Viewing the nodes in the noise: Leveraging Data Science to Discover Persistent Threats

…Sharing Threats with US Based Commercial Critical Infrastructure

D. L. Evenden
Federal Cyber Security Division
720-578-5365
David.Evenden@centurylink.com
Agenda

• Intelligence Collection & Redistribution
• Lessons Learned
• Automating the Work
  – QuarterMaster
    • Exposure
    • DGA
    • Patter
• Questions
1. Interested parties get cleared
2. Intel comes in from federal intel communities
3. DHS collects and redistributes intel to interested && cleared parties
4. Cleared parties identify vehicles to leverage intel in commercial spaces
Intelligence Collection & Redistribution

- Intel Communities
  - Homeland Security
    - CSP: Verizon
    - CSP: CenturyLink
      - Local Gov
      - Energy
      - Finance
      - DIB
Intelligence Collection & Redistribution

Standard Configuration
Lessons Learned

What works, What hasn’t
Mature Organizations

Saying “Because it’s Bad” doesn’t work
Non-Mature Organizations

Identifying discovered threats

For non-mature organizations advanced methods are required to aid in the endeavors of identifying the exact source of discovered & blocked threats.

- **Advanced HoneyPot Tactics**
  - Issuing fake files to retrieve (HII) host identifiable information

- **WebRTC**
  - Leveraging existing JS code to retrieve (HII) host identifiable information

- **Netflow Traffic**
  - The integration and collection of netflow traffic allows for the identification of the source
Findings

• If you are providing IOC’s to clients, maintain an Intel database and send them enriched data

• If you are a client that receives IOCs, maintain your own intel database.
Sourcing your DB

- https://attack.mitre.org/wiki/Groups
- https://www.fireeye.com/current-threats/apt-groups.html
- https://apt.securelist.com/#!/threats/
- https://github.com/CyberMonitor/APT_CyberCriminal_Campaign_Collections
  - https://github.com/kbandla/APTnotes
  - https://github.com/fdiskyou/threat-INTel
Automating the hard work

Advanced analytics - Cultivating a better Cyber World
Intelligence Collection & Redistribution

2018 Q3 ECS BLOCK REPORT

Q3 Hits by TA: All ECS indicators are tied to unique TAs (represented here as TA numbers that obfuscate the group or nation-state responsible for malicious activity). Some TAs conduct targeted campaigns against specific groups while others direct general threats at various targets.

Selected ECS Activity Examples:

ECS alerts allowed a customer to locate malware. The CSP then correlated the traffic against other commercial sensors monitoring that network and ran triage on infected servers.

ECS validated hits on a domain which the NCCIC states is a callback address associated with Remote Access Tool (RAT) activity. NCCIC shared recommended follow-up actions.
Lifecycle

- DNS
- Email

Active Blocking
- In Email
- Out DNS

Exposed Attributed
- Heuristics

Feature Extraction

Additional Analysis
- Passive Algorithmic Analysis
Algorithms

Algorithms designed to identify malicious DNS Names

• DGA Algorithm
  • Identifies Dynamically Generated Domain Names

• Exposure Algorithm
  • Uses Exposed DNS Attributes to identify potentially malicious domains

• Pattern Algorithm
  • Uses callback intervals to identify potentially malicious domains
Operationalizing Models

- **DGA models:**
  - Model only needs to be trained once
  - Only uses domain name for classifying

- **Exposure Model:**
  - Need historical data to train the model
  - Needs to be re-trained every few months
  - Classifying domains requires historical features

- Training utilizes batch-processing using a week’s worth of data

- Scoring requires semi-real time stream processing with different codebase
Leveraging Attributes

This allows us to identify new C2 domains in an unclassified way: using the unclassified attributes of classified indicators sourced from the Intel Community.
DGA Algorithm: Lifecycle

- Attacker Uploads Agent
- Agent uses stored key
- Agent pulls time from public NTP
- ALGO server uses same stored key
- ALGO pulls time from public NTP
- ALGO & Agent calculate Domain
- ALGO Registers Domain/IP mapping
- Agent pulls IP for Domain

DGA Lifecycle

Public NTP

DGA ALGO server

C2 Server

Victim

Attacker
DGA Detection Models

- Analyzing statistical features (randomly-generated)
  - 60680ad5728991c31277cd43f733903d]\net
  - mns34m1qifzti4q7h9qlpik[.]com
  - rxjthjm1pofte[.]com

- Pre-defined word lists (Dictionary-DGA)
  - Using NLP techniques, determine when words “don’t make sense” together:
    - hypophyseal-relativity[.]com
    - machinelike-hypocellularity[.]com
    - imperative-carborundum[.]com

- words and random characters (Hybrid-DGA)
  - yawqthdpanxious[.]download
  - msxvkcijreactivity[.]download
Exposure Model

• Using these features and probabilities from DGA models, we train a model using whitelisted and blacklisted domains:

• WhiteListed:
  • TrendMicro
  • Symantec
  • Verizon

• Blacklisted:
  • bambenekconsulting.com/feeds/
  • OTX
  • AIS
Exposure Model

~30 Attributes

TTL
Registration
IP Variations
Responses
Daily Trends
Cost per Query
Pattern Algo

Scope Bandwidth

Randomized Variable

Callback Interval

15 Minute

~ (+/-) 15 Minutes

7200 Seconds

7200 Seconds ~ (+/-) 15 Minutes
Viewing the Noise

Jeremy said he asked.. just waiting on response/approval.
Viewing the Noise

Jeremy said he asked.. just waiting on response/approval.
Viewing the Noise

configured intervals become apparent

Remove outliers
Data Organization & Alerting
Data Organization

Extended Configuration
CDP, CUSUM, and Pattern Change

Enclave 1

US-Cert/DHS Indicators

Ingest

Store metadata in Database

SQL Query: delete from dns.dnstraffic where customer=’Xs’ and dnsname=’Xs’ and timestamp=’Xs’

Distribute metadata based on customer

Alerts

SQL Query: select customer, dnsname, timestamp from dns.dnstraffic
Modeling & ELK Tagging Configuration

- You can easily use Python to train your models, process your data, and assign relevant tags in your Elastic dataset.
Data Tagging

Visualize / DGA Predictions

Search... (e.g. status:200 AND extension:PHP)

Add a filter

enrich*

Metrics

- Metric Count

Add metrics

Buckets

- Split Rows domain.keyword: Descending
- Split Rows score.keyword: Descending
- Split Rows label.keyword: Descending

Add sub-buckets

<table>
<thead>
<tr>
<th>domain.keyword</th>
<th>score.keyword</th>
<th>label.keyword</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>dfw28s02-in-f3.1e100.net</td>
<td>0.99</td>
<td>dga</td>
<td>2</td>
</tr>
<tr>
<td>dfw28s02-in-f3.1e100.net</td>
<td>0.08</td>
<td>legit</td>
<td>1</td>
</tr>
<tr>
<td>msnbot-64-4-54-18.search.msn.com</td>
<td>0.23</td>
<td>legit</td>
<td>2</td>
</tr>
<tr>
<td>msnbot-64-4-54-18.search.msn.com</td>
<td>0.0</td>
<td>legit</td>
<td>1</td>
</tr>
<tr>
<td>n25.netmark.pl</td>
<td>0.07</td>
<td>legit</td>
<td>2</td>
</tr>
<tr>
<td>n25.netmark.pl</td>
<td>0.0</td>
<td>legit</td>
<td>1</td>
</tr>
<tr>
<td>200-165-92-67.user.veloxzone.com.br</td>
<td>0.0</td>
<td>legit</td>
<td>1</td>
</tr>
<tr>
<td>200-165-92-67.user.veloxzone.com.br</td>
<td>1.0</td>
<td>dga</td>
<td>1</td>
</tr>
<tr>
<td>208.185.50.80.IPYX-063360-004-ZYO.zip.zayo.com</td>
<td>0.01</td>
<td>legit</td>
<td>1</td>
</tr>
<tr>
<td>208.185.50.80.IPYX-063360-004-ZYO.zip.zayo.com</td>
<td>1.0</td>
<td>dga</td>
<td>1</td>
</tr>
</tbody>
</table>
Active Blocking from Passive Threat Intelligence
Modeling & ELK Tagging Configuration

• You then generate alerts from your ELK platform that allow you to respond to identified threats.

• Once new threats are passively identified, add them to your active blocking platform
Thank you

Questions