

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



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

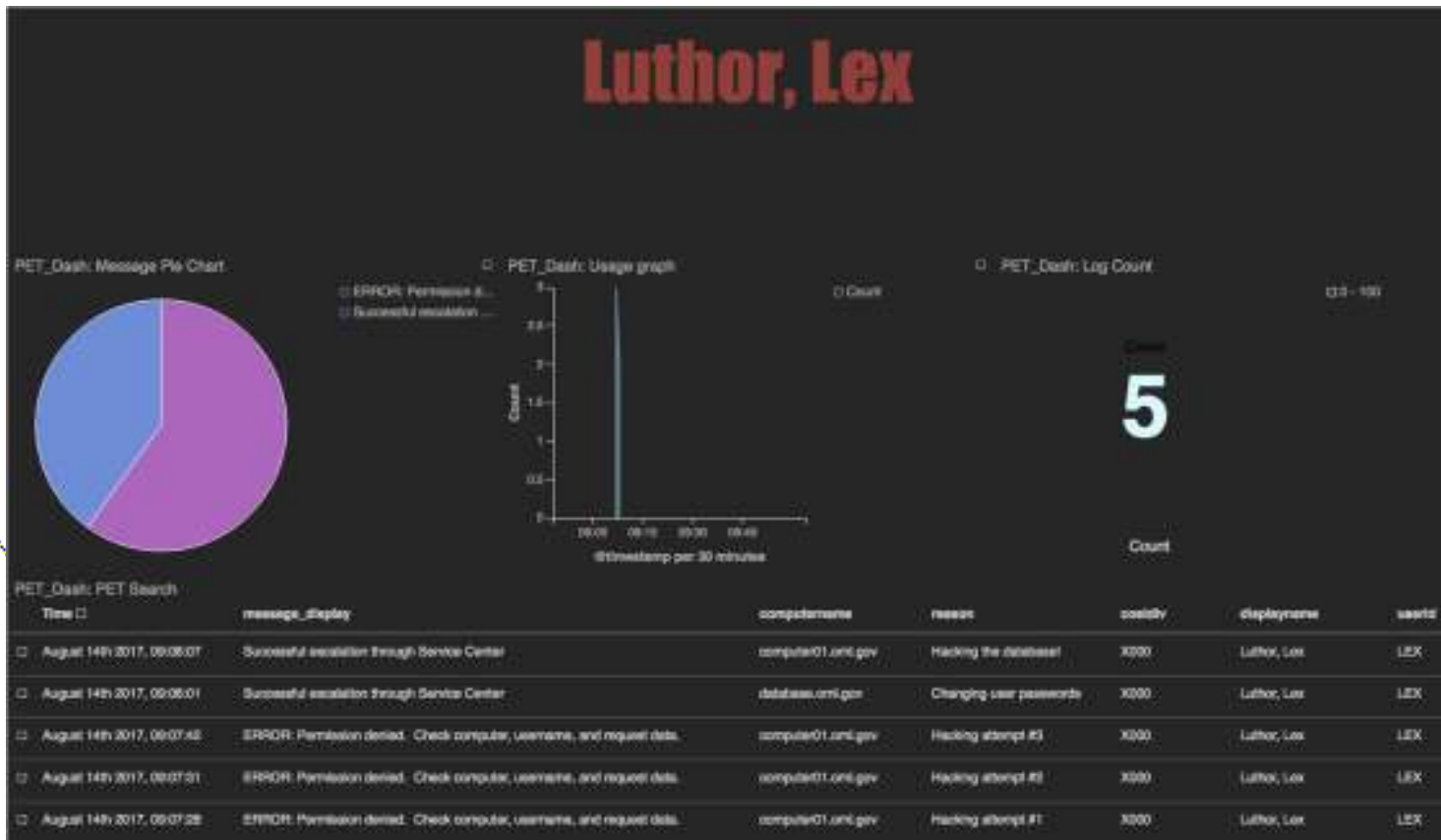
- 6 Elasticsearch nodes (Docker) across 6 physical servers
- 2 Billion documents/ ~1.5TB/Day
- Maintaining ~ 30 days worth of data, ~21 billion documents



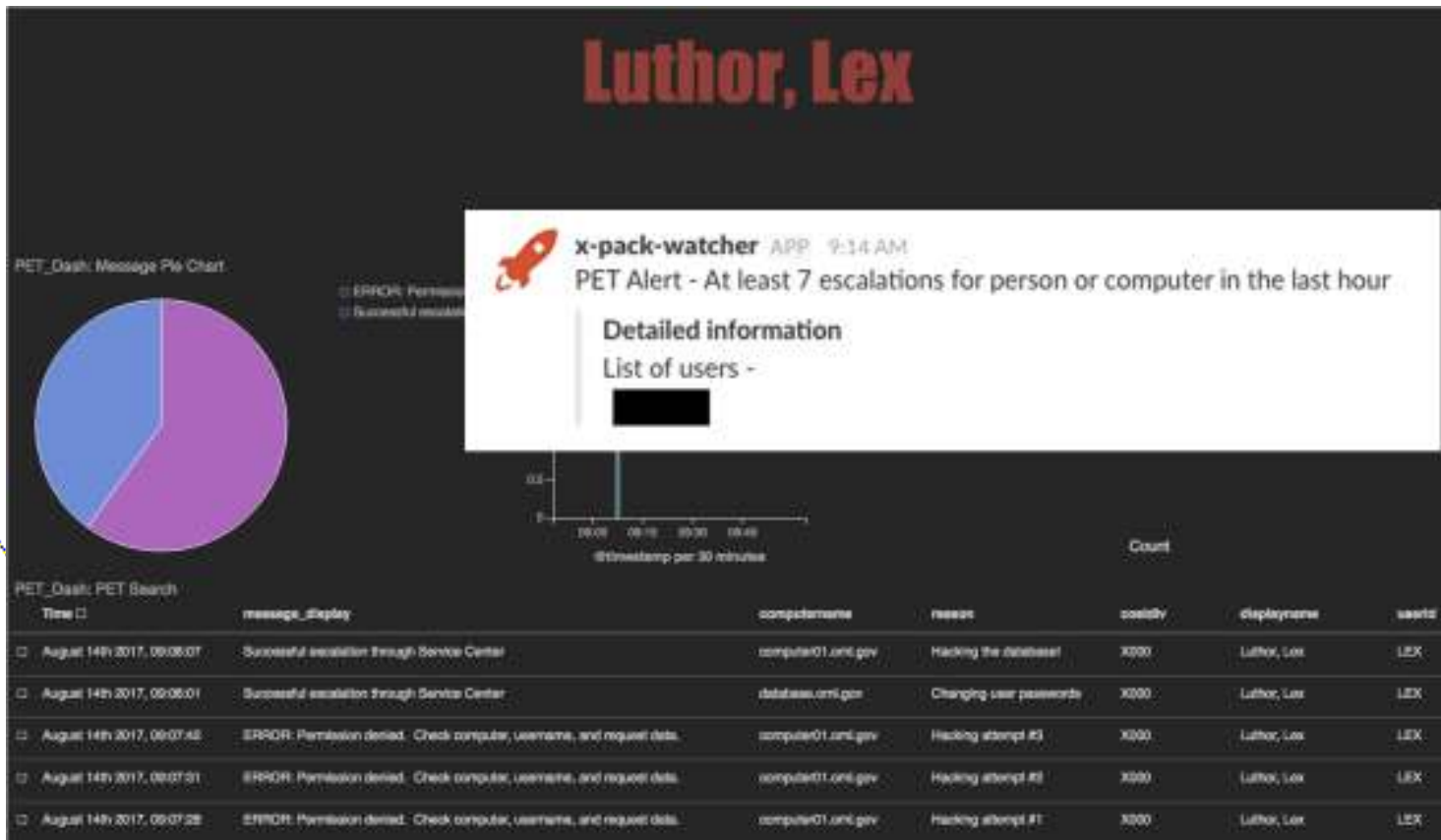
The screenshot shows a Kibana dashboard with three main panels. The left panel displays cluster health metrics: 'Name' (kibana), 'Status' (green), 'Uptime' (1d 10h), and 'Size' (4). The middle panel shows node details for 'Node 0', including 'IP Address' (192.168.1.10), 'Host Name' (192.168.1.10), and 'Type' (master). The right panel shows node details for 'Node 101', including 'IP Address' (192.168.1.101), 'Host Name' (192.168.1.101), and 'Type' (data).

Cluster Health	Node 0	Node 101
Name: kibana	IP Address: 192.168.1.10	IP Address: 192.168.1.101
Status: green	Host Name: 192.168.1.10	Host Name: 192.168.1.101
Uptime: 1d 10h	Type: master	Type: data
Size: 4		

Kibana Dashboards



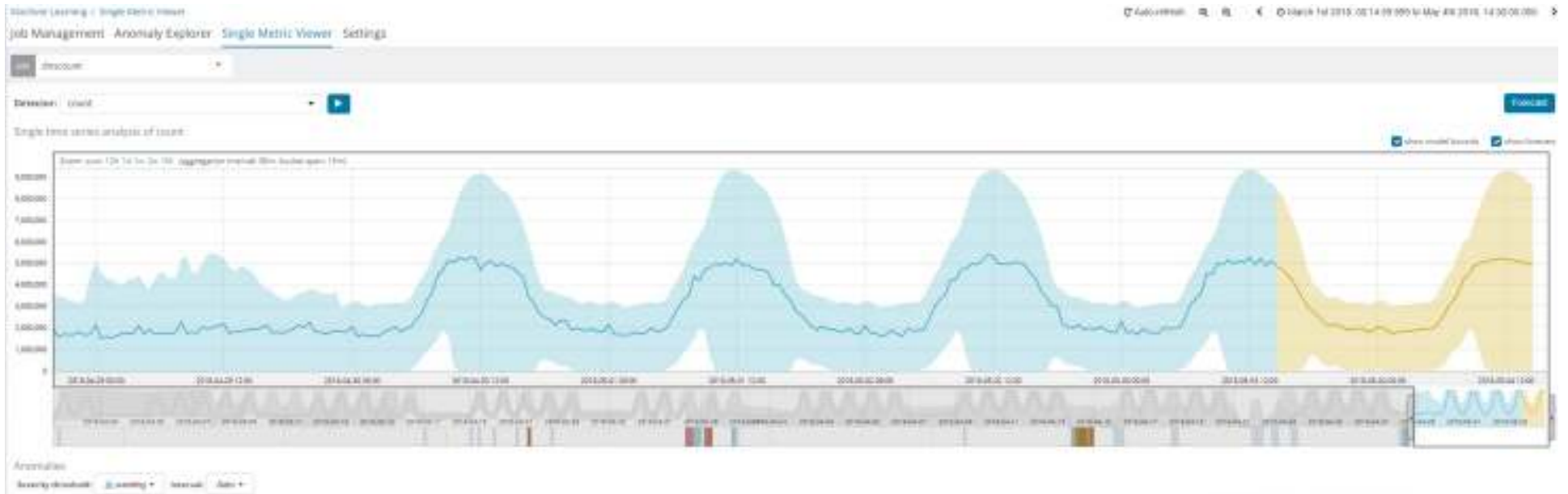
Kibana Dashboards



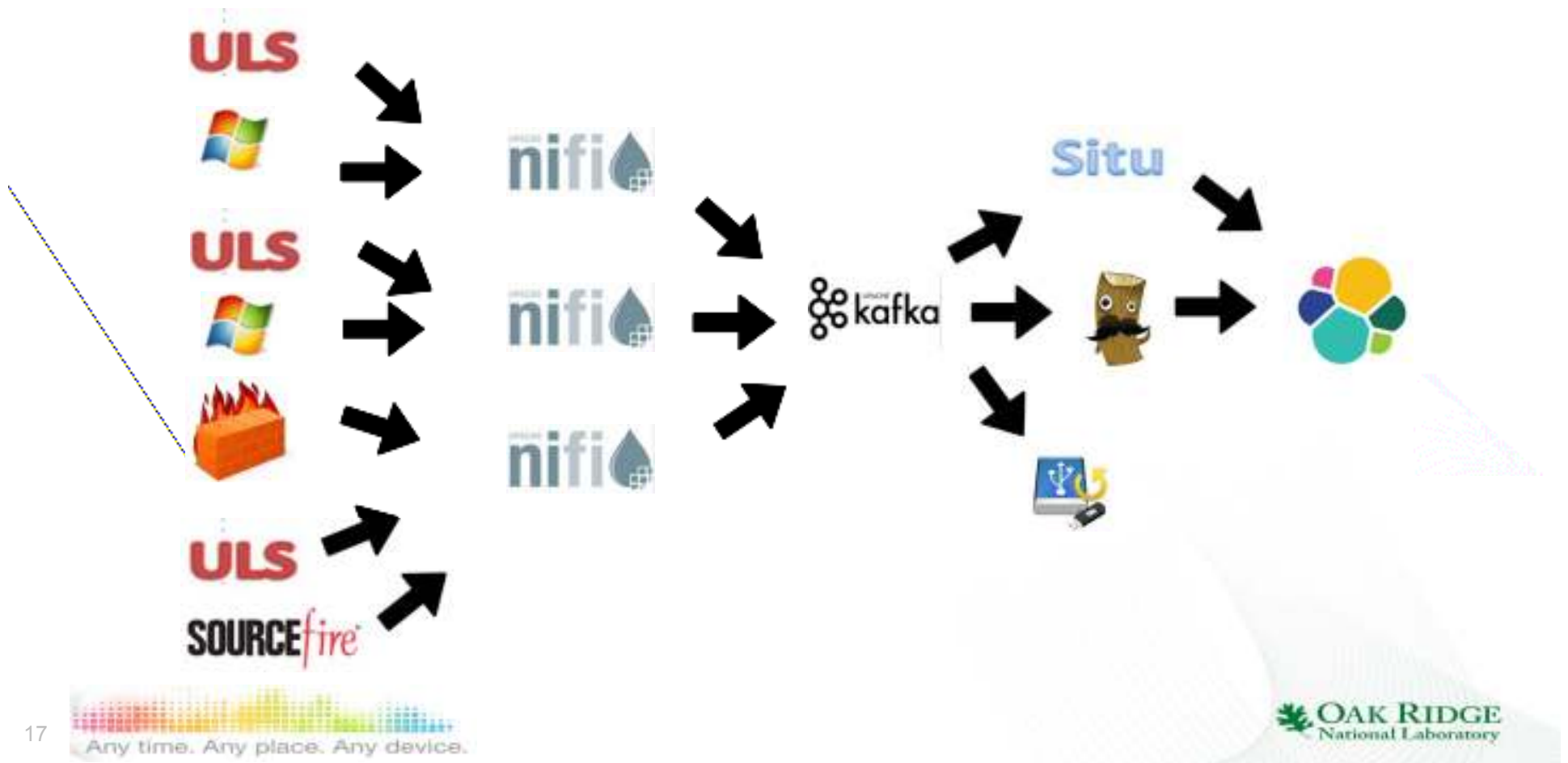
Timelion



Machine Learning



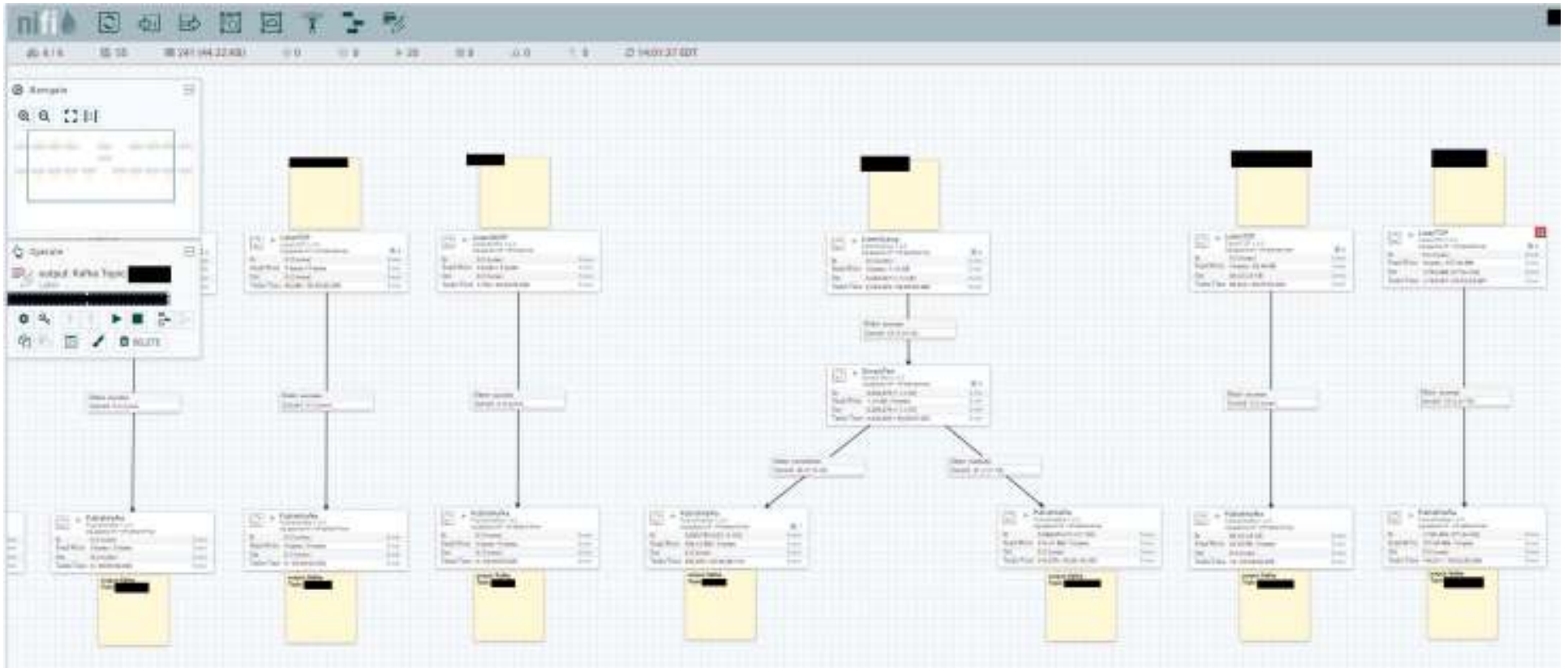
Improved Data Streaming Pipeline



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

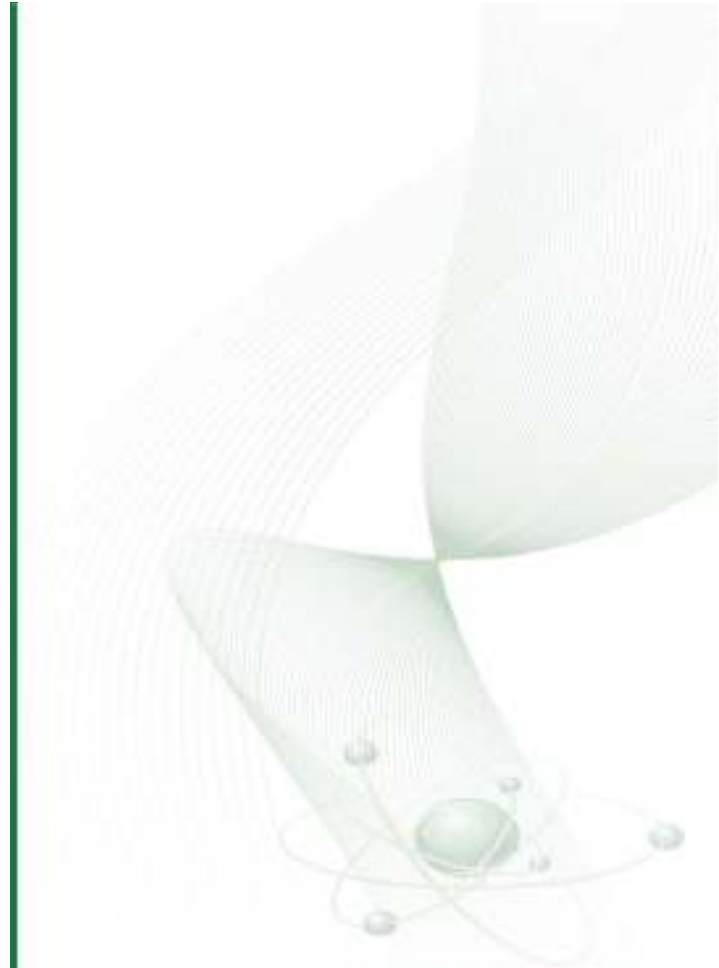


Kafka

- 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



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

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

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

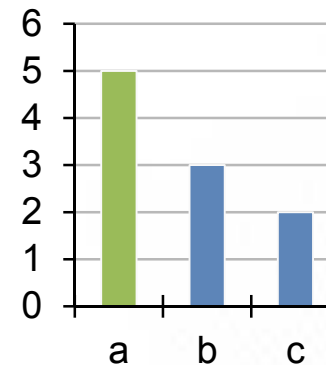
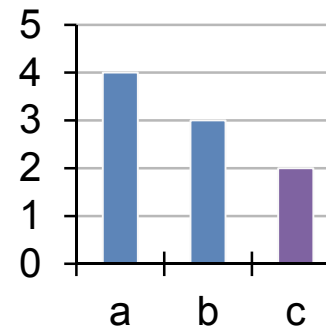
score: 0.237

new event: a

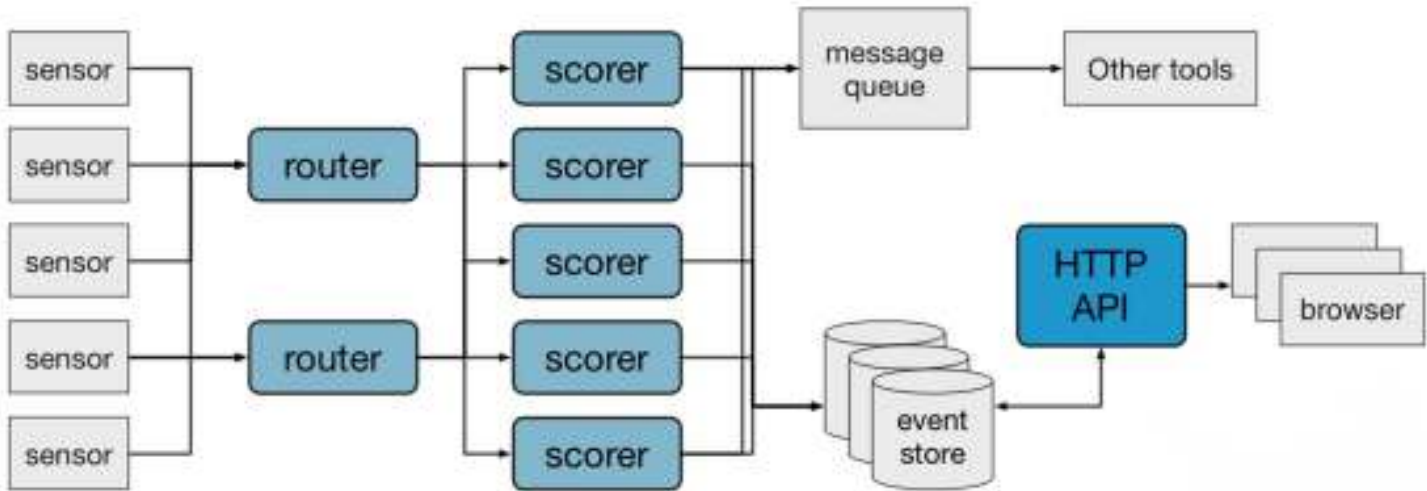
more likely events: []

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

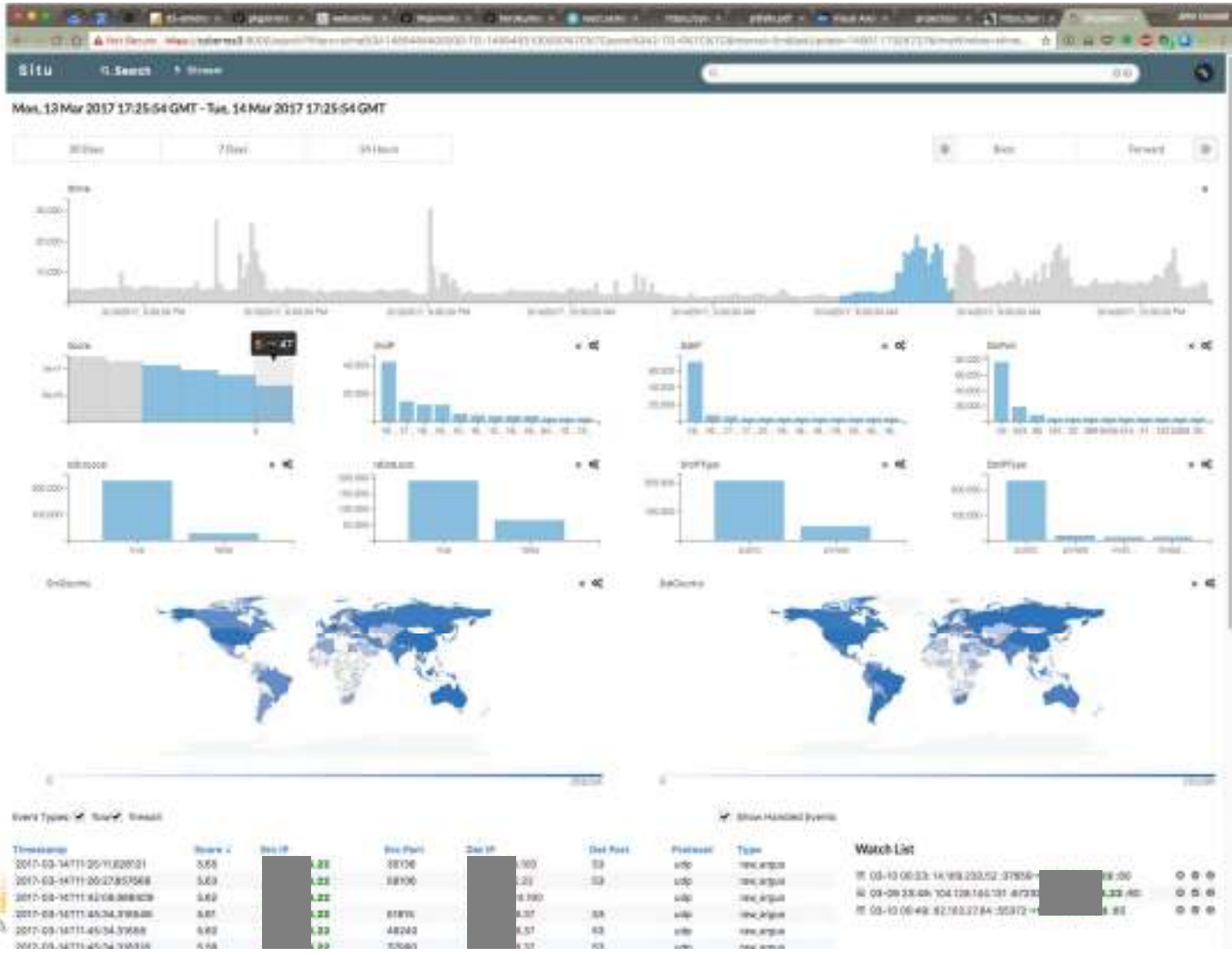
“ Situ is the best tool for detecting and analyzing beaconing traffic known to humankind.

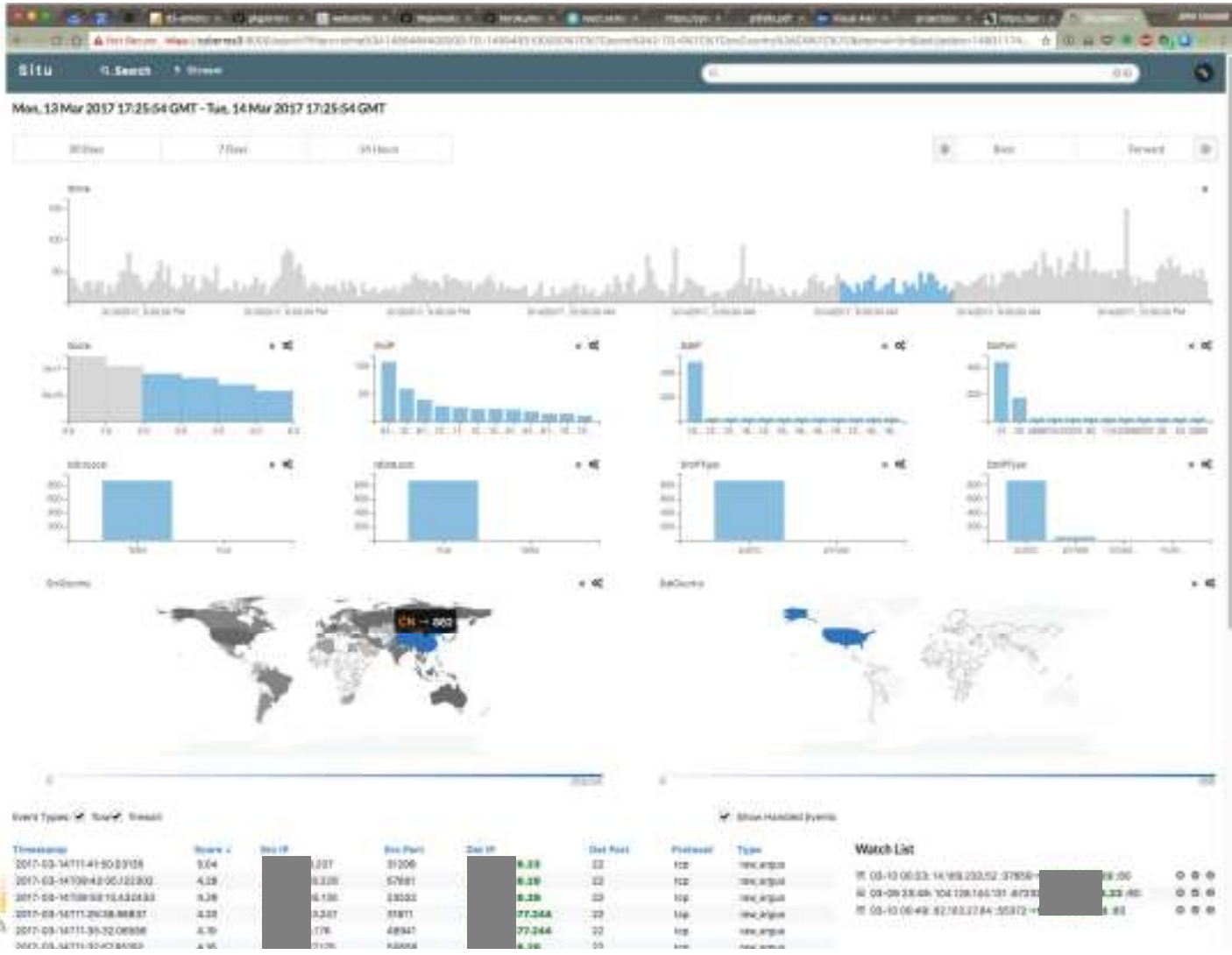
- SOC Senior Engineer, ORNL

“ 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







Event Types: Row Detail

Watch Header Events

Threadname	Rows	Size	Size Part	Date	Date Part	Protocol	Type
2017-03-14711415033105	3,04	36,193,182,281	3,206	19	22	tcp	irc.argus
2017-03-147004200322300	4,18	122,190,193,228	2,761	19	22	tcp	irc.argus
2017-03-1470038131430433	4,28	123,190,226,126	2,833	19	22	tcp	irc.argus
2017-03-14711263818837	3,28	127,238,182,287	3,181	244	22	tcp	irc.argus
2017-03-14711303206606	4,70	123,95,176,179	2,891	244	22	tcp	irc.argus
2017-03-14711324299992	4,16	123,190,227,175	3,004	19	22	tcp	irc.argus
2017-03-147113332196783	4,08	112,238,314	3,623	244	22	tcp	irc.argus
2017-03-147020033099476	4,04	113,123,258,189	3,627	19	22	tcp	irc.argus
2017-03-147020202007884	4,02	22,184,89,87	4,270	19	22	tcp	irc.argus
2017-03-147020202007884	4,02	175,189,184,244	3,895	19	22	tcp	irc.argus
2017-03-14711324807480	4,02	114,222,192,73	3,723	19	22	tcp	irc.argus
2017-03-1471132291806052	3,68	114,231,194	3,995	22	22	tcp	irc.argus
2017-03-147020202007884	3,88	98,226,261,238	3,819	4	30	tcp	irc.argus
2017-03-147020202007884	3,88	81,147,89,18	3,730	4	7080	tcp	irc.argus
2017-03-147004840301803	3,78	81,147,89,18	3,143	4	14890	tcp	irc.argus
2017-03-147004840301803	3,78	81,147,89,18	3,154	4	33400	tcp	irc.argus
2017-03-147004840301803	3,78	81,147,89,18	3,169	4	33500	tcp	irc.argus
2017-03-147004840301803	3,78	81,147,89,18	3,153	4	3600	tcp	irc.argus
2017-03-147004840301803	3,78	81,147,89,18	3,171	4	36765	tcp	irc.argus
2017-03-147004840301803	3,78	81,147,89,18	3,187	4	3787	tcp	irc.argus
2017-03-147004840301803	3,78	81,147,89,18	3,190	4	12760	tcp	irc.argus
2017-03-147004840301803	3,77	81,147,89,18	3,231	4	3637	tcp	irc.argus
2017-03-147004840301803	3,77	81,147,89,18	3,192	4	36816	tcp	irc.argus
2017-03-147004840301803	3,77	81,147,89,18	3,186	4	38318	tcp	irc.argus
2017-03-147004840301803	3,77	81,147,89,18	3,203	4	28309	tcp	irc.argus

Page 1 of 38

Previous Next

Appliance

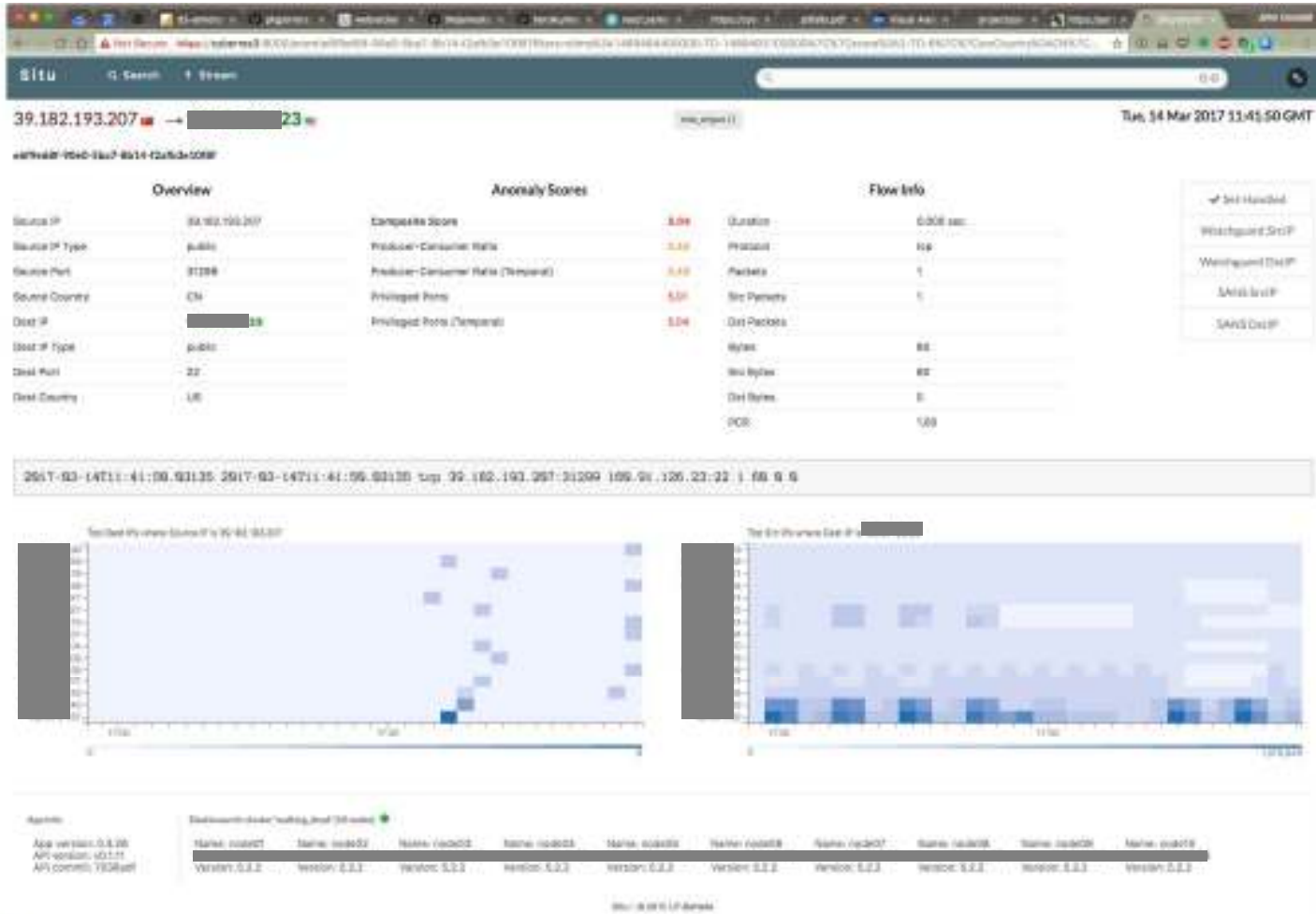
Appliance version: 0.3.20
API version: 40.1.11
API address: 7028aaf

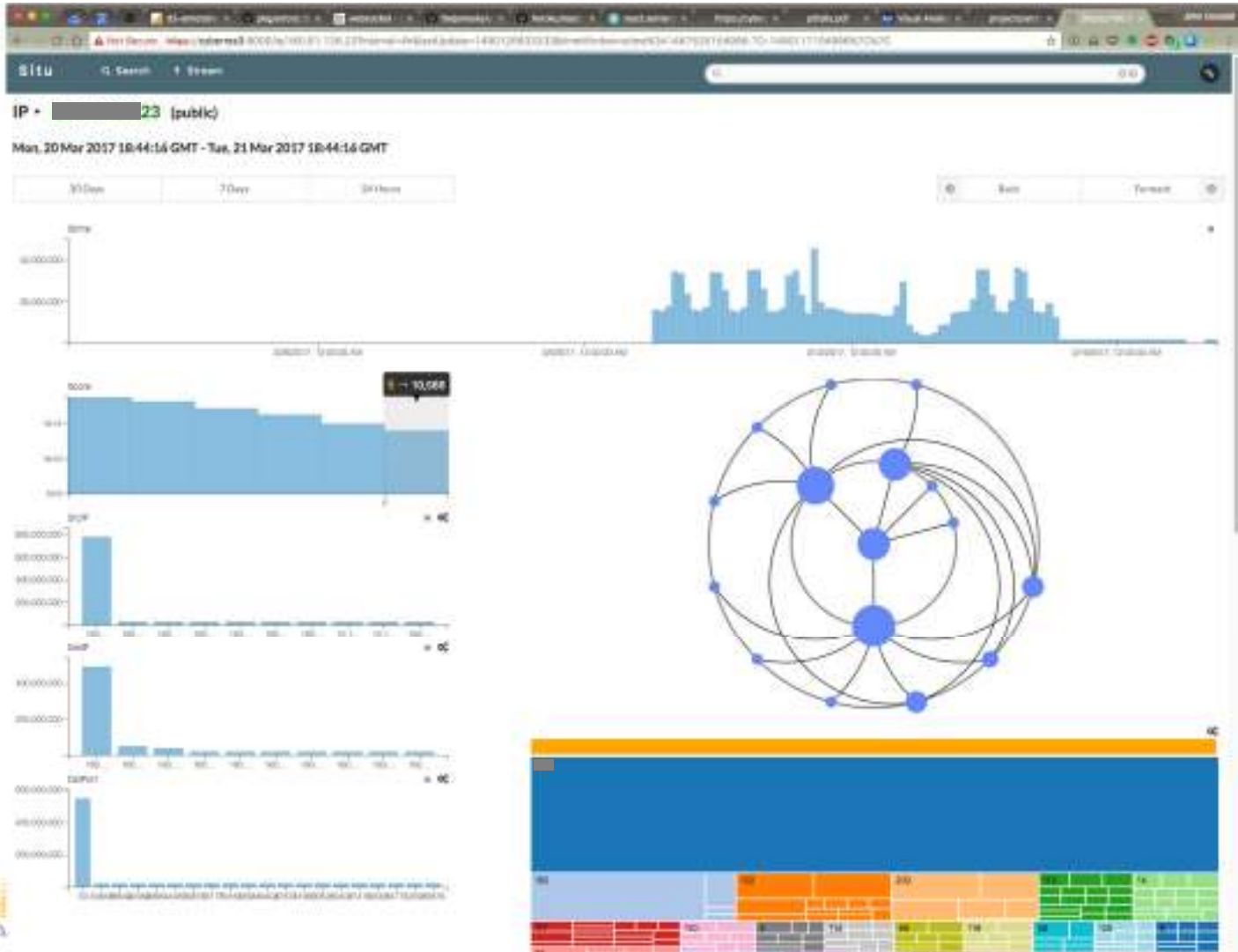
Backend status: healthy (OK) [OK]

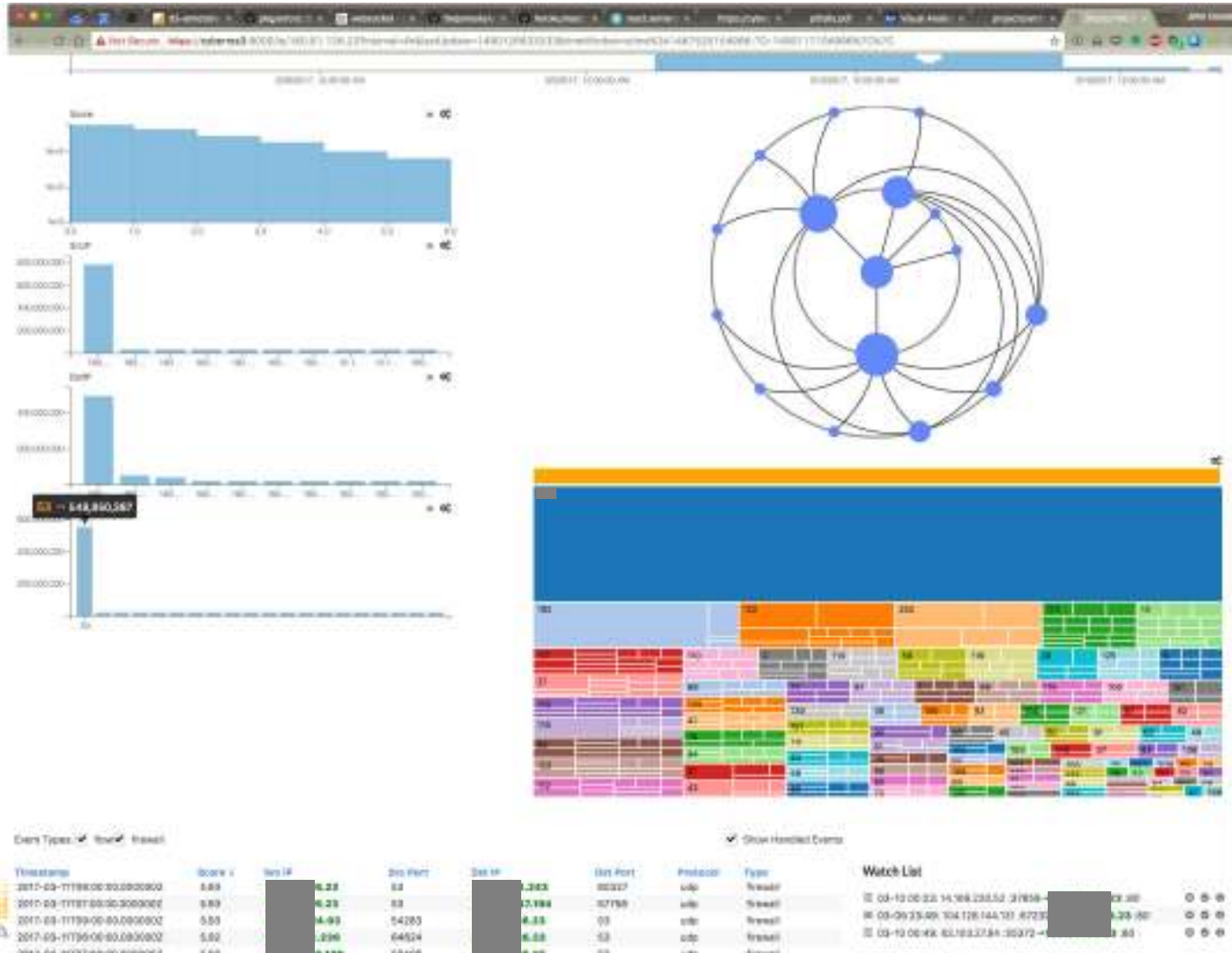
Name: node01	Name: node02	Name: node03	Name: node04	Name: node05	Name: node06	Name: node07	Name: node08	Name: node09	Name: node10
Version: 0.2.2	Version: 0.2.2	Version: 0.2.2	Version: 0.2.2	Version: 0.2.2	Version: 0.2.2	Version: 0.2.2	Version: 0.2.2	Version: 0.2.2	Version: 0.2.2

Be: 0.20% of 64000









Event Types: New Reset Show Handled Events

Timestamp	Score	Src IP	Src Port	Dest IP	Dest Port	Protocol	Type	Watch List
2017-03-07T18:28:03.807387	5.00	132.160.160.118	34754	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-07T18:28:04.488932	5.00	80.181.254.67	84302	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-07T18:49:42.335829	5.00	217.86.88.137	86551	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-08T08:27:18.136002	5.07	85.84.78.38	41204	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-08T13:21:10.28624	5.10	0.143.28.141	9852	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-09T16:17:19.81281	5.11	113.190.268.36	38184	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-09T16:18:38.871888	5.12	113.171.8.233	87688	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-10T01:48:43.921895	5.11	88.181.244.4	30441	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-10T09:47:48.575267	5.11	122.91.182.48	10249	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-10T10:19:58.206957	5.11	88-89.8.274	38281	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-10T08:38:08.647381	5.11	112.17.22.191	6068	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-10T10:32:36.212247	5.10	128.84.87	28822	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-12T09:24:03.478664	5.08	93.778.232	52287	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-12T09:49:50.181821	5.08	92.169.118.19	38122	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-12T07:34:16.706587	5.07	17.20.183.81	7163	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-12T18:03:58.473816	5.08	79.82.150.222	7402	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-13T11:57:52.068829	5.08	89.176.216.8	44278	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-13T10:37:36.872389	5.08	163.12.78.177	37111	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-14T18:30:23.807216	5.01	17.84.184.782	28864	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-14T10:32:30.185918	5.02	217.84.128.28	37037	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-15T07:30:31.814718	5.03	88.11.61.188	3781	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-15T18:03:20.826711	5.07	88.91.73.168	4478	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-16T16:31:28.482876	5.12	112.198.80.117	47816	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-16T12:48:37.888832	5.10	176.229.80.71	6788	[REDACTED]	22	tcp	new.alert	[REDACTED]
2017-03-16T18:34:09.037864	5.10	96.180.228.178	28060	[REDACTED]	22	tcp	new.alert	[REDACTED]

Page 1 of 88

Previous Next

Agents

Backend - status: 'waiting client' (20 nodes)

Name: node01	Name: node02	Name: node03	Name: node04	Name: node05	Name: node06	Name: node07	Name: node08	Name: node09	Name: node10
Version: 5.2.2	Version: 5.2.2	Version: 5.2.2	Version: 5.2.2	Version: 5.2.2	Version: 5.2.2	Version: 5.2.2	Version: 5.2.2	Version: 5.2.2	Version: 5.2.2

64 / 6 2017-07-06 09:00



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

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?

