A Data Scientist's Swiss Army Knife

A Cyber Defense System to Detect Low Footprint Campaigns in Multi-petabyte Network Data

Lee Blum
Big Data Architect
Verint Systems
About Me

Lee Blum

@theLeeBlum

Lee.Blum@Verint.com
What We’re Up Against

Fancy Bear: German attack "by Russians"

The Equifax Hack Has the Hallmarks of State-Sponsored Pros

Investigations into the massive breach aren't complete, but the intruders used techniques that have been linked to nation-state hackers in the past.
Cyber Crash Course!
Cyber Defense - Sensors

Define Goals & Target
Gather Intelligence
Weaponize & Infiltrate
Call Home
Move Laterally
Exfiltration & Damage

Internet
Firewall
Logs
C&C
WWW
DMZ
AV
Mail
FTP
Honeypot
Firewall
Private LAN
NF
DBs
EP
Client EP
Directory
RADIUS
LM
Logs
Firewall

Confidential and proprietary information of Verint Systems Inc. © All rights reserved worldwide
Internet Service Provider

Organization 1

Organization 2

... Organization N

Alerts, Metadata, NetFlows

Internet

Cyber Defense System
Meet Our Personas!

Mark
CSOC Operator (Tier 1)

Rajesh
Cyber Analyst (Tier 2)

Amy
Data Scientist (Tier 3)

Violet
DevOps

Linda
Director

Seconds
Minutes
Hours
Unified
Low Footprint
# How Much Data?

## For 100Gbps network with 100k Endpoints

<table>
<thead>
<tr>
<th>Structured</th>
<th>Unstructured</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Alerts</td>
<td>+ Logs</td>
</tr>
<tr>
<td></td>
<td>+ Alert Files &amp; Raw</td>
</tr>
<tr>
<td>C&amp;C</td>
<td>1 billion</td>
</tr>
<tr>
<td>AV</td>
<td>1 billion</td>
</tr>
<tr>
<td>EP</td>
<td>8 billion</td>
</tr>
<tr>
<td>LM</td>
<td>1 billion</td>
</tr>
<tr>
<td>NetFlows</td>
<td>4 trillion</td>
</tr>
</tbody>
</table>

- **650k Events p/s | 90 days | 5 Trillion events | 4 PB**
Our Solution
High Level Architecture

Applications
- Enrichments & Correlations
- Large Scale Repository
- Entity Analytics
  DNS, IP, Cookie, etc.
- Graph Analytics
- Rule Engine
- Encrypted Event Analytics
  SparkML

Research
Data Exploration

Tools
- Logs
- Statistics
- Monitoring

Devops
- Installation
- Cluster Management
- Patching
- Upgrade
- InfoSec

Infrastructre

Alerts – C&C, AV, EP, LM – MD and Files
Network Forensics – NetFlows, Raw NW Logs – Text (Semi-structured)
Entity Analytics – Timely Aggregation

**Streaming**
15 minute aggregations

- 00:00 - 00:15
- 00:15 - 00:30
- ...

**Batch**
Hierarchical Time Based Aggregations

- Daily, Day 1 of month
- Day 2
- ...
- Weekly, Sunday
- Monday
- ...
- Monthly, January
- ...

**Time**
# Entity Analytics – Knowledge Layers

<table>
<thead>
<tr>
<th>Entity</th>
<th>Host</th>
<th>URL</th>
<th>IP</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregation</td>
<td>Hourly</td>
<td>Morning, Evening</td>
<td>Daily &amp; Day in Week</td>
<td>Monthly &amp; Month in Year</td>
</tr>
<tr>
<td>Patterns</td>
<td>Daily Peak</td>
<td>Traffic Profile</td>
<td>Destination Geo Location</td>
<td>Accessed EP by time of day</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Employee Endpoint</td>
<td>News website</td>
<td>C&amp;C Server</td>
<td>SysAdmin</td>
</tr>
<tr>
<td>Anomalies</td>
<td>Surf out of hours</td>
<td>Irregular Protocol</td>
<td>Irregular UL/DL</td>
<td>Irregular login to EP</td>
</tr>
</tbody>
</table>
Entity Analytics - Big Data Pipeline

Alerts, MD

Stream Processing: 15 min agg.

Orchestrator

Persistency (Entities)

Batch Processing (Data Scientists)

Query (Thrift Server)

Batch Processing: Long agg.

Result Set Exploration
The Data Scientist’s Swiss Army Knife

Train Models, Test Detections

Amy
Data Scientist
(Tier 3)

Spark
SQL / ML

Entities
HTTP / SSL
Connections

Alerts
NetFlows

Logs
Raw Network
Packets
The new model is integrated in the system to generate new alerts.

A group of suspicious alerts forms an incident.

Incident analysis may suggest a deeper pattern.

Research of stored MD and Raw generates new model for detection on Streaming / Batch job.
Summary

+ Transformation from Collection to Analytics

+ Democratization of Big Data
  - Rapid R&D
  - Single platform
  - High performance
  - Elasticity

+ Rapid Research to Production Cycle
Thank you @theLeeBlum