 3600)  
} GROUP BY ?target ?dst  
} GROUP BY ?target  
HAVING (MAX(?outRate) < 100000)  
}
```

Querying for Chains of Activity



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Path Query



Time Window ➤

00 : 00 : 00

Message Count ➤

HOST 1

IP...

MSG 1

HOST 2

IP...

MSG 2

HOST 3

IP...

MSG 3

HOST 4

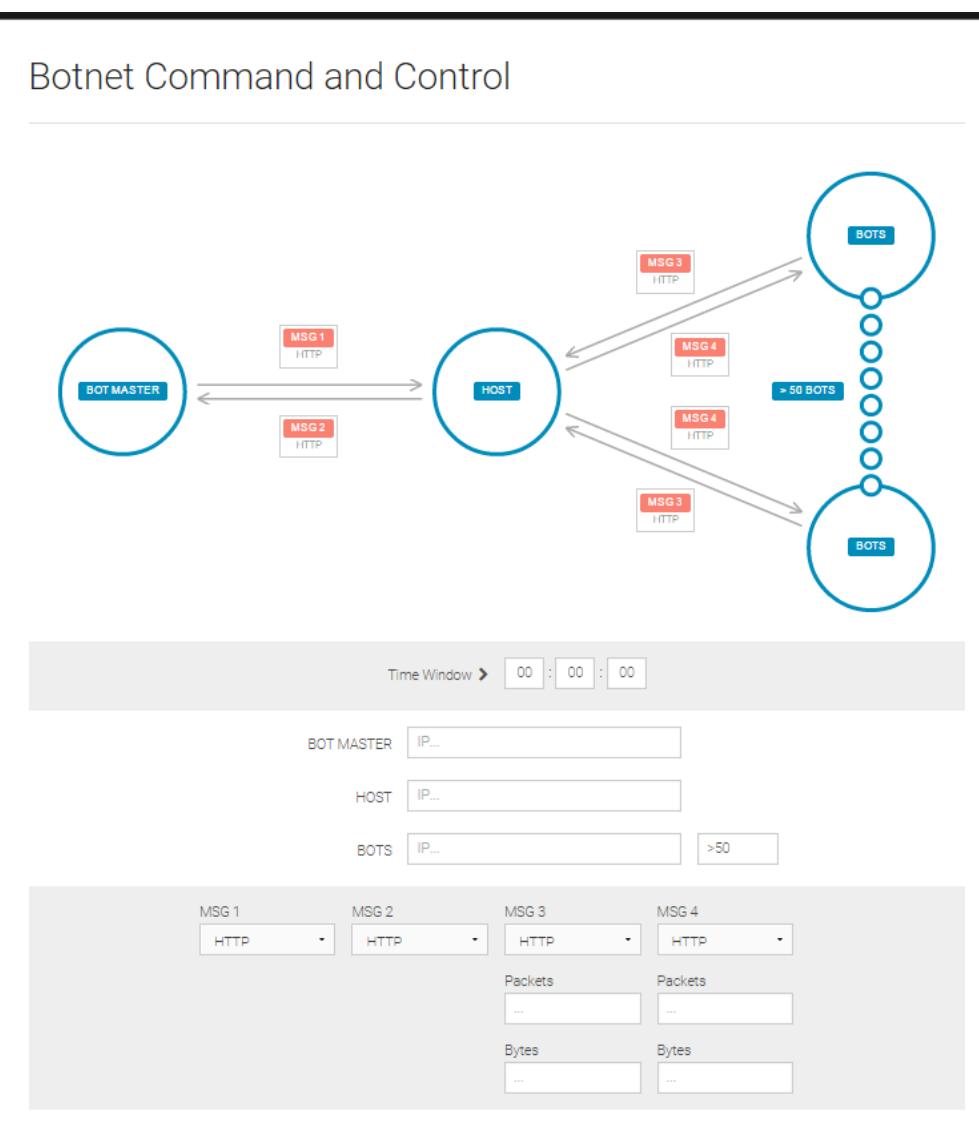
IP...

Botnet Command and Control



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

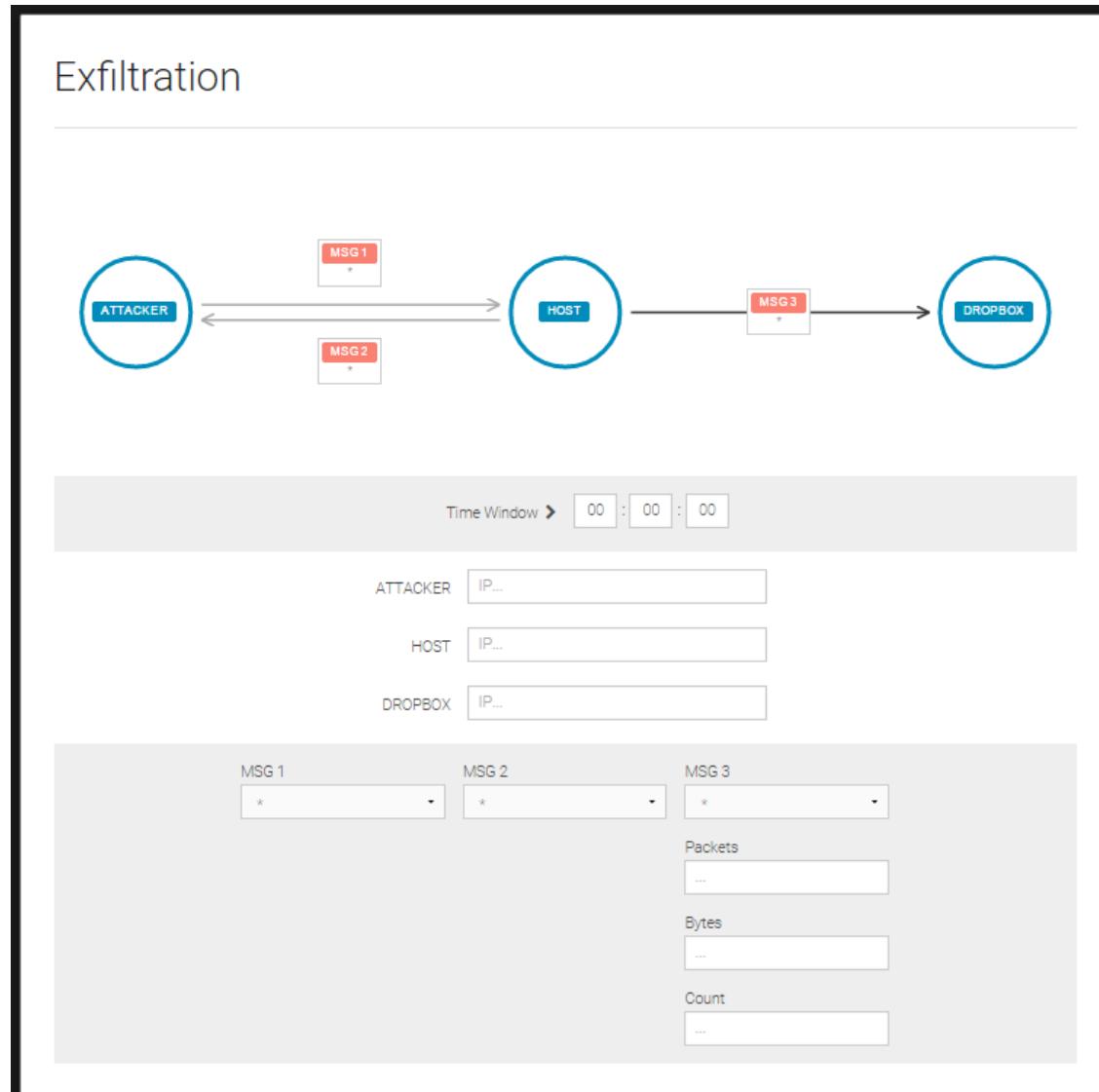


Exfiltration



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965



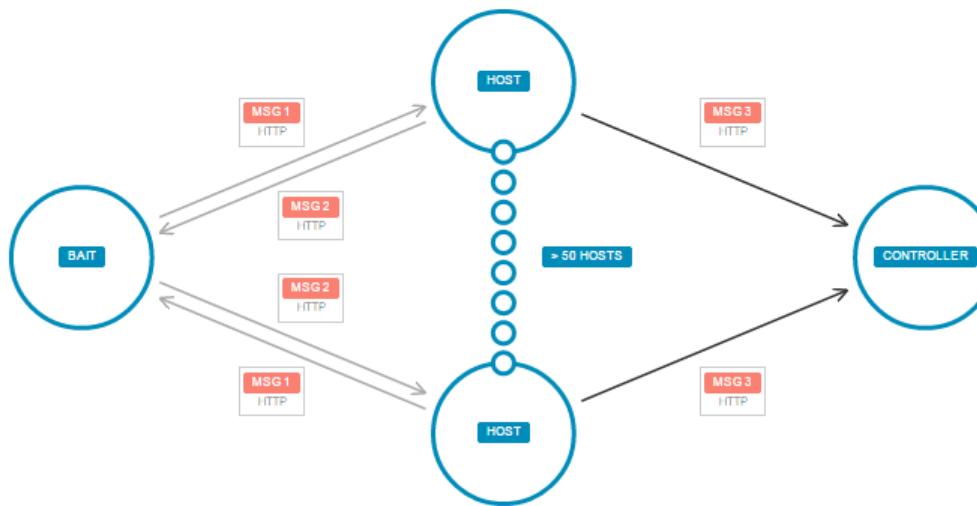
Watering Hole



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Watering Hole



Time Window > 00 : 00 : 00

HOSTS >50

BAIT

CONTROLLER

MSG 1 MSG 2 MSG 3
HTTP HTTP HTTP

Unclassified

August 15, 2017

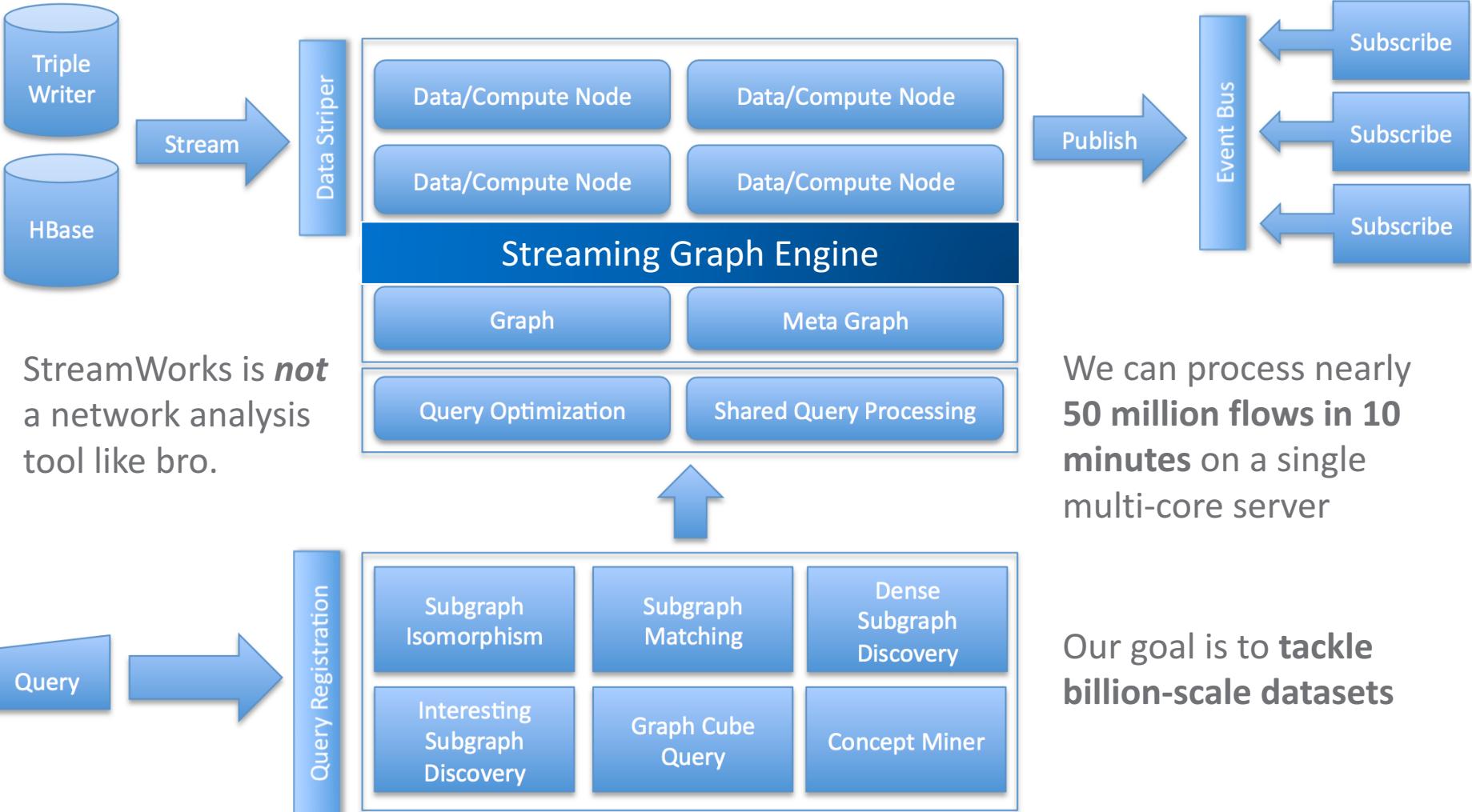
11



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

The StreamWorks Architecture

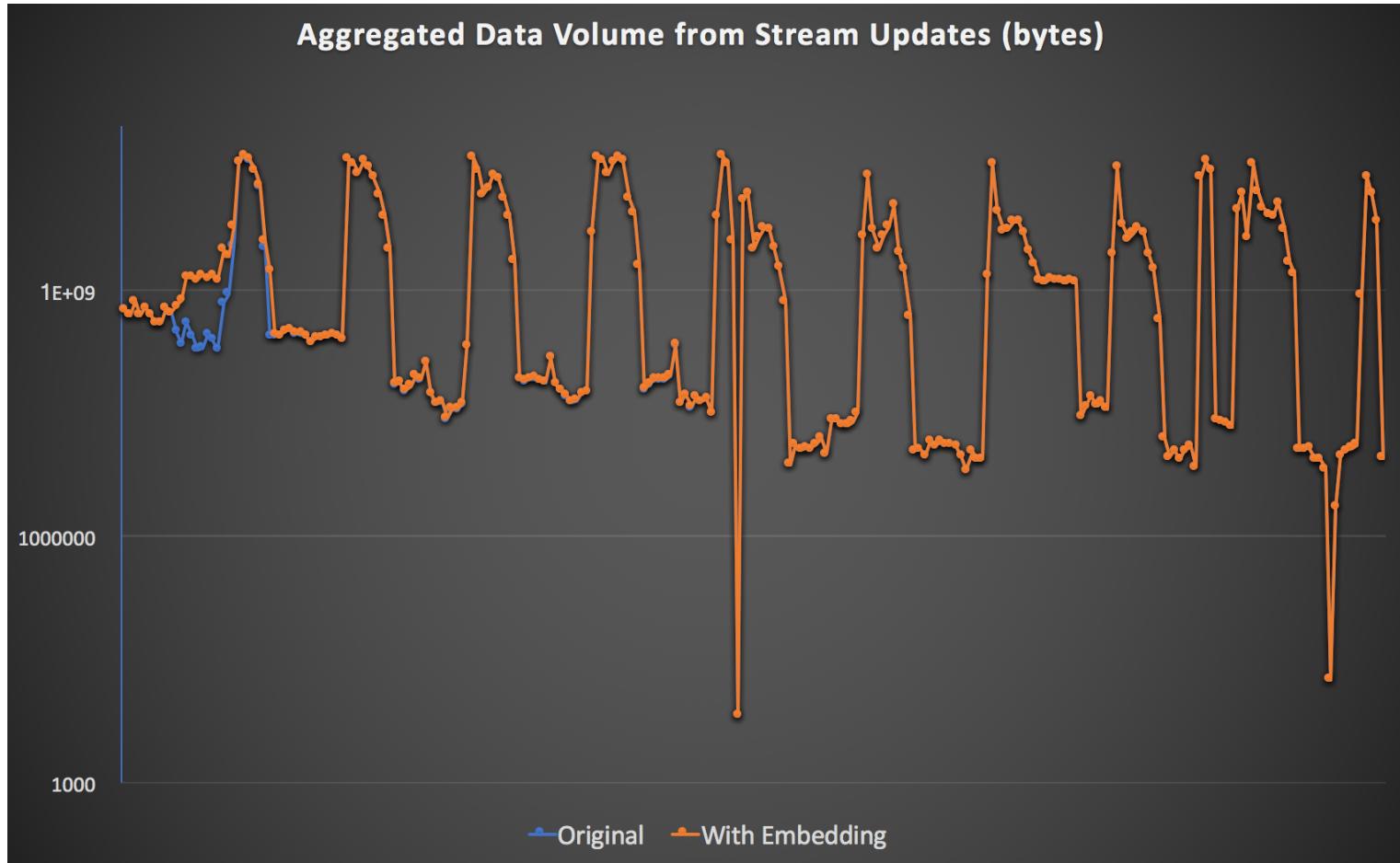




Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Finding the Needle in a Haystack

- ▶ Embedded multiple embeddings of exfiltration into a large-scale dataset

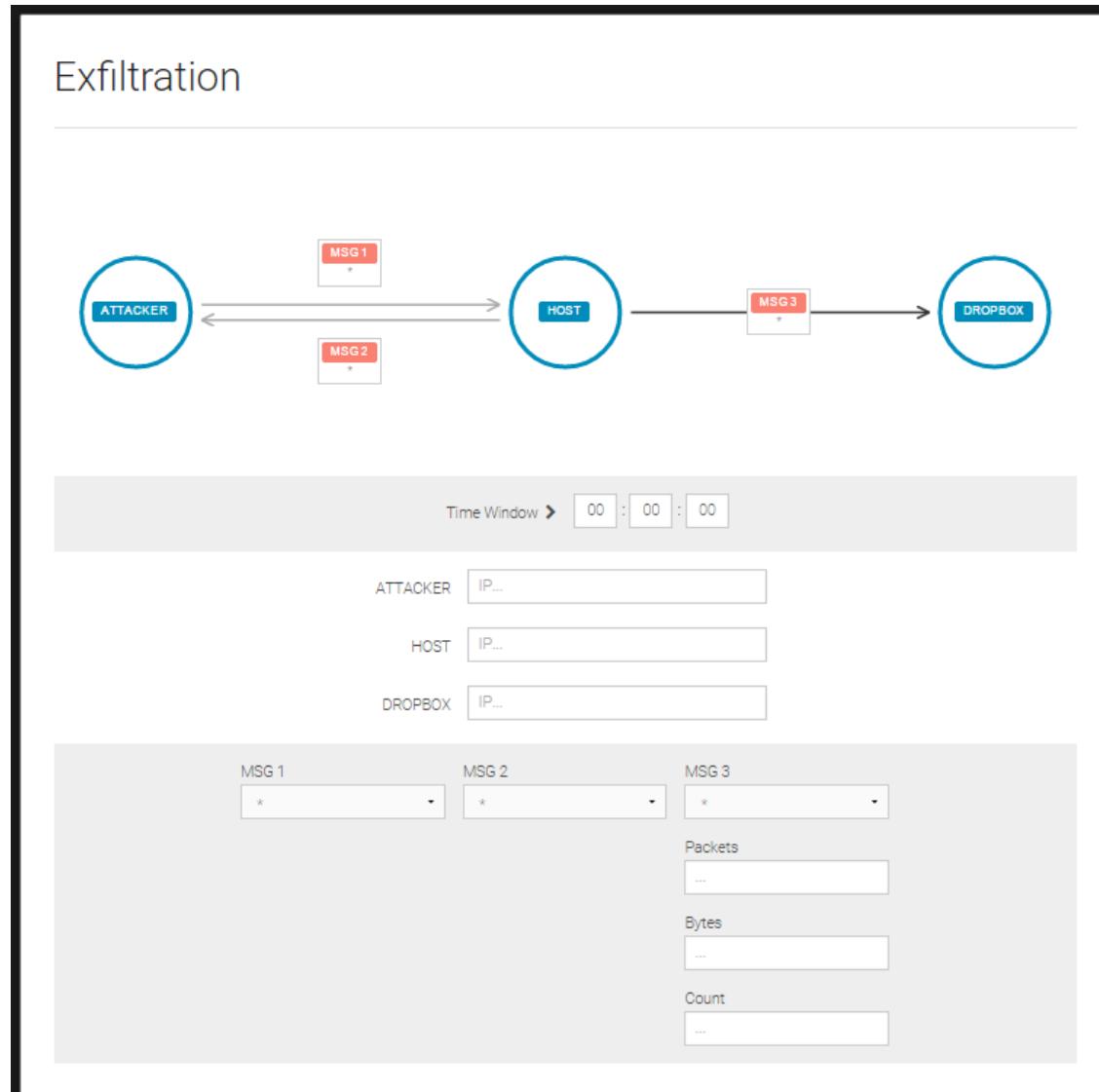


Exfiltration

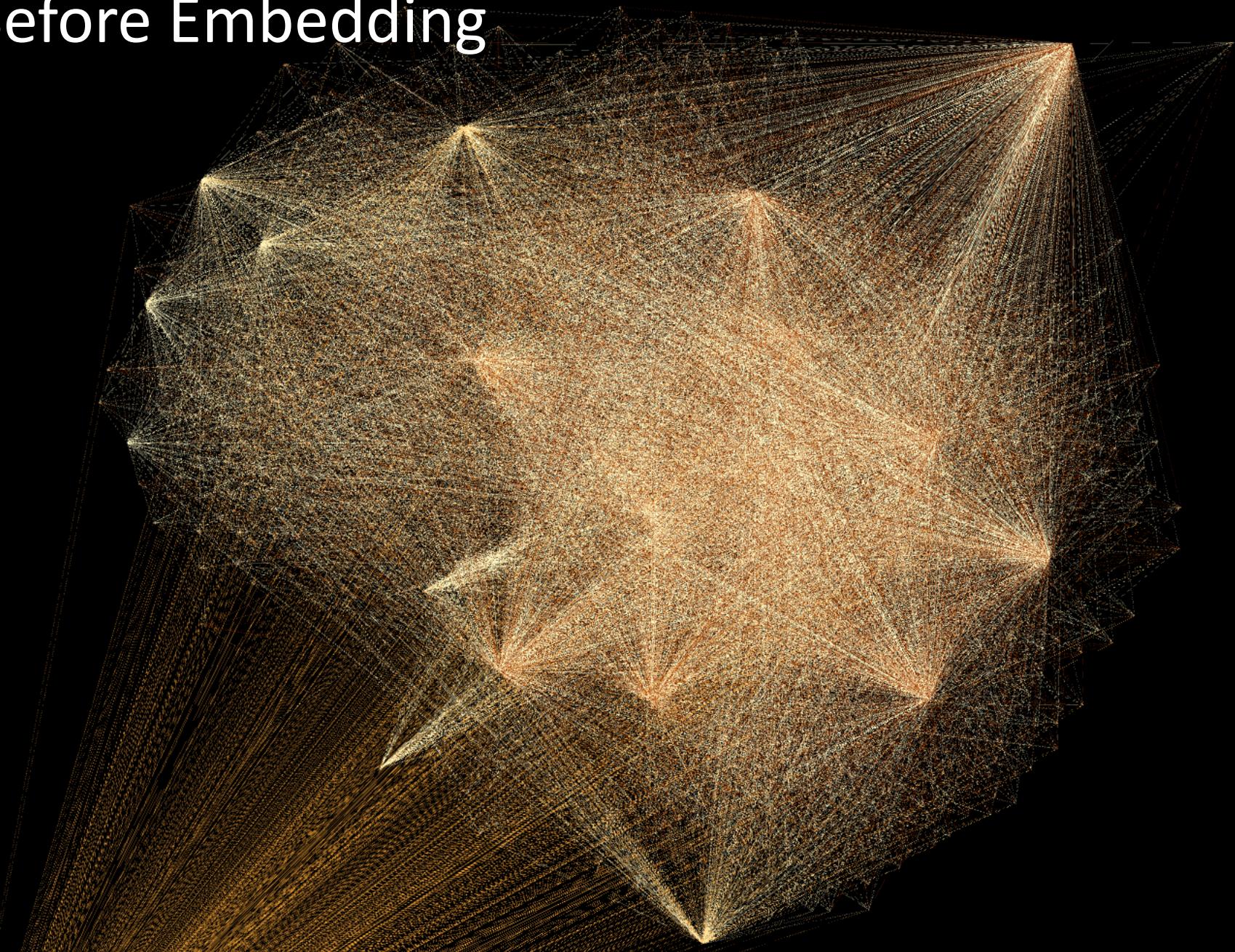


Pacific Northwest
NATIONAL LABORATORY

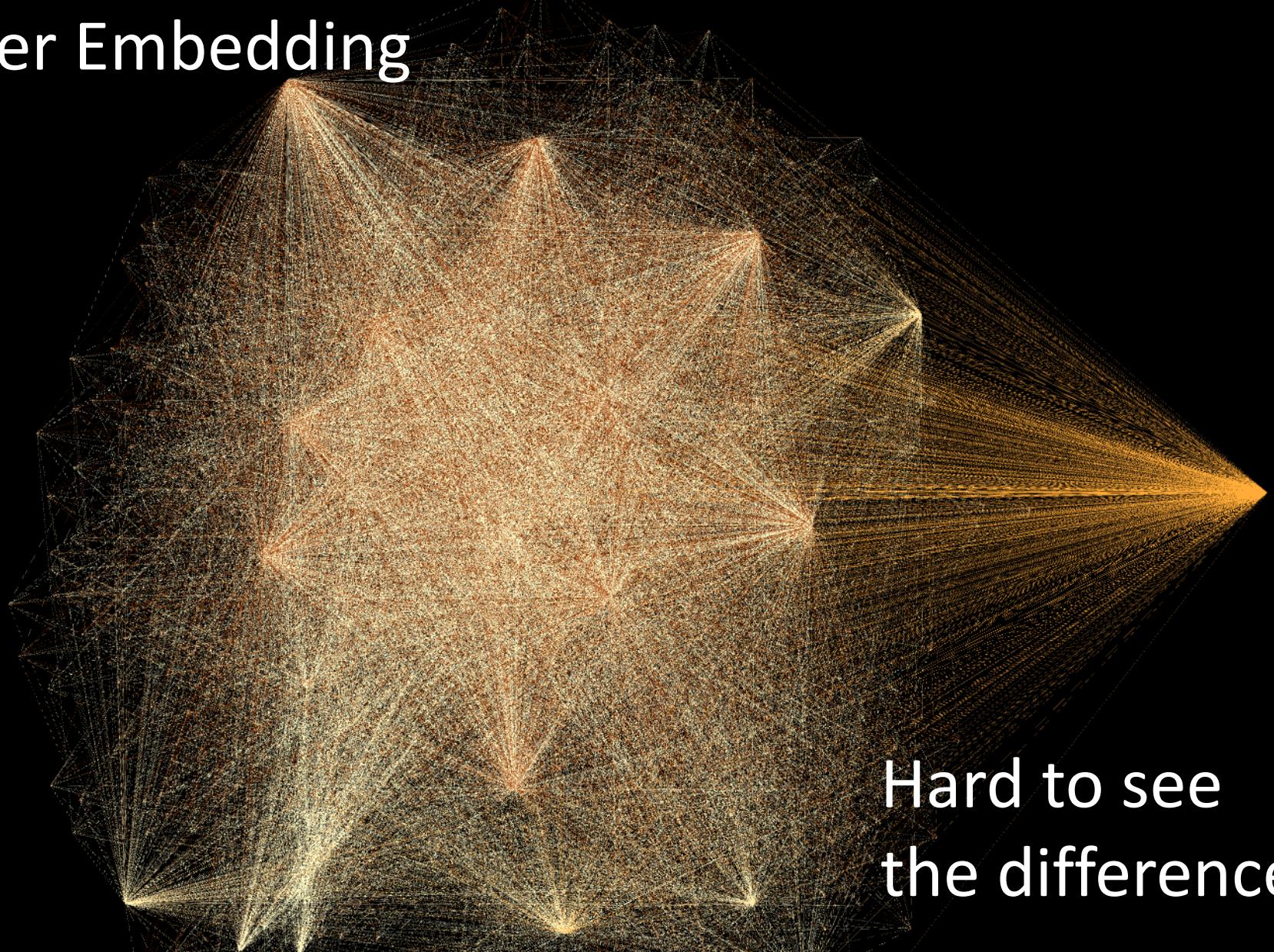
Proudly Operated by Battelle Since 1965



Before Embedding



After Embedding



Hard to see
the difference!

Visualization of Matching Patterns



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

The screenshot shows a StreamWorks visualization interface titled "STREAMWORKS". The main area displays a complex graph of nodes (represented by colored circles) and directed edges (represented by arrows). A specific path or pattern is highlighted in blue. The graph is highly interconnected, with many nodes having multiple incoming and outgoing edges. In the bottom right corner of the visualization window, there is a small control panel with a play/pause button, a progress bar from 01:43 to -01:57, and two circular status indicators labeled "164" and "123".

To the right of the visualization is a terminal window displaying command-line logs. The logs show the execution of a script named "runBroker.sh" and the processing of a file named "vast_3799.tsv". The logs include numerous "INFO" messages indicating the submission of a batch job, the sending of messages to topics like "/topic/cass.demo-result", and the processing of streaming search queries with various LHS and RHS parameters.

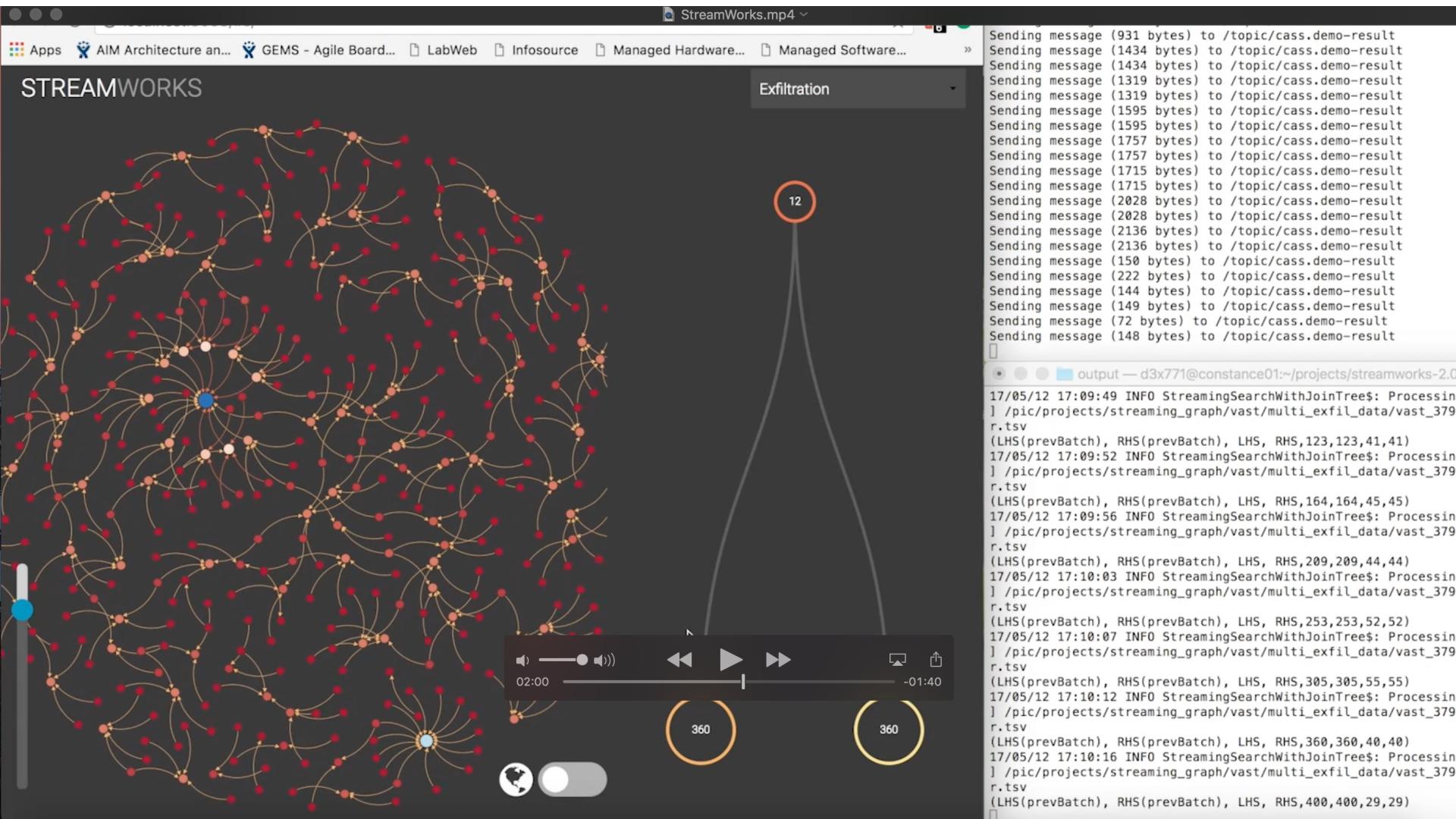
```
[d3x771@constance01 streamworks]$ ./runBroker.sh
Starting!
Submitted batch job 1397630
Sending message (424 bytes) to /topic/cass.demo-result
Sending message (424 bytes) to /topic/cass.demo-result
Sending message (660 bytes) to /topic/cass.demo-result
Sending message (660 bytes) to /topic/cass.demo-result
Sending message (931 bytes) to /topic/cass.demo-result
Sending message (931 bytes) to /topic/cass.demo-result
Sending message (1434 bytes) to /topic/cass.demo-result
Sending message (1434 bytes) to /topic/cass.demo-result
Sending message (1319 bytes) to /topic/cass.demo-result
Sending message (1319 bytes) to /topic/cass.demo-result
Sending message (1595 bytes) to /topic/cass.demo-result
[...]
17/05/12 17:09:32 INFO StreamingSearchWithJoinTree$: Processing /pic/projects/streaming_graph/vast/multi_exfil_data/vast_3799.tsv
(LHS(prevBatch), RHS(prevBatch), LHS, RHS,0,0,0,0)
17/05/12 17:09:35 INFO StreamingSearchWithJoinTree$: Processing /pic/projects/streaming_graph/vast/multi_exfil_data/vast_3799.tsv
(LHS(prevBatch), RHS(prevBatch), LHS, RHS,0,0,11,11)
17/05/12 17:09:37 INFO StreamingSearchWithJoinTree$: Processing /pic/projects/streaming_graph/vast/multi_exfil_data/vast_3799.tsv
(LHS(prevBatch), RHS(prevBatch), LHS, RHS,11,11,17,17)
17/05/12 17:09:40 INFO StreamingSearchWithJoinTree$: Processing /pic/projects/streaming_graph/vast/multi_exfil_data/vast_3799.tsv
(LHS(prevBatch), RHS(prevBatch), LHS, RHS,28,28,24,24)
17/05/12 17:09:43 INFO StreamingSearchWithJoinTree$: Processing /pic/projects/streaming_graph/vast/multi_exfil_data/vast_3799.tsv
(LHS(prevBatch), RHS(prevBatch), LHS, RHS,52,52,37,37)
17/05/12 17:09:46 INFO StreamingSearchWithJoinTree$: Processing /pic/projects/streaming_graph/vast/multi_exfil_data/vast_3799.tsv
(LHS(prevBatch), RHS(prevBatch), LHS, RHS,89,89,34,34)
17/05/12 17:09:49 INFO StreamingSearchWithJoinTree$: Processing /pic/projects/streaming_graph/vast/multi_exfil_data/vast_3799.tsv
(LHS(prevBatch), RHS(prevBatch), LHS, RHS,123,123,41,41)
```

Visualization of matching patterns



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1945

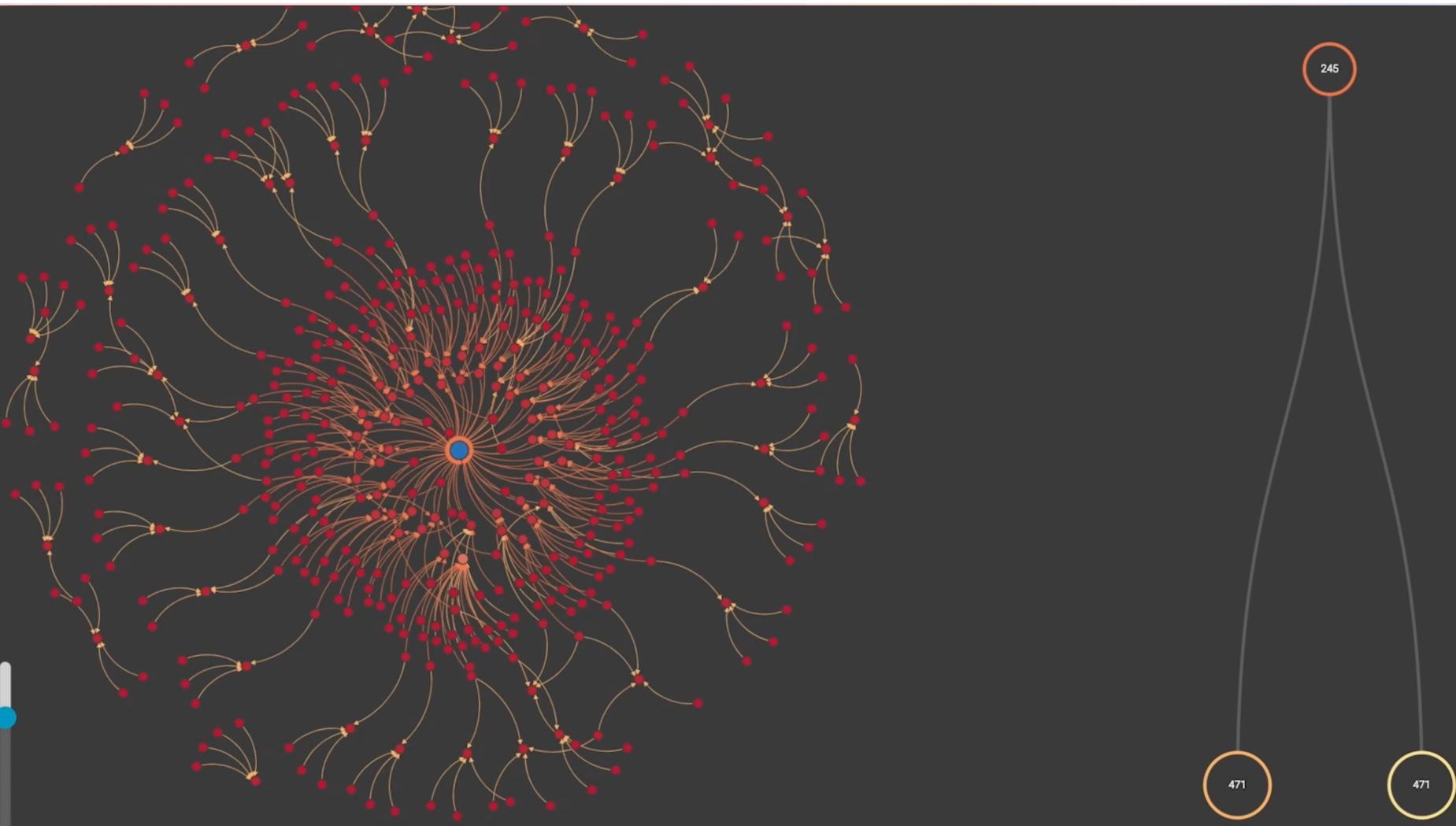


Visualization of Matching Patterns



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965



Visualization of Matching Patterns

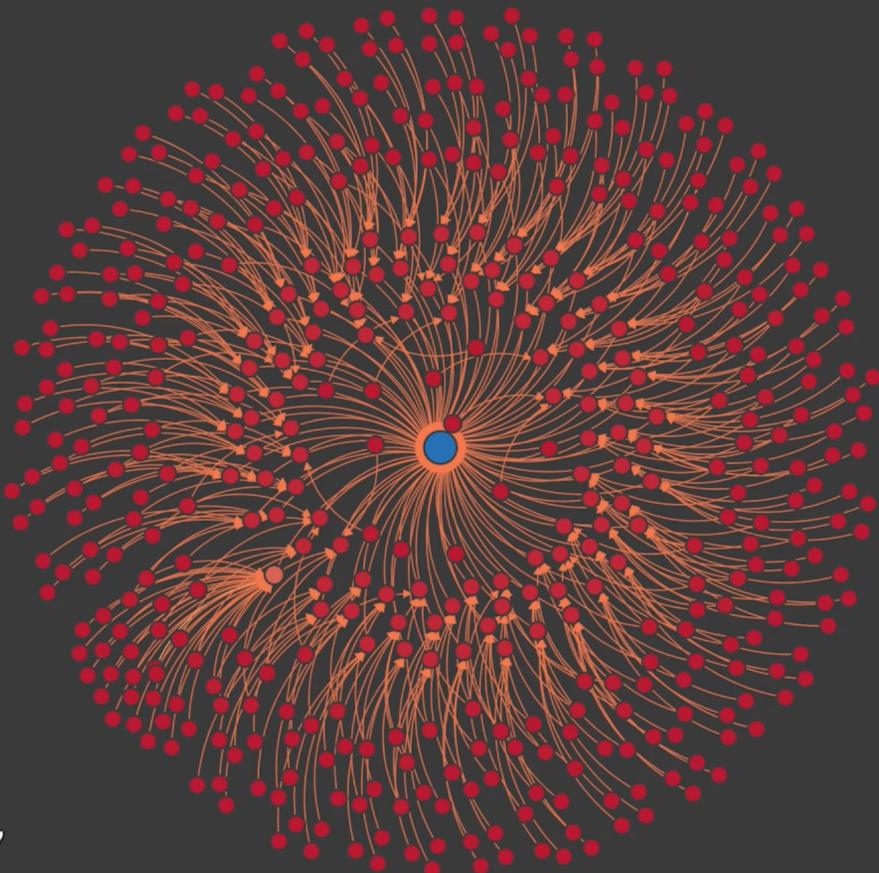


Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

STREAMWORKS

Exfiltration



Providing a geographical perspective



Another example of Geo-View



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965





Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Competition

Product	Streaming	Graph Search	Visual Analytics
StreamWorks	✓	✓	✓
SQRRL Enterprise	✗	✓	✓
Apache Spark	✓	✗	✗
Neo4J	✗	✓	✗

- We obtained 10-100x improvement in runtime on an internet backbone traffic flow dataset.
- Filed US Patent on graph based pattern matching technology



Pacific Northwest
NATIONAL LABORATORY



TRANSITION TO PRACTICE

THANK YOU!

StreamWorks

sutanay.choudhury@pnnl.gov



Homeland
Security
Science and Technology

This technology has been brought to you by the DHS S&T Cyber Security Division Transition to Practice (TTP) Program. For more information, contact ST.TTP@hq.dhs.gov