Who am I?
Why Apache Metron?
Months until breach noticed vs Avg. months log retention vs Months missing

9 vs 6 vs 3
Time until breach actually noticed

- **Police One/Berkut**: 28 Months
- **Yahoo/?**: 35 Months
- **FB/Cambridge Analytica**: 48 Months
“Sometime in the next few years we're going to have our first category-one cyber-incident; one that will need a national response.”

Ian Levy
Technical Director
National Cyber Security Centre
Andhra Pradesh Police, India
Aristotle University of Thessaloniki, Greece
Automobile Dacia, Romania
Cambrian College, Canada
Chinese public security bureau
CJ CGV
Dalian Maritime University
Deutsche Bahn
Dharmais Hospital, Indonesia
Faculty Hospital, Nitra, Slovakia
FedEx
Garena Blade and Soul
Guilin University Of Aerospace Technology
Guilin University Of Electronic Technology
Harapan Kita
Hospital[disambiguation needed], Indonesia
Hezhou University
Hitachi
Honda
Instituto Nacional de Salud, Colombia
Lakeridge Health
LAKS
LATAM Airlines Group
MegaFon
Ministry of Internal Affairs of the Russian Federation
Ministry of Foreign Affairs (Romania)
National Health Service (England)
NHS Scotland
Nissan Motor Manufacturing UK
O2, Germany
Petrobrás
PetroChina
Portugal Telecom
Pulse FM
Q-Park
Renault
Russian Railways
Sandvik
São Paulo Court of Justice
Saudi Telecom Company
Sberbank
Shandong University
State Governments of India
Government of Gujarat
Government of Kerala
Government of Maharashtra
Government of West Bengal
Suzhou Vehicle Administration
Sun Yat-sen University, China
Telefónica
Telenor Hungary, Hungary
Telkom (South Africa)
Timrå Municipality, Sweden
Universitas Jember, Indonesia
University of Milano-Bicocca, Italy
University of Montreal, Canada
Vivo, Brazil
2018 so far...

**EXACTIS**

340M Records

**Under Armour**

150M Records

**MyHeritage**

92M Records

And many, many, many more..

What Does Apache Metron Look Like?
Security telemetry source: authentication logs

```
cat data.txt
1,C625@$DOM1,U147@$DOM1,C625,C625,Negotiate,Batch,LogOn,Success
1,C653@$DOM1,SYSTEM@C653,C653,C653,Negotiate,Service,LogOn,Success
1,C660@$DOM1,SYSTEM@C660,C660,C660,Negotiate,Service,LogOn,Success
```
<table>
<thead>
<tr>
<th>Score</th>
<th>timestamp</th>
<th>source/type</th>
<th>alert_status</th>
<th>ip_dst_host</th>
<th>ip_src_host</th>
<th>threat_reason</th>
<th>user</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2017-05-25 03:56:55</td>
<td>auth</td>
<td>NEW</td>
<td>C467</td>
<td>C506</td>
<td>The distinct lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 02:56:56</td>
<td>auth</td>
<td>NEW</td>
<td>C612</td>
<td>C965</td>
<td>The distinct lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 03:57:00</td>
<td>auth</td>
<td>NEW</td>
<td>C467</td>
<td>C506</td>
<td>The distinct lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 04:03:15</td>
<td>auth</td>
<td>NEW</td>
<td>C612</td>
<td>C965</td>
<td>The distinct lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 04:08:09</td>
<td>auth</td>
<td>NEW</td>
<td>C467</td>
<td>C506</td>
<td>The distinct lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 03:56:55</td>
<td>auth</td>
<td>NEW</td>
<td>C528</td>
<td>C477</td>
<td>The distinct lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 03:57:00</td>
<td>auth</td>
<td>NEW</td>
<td>C528</td>
<td>C477</td>
<td>The distinct lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 03:57:36</td>
<td>auth</td>
<td>NEW</td>
<td>C611</td>
<td>C611</td>
<td>The distinct lan (1.00)</td>
<td>U66</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 03:59:30</td>
<td>auth</td>
<td>NEW</td>
<td>C528</td>
<td>C477</td>
<td>The distinct lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 04:08:24</td>
<td>auth</td>
<td>NEW</td>
<td>C612</td>
<td>C1143</td>
<td>The distinct lan (1.00)</td>
<td>U534</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 03:57:13</td>
<td>auth</td>
<td>NEW</td>
<td>C528</td>
<td>C477</td>
<td>The distinct lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 03:59:30</td>
<td>auth</td>
<td>NEW</td>
<td>C528</td>
<td>C477</td>
<td>The distinct lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 04:04:15</td>
<td>auth</td>
<td>NEW</td>
<td>C612</td>
<td>C1191</td>
<td>The distinct lan (1.00)</td>
<td>U534</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 03:59:23</td>
<td>auth</td>
<td>NEW</td>
<td>C523</td>
<td>C1968</td>
<td>The distinct lan (1.00)</td>
<td>U534</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 03:59:23</td>
<td>auth</td>
<td>NEW</td>
<td>C523</td>
<td>C1968</td>
<td>The distinct lan (1.00)</td>
<td>U534</td>
</tr>
<tr>
<td>10</td>
<td>2017-05-25 03:59:45</td>
<td>auth</td>
<td>NEW</td>
<td>C467</td>
<td>C10123</td>
<td>The distinct lan (1.00)</td>
<td>C10123</td>
</tr>
<tr>
<td>Score</td>
<td>timestamp</td>
<td>source</td>
<td>alert_status</td>
<td>ip_dist_host</td>
<td>ip_src_host</td>
<td>threat</td>
<td>reason</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>--------</td>
<td>--------------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25</td>
<td>auth</td>
<td>NEW</td>
<td>C467</td>
<td>C506</td>
<td>The dist...lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25</td>
<td>auth</td>
<td>NEW</td>
<td>C612</td>
<td>C965</td>
<td>The dist...lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25</td>
<td>auth</td>
<td>NEW</td>
<td>C467</td>
<td>C506</td>
<td>The dist...lan (1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25</td>
<td>auth</td>
<td>NEW</td>
<td>C612</td>
<td>C965</td>
<td>The dist...lan (1.00)</td>
<td>U22</td>
</tr>
</tbody>
</table>

The distinct number of machines that user U22 attempted to login to (2) is more than 6 standard deviations (0.23) from the median (1.00).
<table>
<thead>
<tr>
<th>Score</th>
<th>timestamp</th>
<th>source:type</th>
<th>threat:...</th>
<th>reason</th>
<th>user</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>2017-05-25 03:56:55</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 03:56:56</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 03:57:00</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 03:57:13</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 03:57:36</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 03:57:39</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:03:15</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:08:24</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:09:09</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:09:30</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:09:45</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:30:15</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:30:45</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:31:15</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:31:45</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:32:15</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:32:45</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:33:15</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
<tr>
<td>100</td>
<td>2017-05-25 04:33:45</td>
<td>auth</td>
<td>The distinct...</td>
<td>(1.00)</td>
<td>U22</td>
</tr>
</tbody>
</table>

The distinct number of machines that user U22 attempted to login to (2) is more than 5 standard deviations (0.29) from the median (1.00).
What is Apache Metron?
Built on top on proven open source big data technology
An architecture for real-time cybersecurity analytics
Telemetry Data Source
Telemetry Data Collectors
Cyber Security Stream Processing Pipeline
Profiling by time

Wide range of algorithms including:

- HyperLogLogPlus
- Bloom filters
- T-digests
- Statistical Baselining
- Hashing functions
- Outlier detection
- GeoHashing over time
- Locality Sensitive Hashing
Cyber Security Stream Processing Pipeline
Apache Metron Modules
Who is Using Apache Metron (Part 1)
KPN neemt QSight IT over

Lees het hele nieuwsbericht
The Wider Apache Metron Ecosystem
PSSC Labs and CyberSecurity Malaysia Team Up to Crunch Data and Crush Hackers

#HPC #CyberSecurity #supercomputer #bigdata #PSSCLabs

prweb.com/releases/2018/ ...
Who is Using Apache Metron (Part 2)
Deploying Apache Metron
Phase 0 – Current State

Security Assets

ADP Event Broker (Kafka)
ADP Smart Connectors
ADP Logger
ArcSight ESM

HDF
NiFi - Ingest

ADP Event
Broker (Kafka)
ADP Smart
Connectors
ADP Logger
ArcSight ESM

HDF
HDP
HDFS

AD/AssetDB/HR/Threat
Phase 1 - Ingest and Archive
Phase 2 – Enrich and Threat Intel

Security Assets

AD/AssetDB/HR/Threat

ArcSight ESM

ADP Event Broker (Kafka)

ADP Smart Connectors

ADP Logger

HDF

NiFi - Ingest

Kafka MQ

Storm Parse / Enrich / GeoIP / Index

Enrichment Data

HDF

Solr

Investigator UI

Spark

Historical Analysis

HDP

HDFS

Zeppelin

Banana / Kibana / ZoomData

Security Assets

AD/AssetDB/HR/Threat

ArcSight ESM

ADP Event Broker (Kafka)

ADP Smart Connectors

ADP Logger

HDF

NiFi - Ingest

Kafka MQ

Storm Parse / Enrich / GeoIP / Index

Enrichment Data

HDF

Solr

Investigator UI

Spark

Historical Analysis

HDP

HDFS

Zeppelin

Banana / Kibana / ZoomData
Phase 3 – NiFi Data Ingestion + Analytics / UEBA Profiling

Security Assets

AD/AssetDB/HR/Threat

ADP Event Broker (Kafka)

ADP Smart Connectors

ADP Logger

ArcSight ESM

HDF

NiFi - Ingest

Kafka MQ

Storm Parse / Enrich / GeoIP / Index

Enrichment Data

Source Data (via NiFi)

Investigator UI

Solr

HDP

Historical Analysis

Spark

Metron Profiler

Alert

Triage

15

HDF5

Zeppelin

Banana

16

15

ADP Logger
Phase 4 – ArcSight Logger Migration + New Data Sources

Security Assets

AD/AssetDB/HR/Threat

HDF

Source Data (via NiFi)

HDF

Storm Parse / Enrich / GeoIP / Index

Kafka MQ

NiFi - Ingest

Enrichment Data

HDP

HDP

Spark

Investigator UI

Historical Analysis

Model as a Service

Alert

Triage

Metron Profiler

Automated Response

Zeppelin

ADP Event Broker (Kafka)

ADP Smart Connectors

ADP Logger

HDF

Netflow / PCAP / Snort (Kafka direct)

17

18

19

20

Source Data (via NiFi)

Banana

© Hortonworks Inc. 2011–2018. All rights reserved
Questions?
Appendix
Considerations for Sizing Apache Metron
Sizing an HCP deployment

• Events per second (average and peak)
• Retention time for Hot / Warm / Cold zones
• Enrichments
• Node sizing
• I/O Considerations
• PCAP?
3 Months

- Fast indexed layer (Solr / ES) ~3 months
- Warm HDFS layer ~3 months
12 Months

- Fast indexed layer (Solr / ES) ~3 months
- Warm HDFS layer ~12 months
24 Months

- Fast indexed layer (Solr / ES) ~3 months
- Warm HDFS layer ~12 months
- Cold HDFS layer +12 months
Beyond 24 months

- Fast indexed layer (Solr / ES) ~3 months
- Warm HDFS layer ~12 months
- Cold HDFS layer +12 months
Apache Metron Data Sheet

- **Ingest:**
  - Apache NiFi: syslog, socket, file, web services, SQL, RDBMS, Windows Event Log, FTP, MQ, JMS, Splunk and others
  - High-performance DPDK Packet Capture

- **Parsers:**
  - Cisco ASA
  - Bluecoat
  - Fireeye
  - Palo Alto
  - SourceFire
  - WebSphere
  - Snort IDS
  - Bro DPI
  - Netflow, IPFIX
  - Grok (Custom)

- **Java (Custom)**
- **JSON**
- **CEF, LEEF (ArcSight, Qradar compat.)**
- **Applications: DHCPD, AD**

- **Enrichments and threat feeds:**
  - Geo
  - Whois
  - HBase
  - JDBC
  - Stellar
  - CSV
  - Stix, Taxii threat intel feeds

- **Analytics features:**
  - Profiler and statistical baselining engine
  - Model Services for advanced ML

- **Threat Triage rules and scoring engine**

- **Indexing and search:**
  - Elasticsearch, Kibana
  - Solr
  - HDFS
  - Kafka

- **Data science features:**
  - Spark Machine Learning
  - Zeppelin notebooks and reporting
  - Wide partner eco-system

- **Forensic features:**
  - PCAP inspector
  - PCAP query
  - Long term data store