Enabling Secure Research and Operations with Elastic Stack and multi-faceted Data Streaming at ORNL

Larry Nichols • nicholslc@ornl.gov
Cyber Security Operations & Engineering
Cyber Security Engineer/ SIEM Admin

John Goodall • jgoodall@ornl.gov
Cyber Security Research Group
Team Lead / Research Scientist

ORNL is managed by UT-Battelle
for the US Department of Energy

Goals

- Deploy a SIEM that increases speed and security
- Establish collaboration with security researchers, leveraging unique capabilities and knowledge
- Reduce overall costs to ORNL
- Increase throughput of data ingestion infrastructure
- Reduce complexity of adding new data sources
- Increase situational awareness through enhanced insights
- Identify anomalous activity that may indicate malicious activity

Agenda

- Splunk to Elastic transition
- Production Architecture
- Research/OPS Development Cluster Architecture
- X-Pack
- Situ
- Q&A

Splunk vs. Elastic

Splunk

Cost: $$$$$
Size: License based on indexing needs
Speed: Searches in minutes+
Security: Roles can lock down index types
Dev. Instances: 50G per day for 30 day increments
Integration: Data sources mean additional costs, limited versatility

Elastic

Cost: $$
Size: License based on production node requirements
Speed: Searches in seconds+
Security: (X-Pack) Lock down to field level per user
Dev. Instances: Unlimited, non-expiring including full support
Integration: Utilized by researches and other Labs, will support needs for expanded searching
Intelligence: X-Pack machine learning, graph, and more

Current Production Architecture

- 25 Elasticsearch nodes (Docker) across 25 VMs on HP Converged Infrastructure
- >2 Billion documents/ ~1.5TB/Day
- Maintaining over 150 days worth of data, ~279+ billion documents

Any time. Any place. Any device.

OAK RIDGE National Laboratory

Why Docker?

- Allows for easy updates and management
- Quickly update Elastic versions/configs and deploy
- Creation of a single template elasticsearch.yml by node type instead of 25!

Current Production Architecture Cont.

- Encryption of data in transit: intra-node, client to Kibana, Kibana to Elasticsearch, Logstash to Elasticsearch, and all API calls
- X-Pack security, Roles-based access control down to field level (read, write, delete)
- LDAP authentication
- SAML SSO capable

Search Time Comparisons

- Used the FW index as a test over 1 billion documents indexed daily
- Due to long search times for a * search over a 24 hour period, “deny” was queried
- Splunk-1Min 26 Seconds
- Elasticsearch- 1.6 Seconds

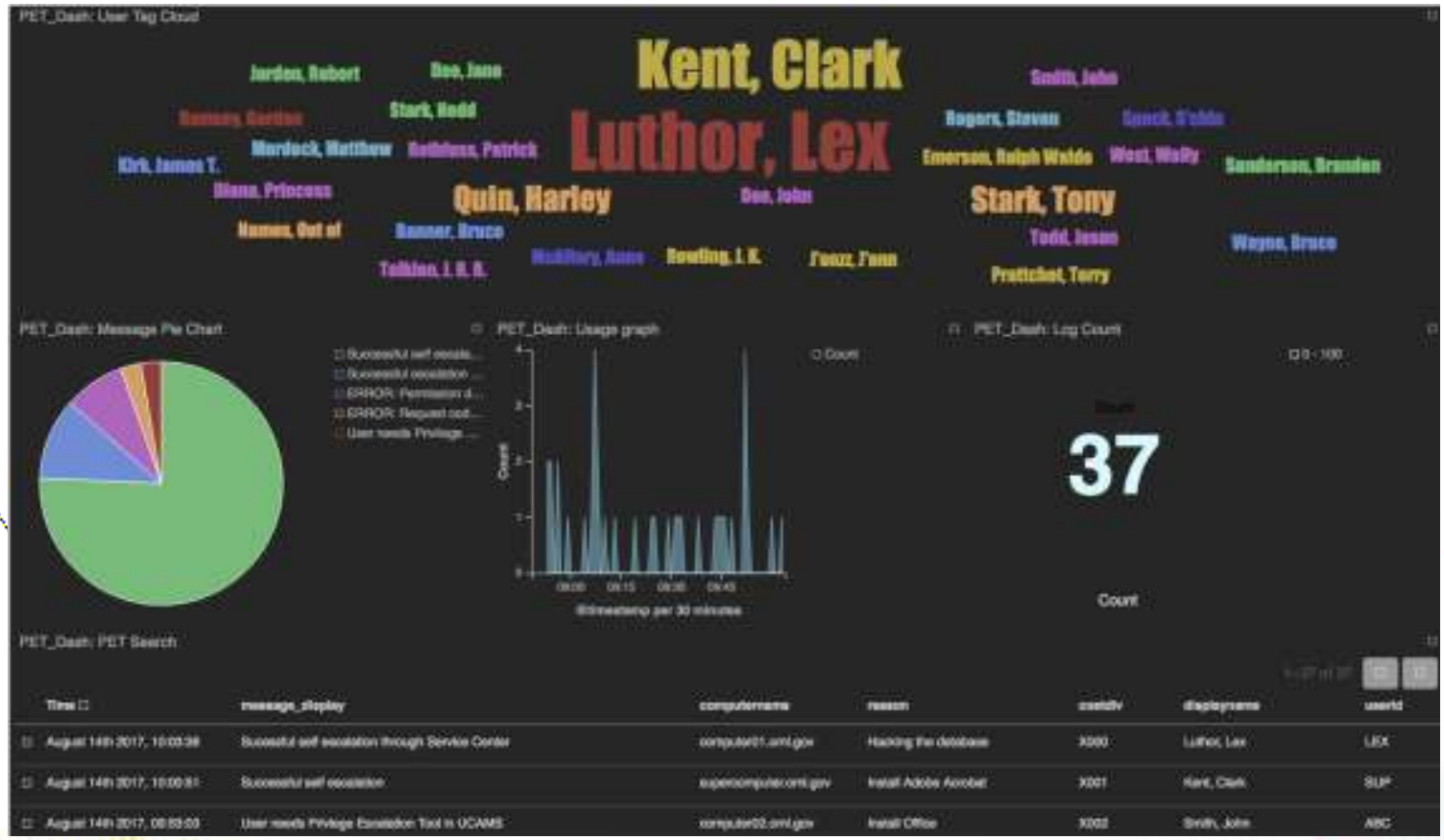
The Journey to production

- Learning Curve
- Elasticsearch Pitfalls
  - Heap memory
  - Shard allocation
  - Hot/Warm Nodes
  - Usage of Master and Coordination Nodes

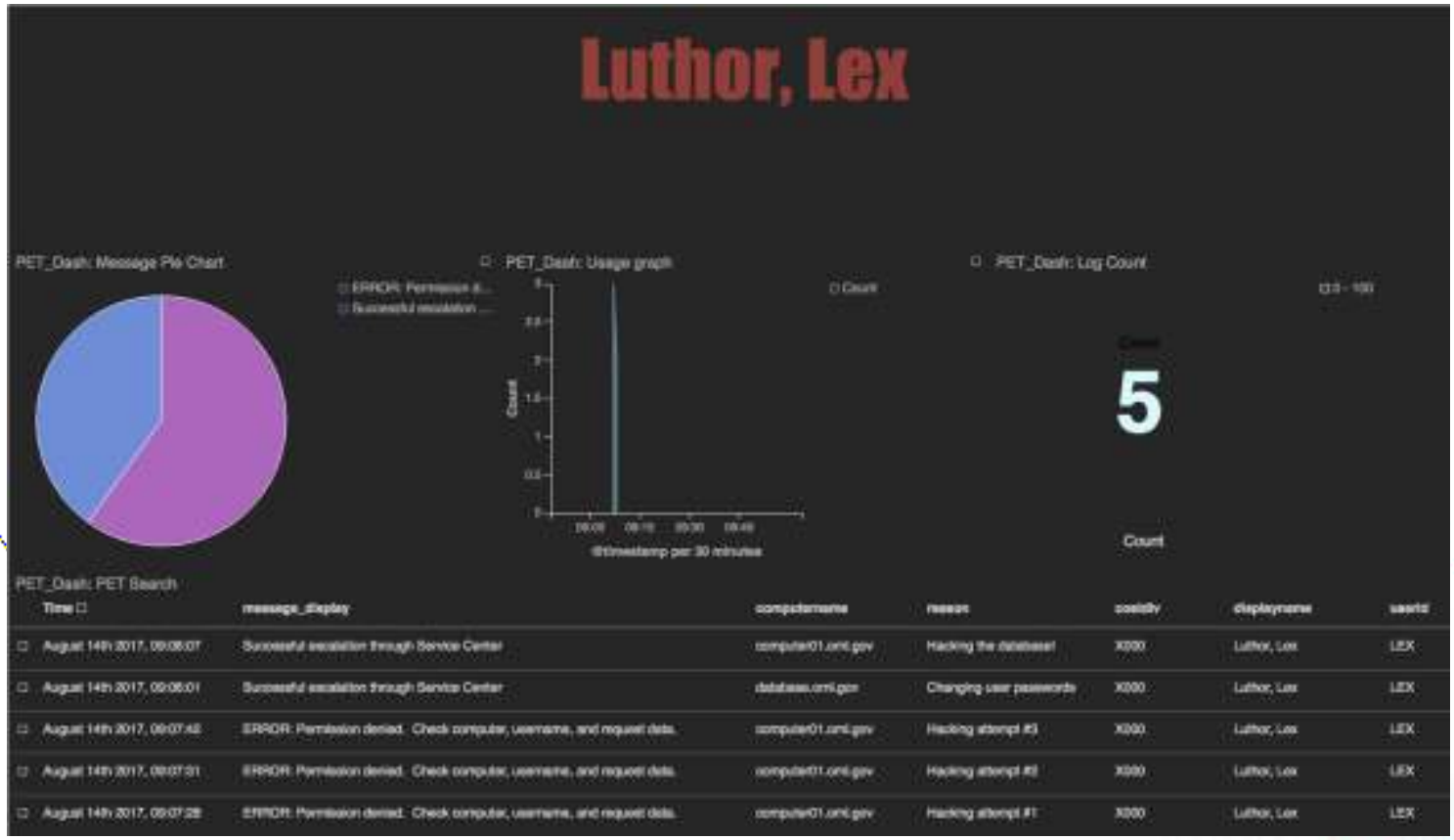
Current Research DEV Architecture

- 6 Elasticsearch nodes (Docker) across 6 physical servers
- 2 Billion documents/ ~1.5TB/Day
- Maintaining ~ 30 days worth of data, ~21 billion documents

Kibana Dashboards



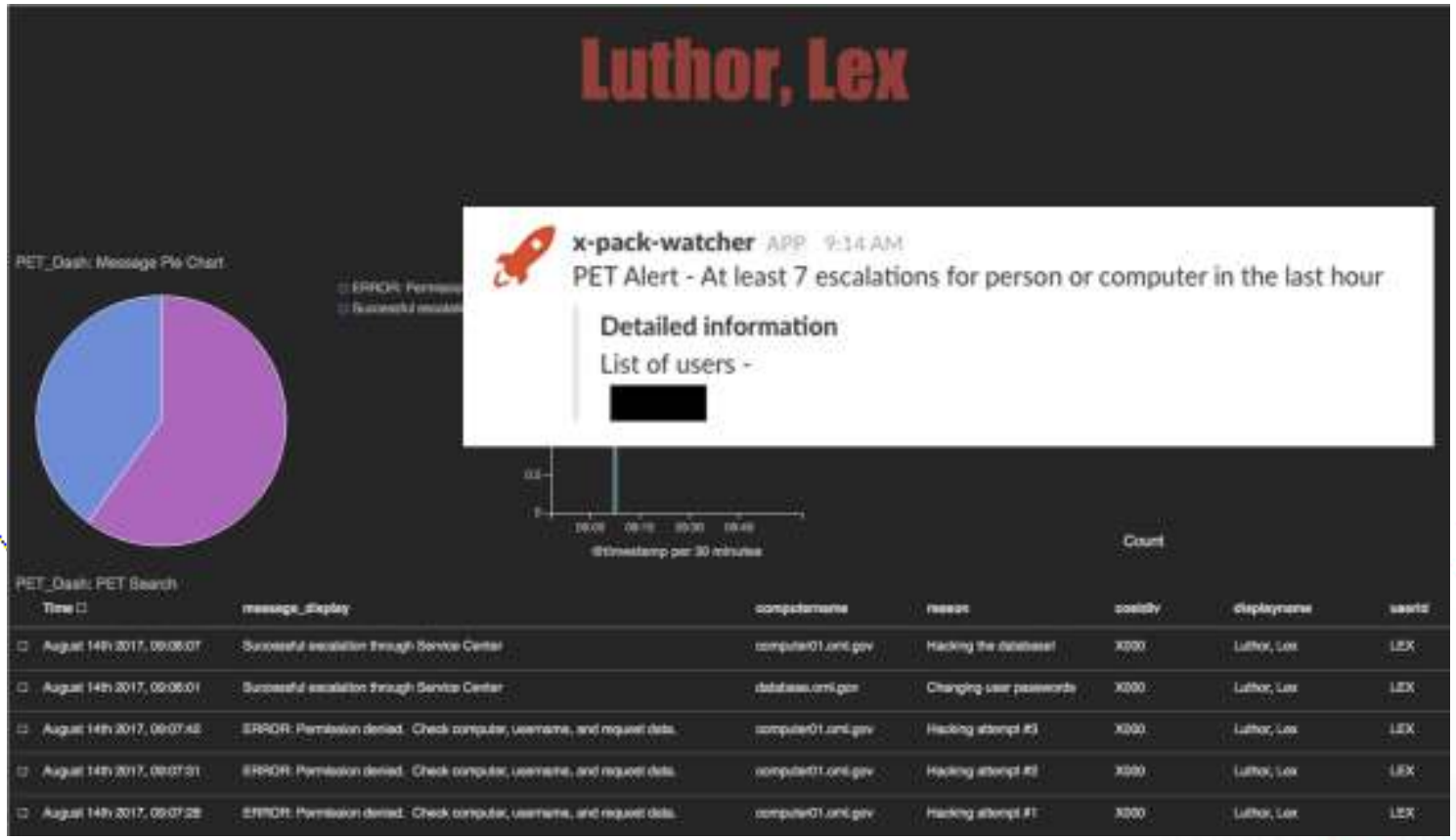
Kibana Dashboards



Any time. Any place. Any device.

OAK RIDGE National Laboratory

Kibana Dashboards

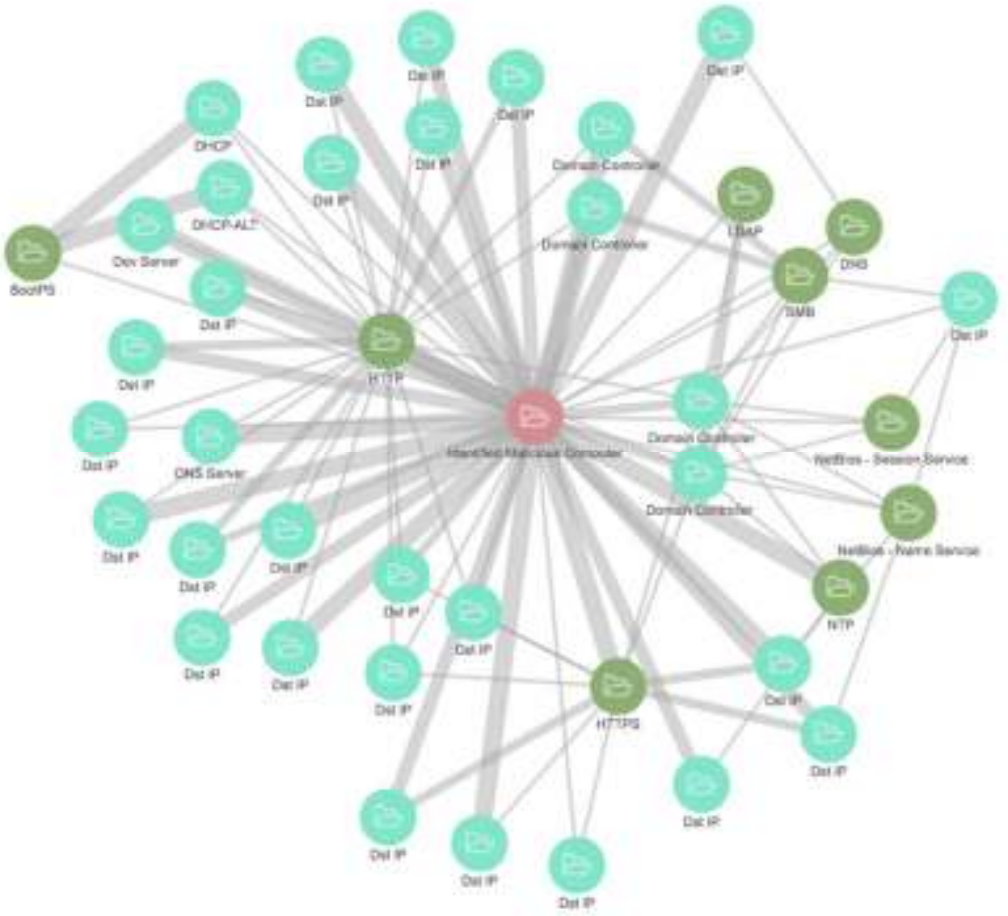


Any time. Any place. Any device.

OAK RIDGE National Laboratory

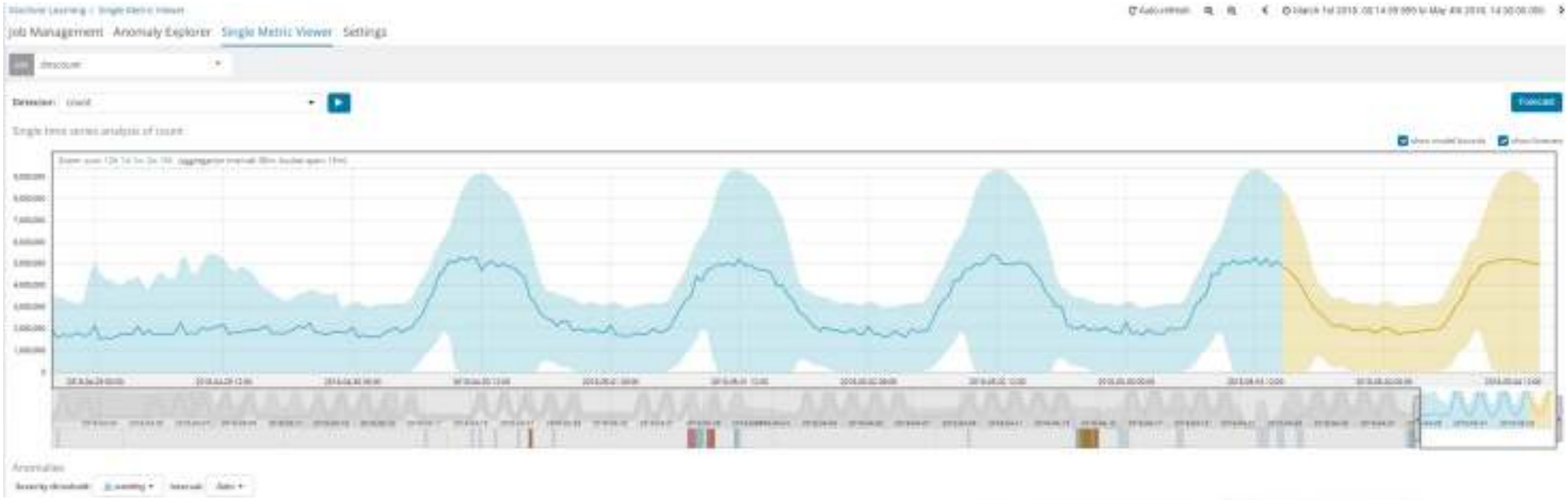
Graph

- Currently used with security focus
- Used to correlate relationships



Timelion

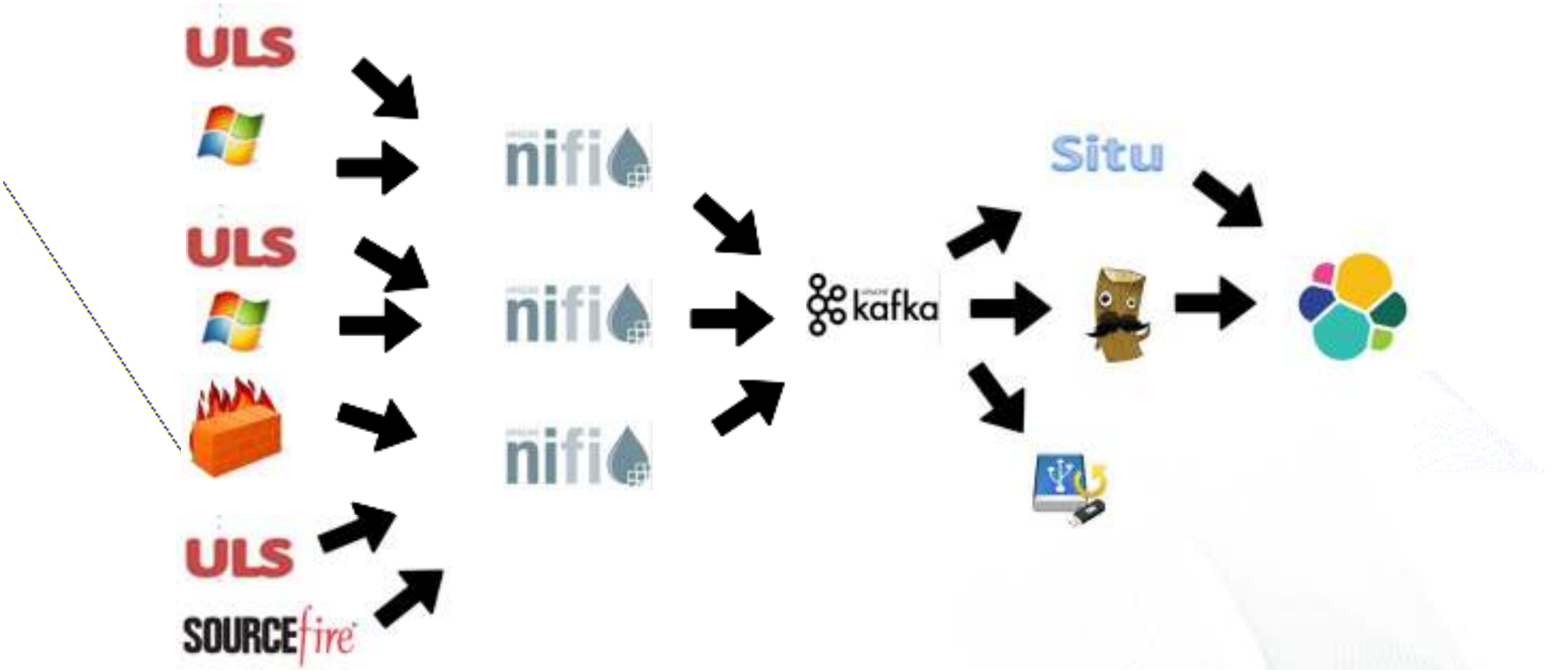
Machine Learning



Any time. Any place. Any device.

OAK RIDGE National Laboratory

Improved Data Streaming Pipeline



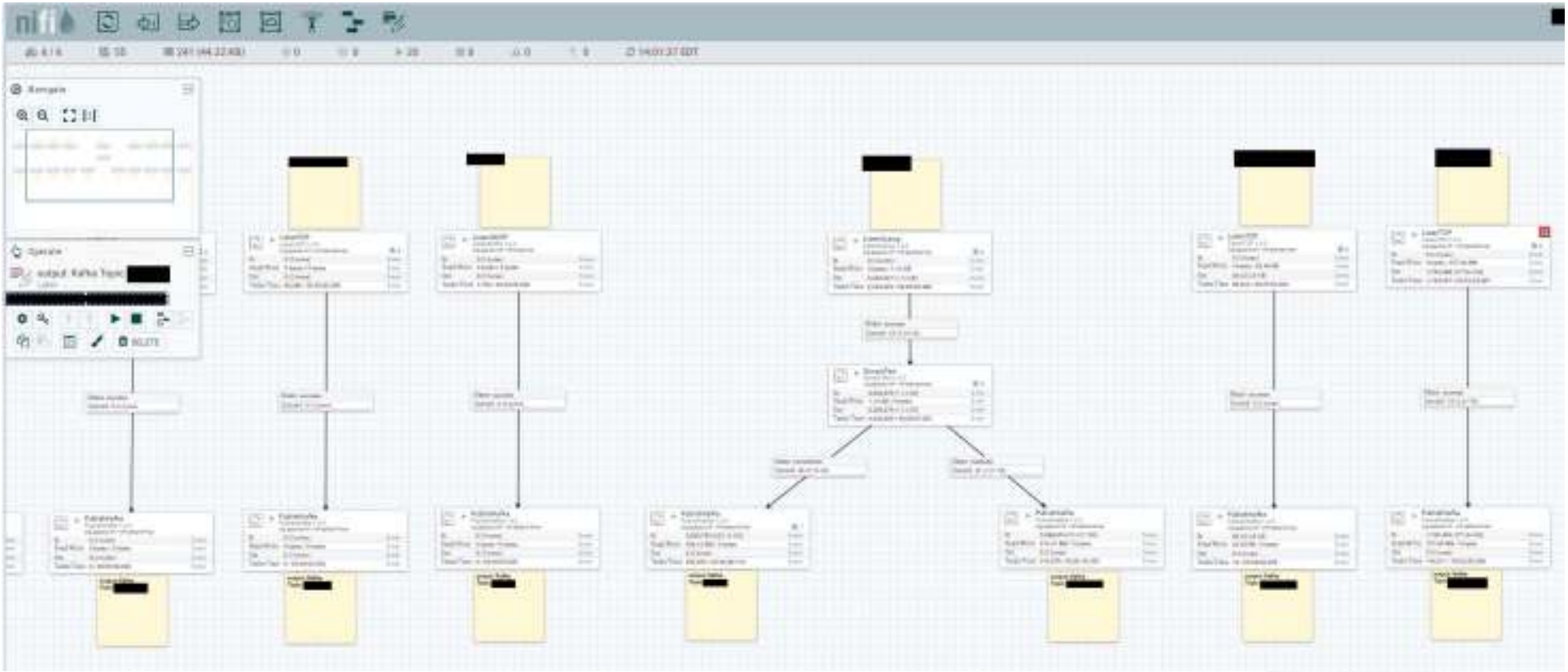
Any time. Any place. Any device.

OAK RIDGE National Laboratory

Nifi

- Web based UI
- Low latency, high throughput
- Data tracking from beginning to end
- Flow modification on the fly
- 50,000 messages a second guaranteed (single node)

Nifi



Kafka

- Similar to a message queue +
- Store message streams in a fault-tolerant way
- Near real-time streaming data pipelines that reliably deliver data
- 100,000 Messages a second (guaranteed single node)

Situ

Anomaly detection for discovering suspicious behavior

OAK RIDGE National Laboratory

Security systems will never be capable of detecting **all** attacks

OAK RIDGE National Laboratory

There is **too much data**
for operators to look at all of it
or to know where to begin

Situ provides operators with a starting point to begin their analysis

OAK RIDGE National Laboratory

Situ

scalable, streaming
anomaly detection
to highlight suspicious activity
within high data rates

Technical approach

- Unsupervised, probabilistic learning
- Heterogeneous data
- Multiple behavior models updated online
  - Privileged Ports
  - DNS activity
  - Producer-Consumer Ratio
- Streaming and search visualizations

- Streaming Demo

Behavior Models

- Each new event is scored according to previous activity for the internal IPs in that event
- Multiple behavior models are created for each IP address
- Each model has a temporal variation
- Scores are on a 0-6 scale
  - 90% of events are 0-1.0
  - Scores of 5 are similar to a 1 in 100,000 event

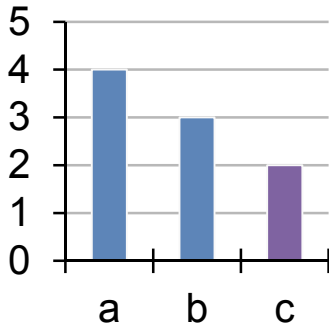
Anomaly detection

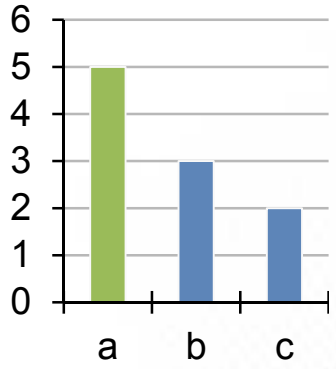
Behavior model is continuously learned from observed data

data: a b c a a a b b **c** **a**

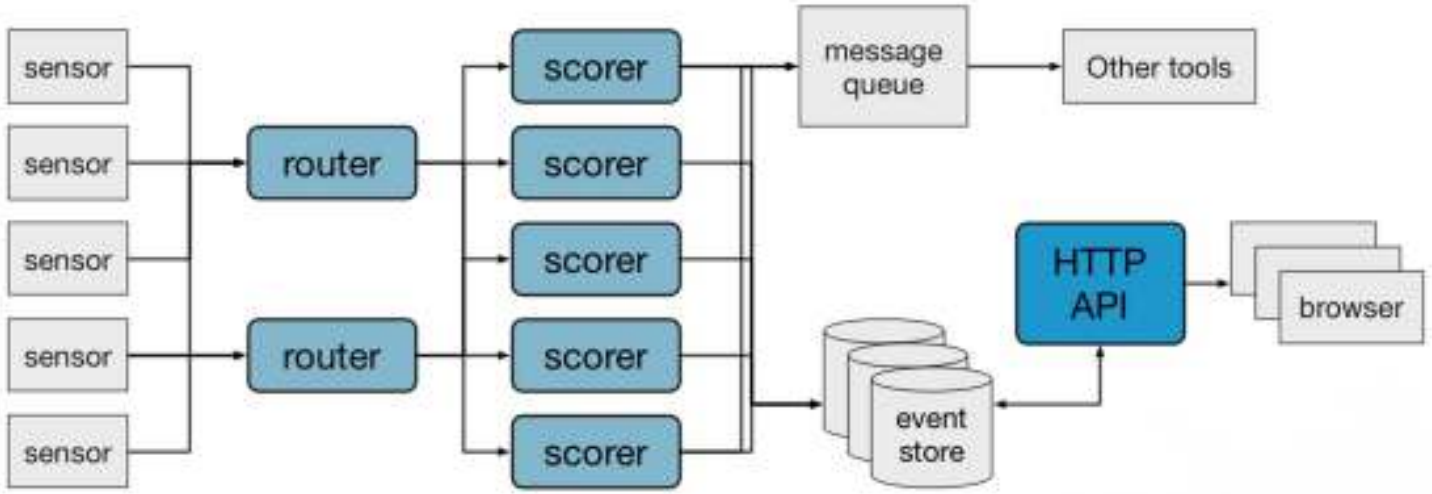
new event: c
more likely events: [b: 3, a: 4]
more likely: 7 / total: 9
score: 0.237

new event: a
more likely events: [ ]
more likely: 0 / total: 10
score: 0.022





Distributed Architecture



Results

- Skaion IARPA Dataset v5
- Background (5b5): 673,141 flows
- Simple attack scenario (5s12): 155,726 flows
- Multiple stepping stones (5s20): 224,859 flows
- Counted number of attack IPs in top 1,000 scores

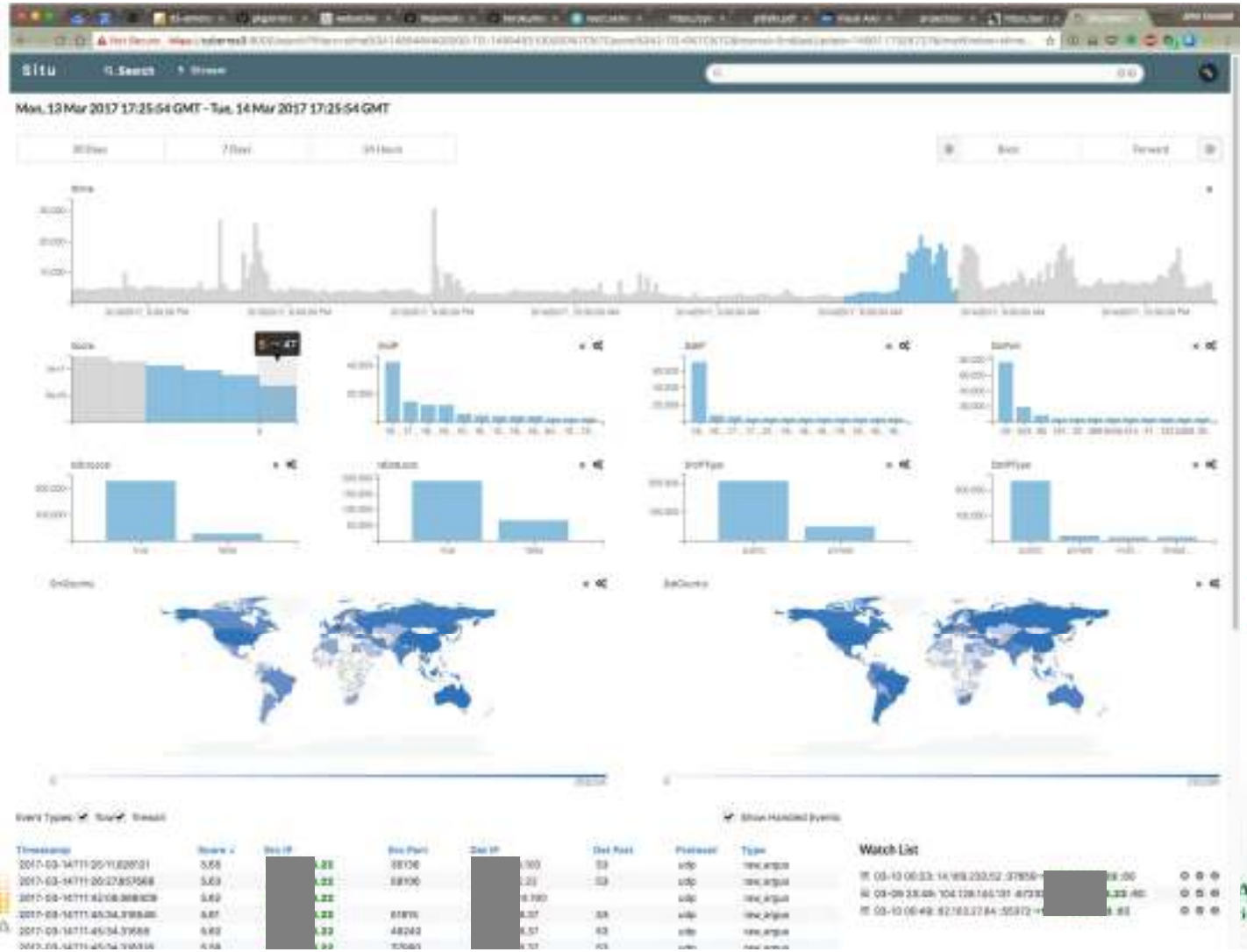
	Part of known attack	Not part of attack
5s12	925	75
5s20	978	22

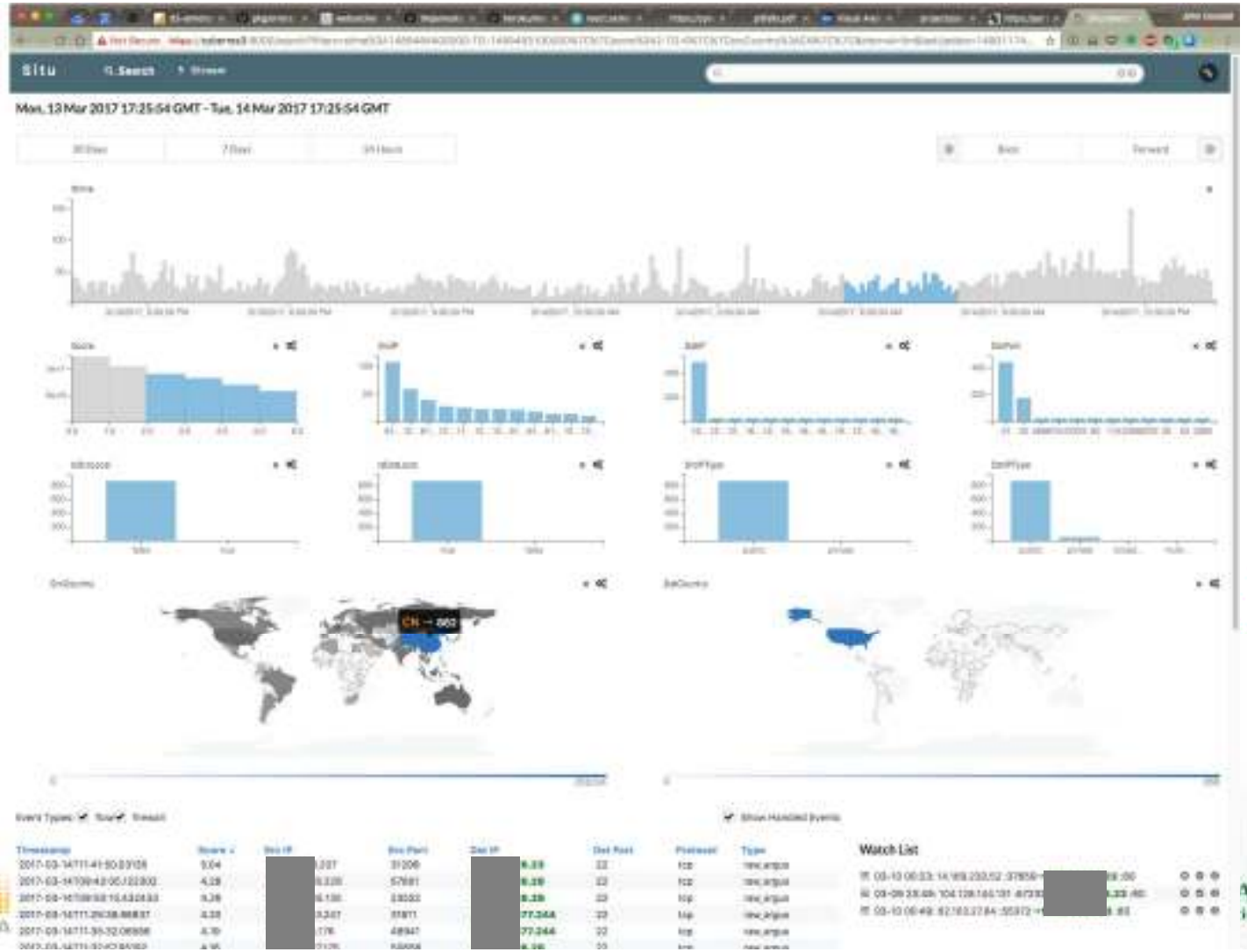
Results

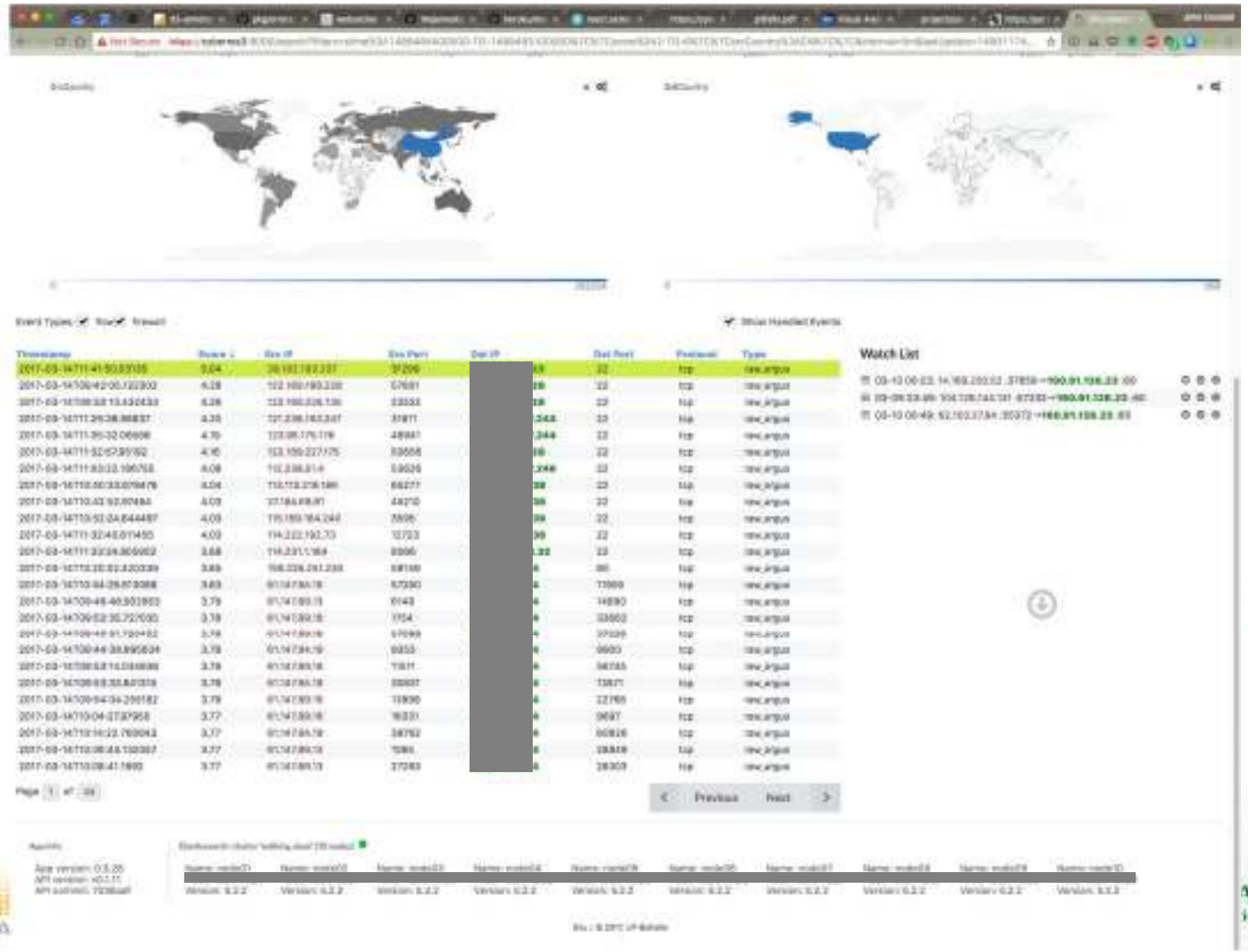
- ORNL's SOC is using Situ on a daily basis
- They report that Situ has uncovered malicious IPs that they were not currently blocking

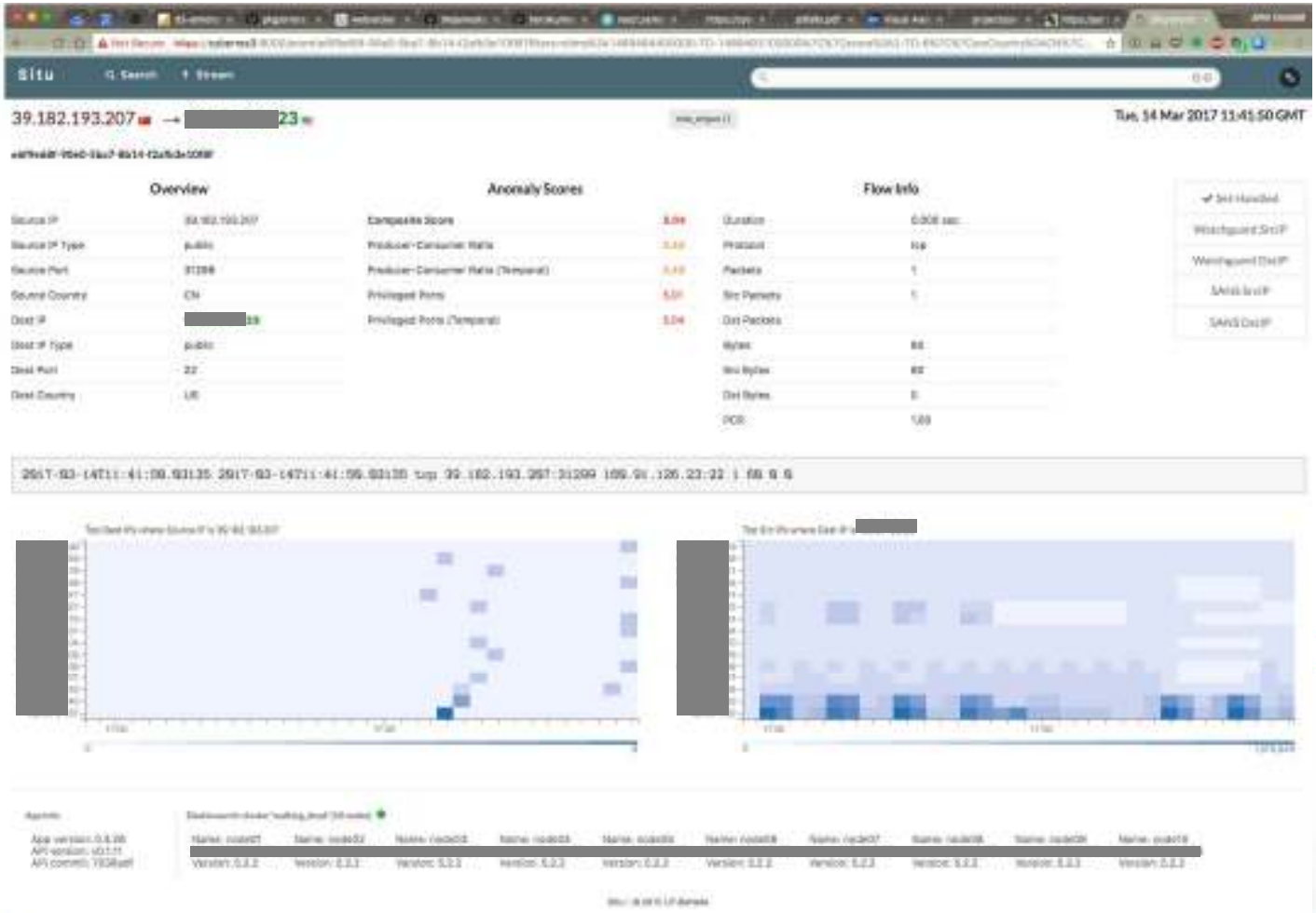
> "Situ is the best tool for detecting and analyzing beaconing traffic known to humankind.
>
> - SOC Senior Engineer, ORNL

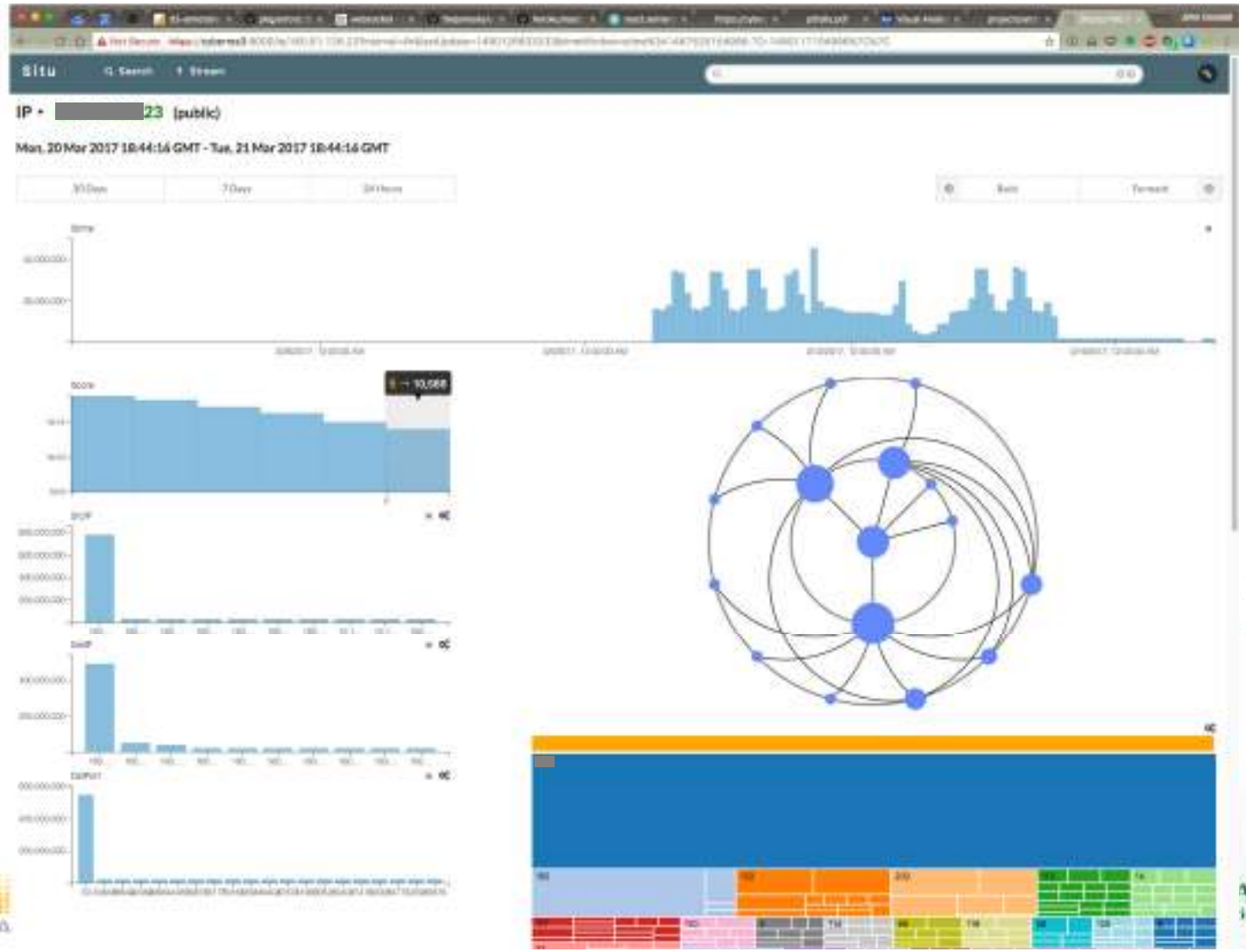
> "Situ has been useful in identifying threats that our commercial tools do not always see or block, assists in identifying deficiencies in our current processes and allows for rapid response to threats by analysts in the Cyber Security Operations Center.
>
> - SOC Team Lead, ORNL

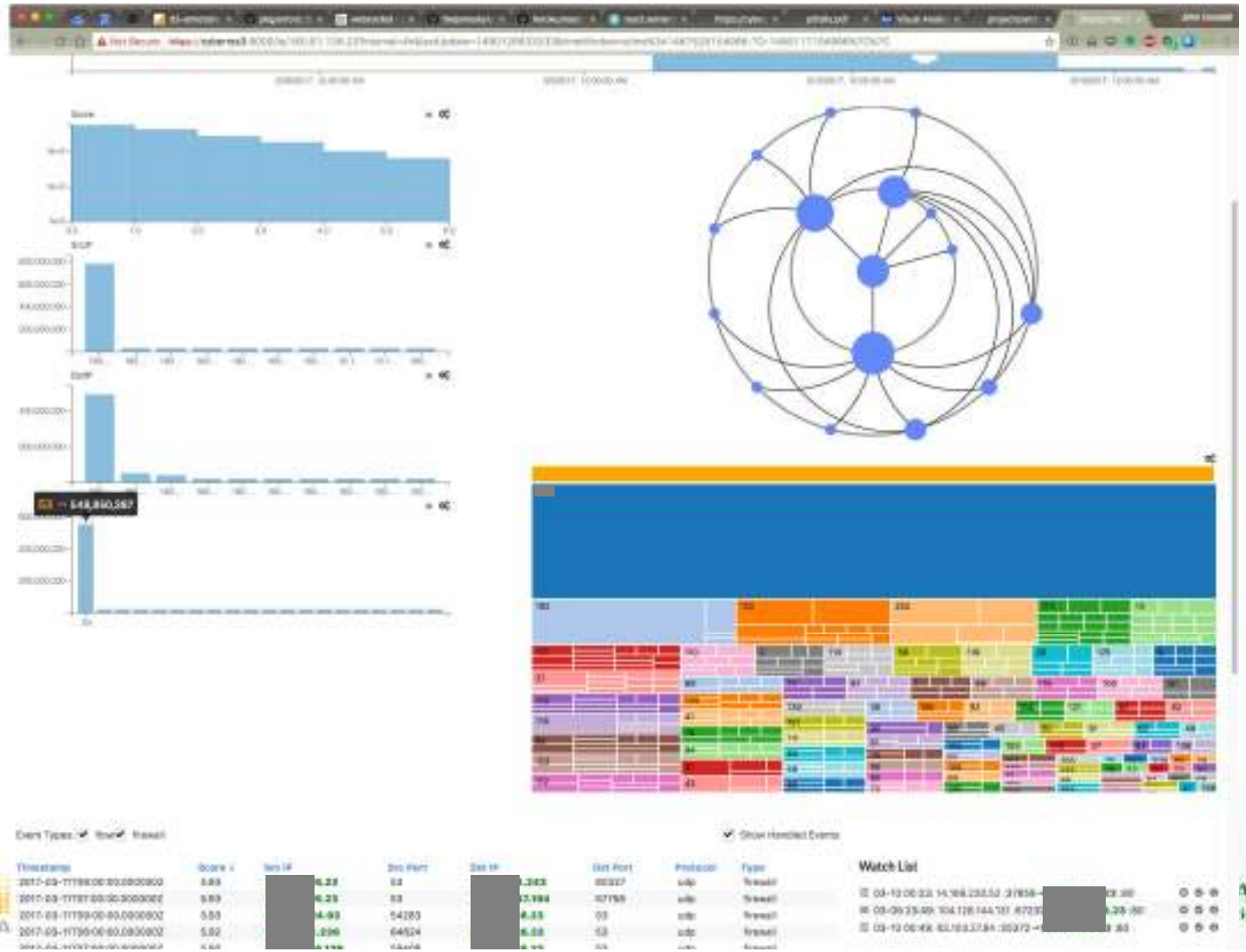


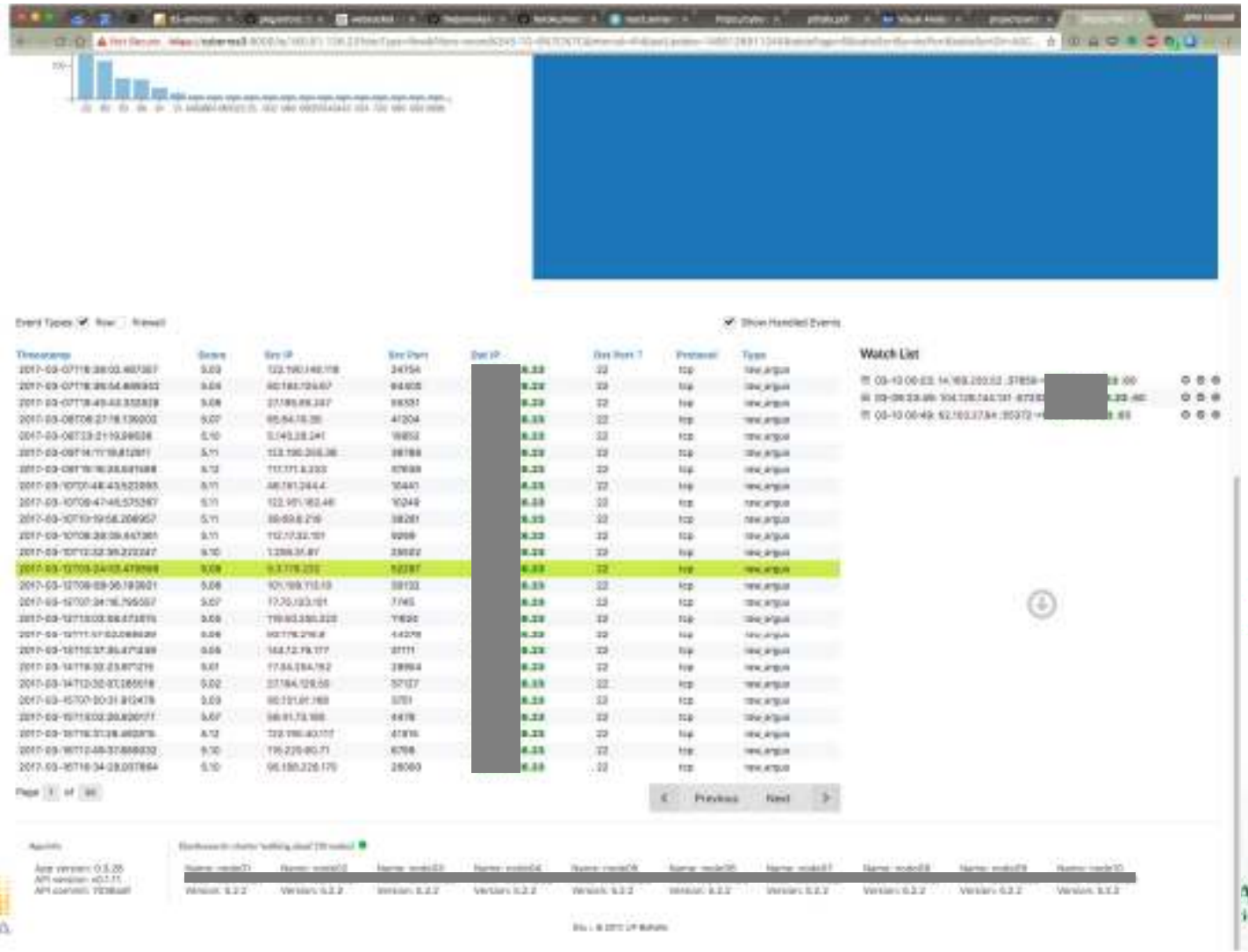





Situ
Search
Stream
39.182.193.207
Tue, 14 Mar 2017 11:41:50 GMT
Overview
Anomaly Scores
Flow Info




Watch List



Situ reduces huge data flows
to a manageable subset
for review by domain experts

OAK RIDGE National Laboratory

Advantages

- Detects unknown attacks
- Requires no labeled data
- Trains and scores online
- Helps operators understand why

Current activity

- Deployment at ORNL in SOC
  - Network flows: 400 million flows / day
  - Firewall logs: 1 billion logs / day
  - Used by tier 1 and 2 analysts in Cyber SOC
- Pilot at large company is ongoing
- Deployment at ORNL in NCCS
  - Network flows: 5.2 million flows / day
  - Part of analysts' routine triage process
- To be piloted at Department of the Interior

Looking for organizations to **pilot** or **license** Situ

Questions?