Choose Your Battles
How To Fight The Right Wars

Eyal Paz, Security Researcher

Check Point
SOFTWARE TECHNOLOGIES LTD
whoami

• Security Researcher at Check Point
• B.Sc. in Software Engineering, studying towards M.Sc. in Computer Science
• Information Security lecturer
• Father
Agenda

• Research Motivation & Goals
• Under The Hood - Algorithmic Overview
  ▪ Aggregating events to incidents
  ▪ Differentiating incidents on host
  ▪ In-house TI feed
  ▪ Threat context
Motivation
Staying a Step Ahead of Threats

Make every effort to **PREVENT** attacks

Detection is not enough. The only way to avoid the cost of an attack is to prevent it altogether.

**DETECT** and **CONTAIN** attacks as soon as possible

Once infected, the cost of the attack just keeps on rising.

Effectively **RESPOND** and **REMEDiate**

Address the real business impact
Make sure the intrusion doesn’t come back.
Timing is Everything

Source: 2015 cost of data breach study: global analysis, Ponemon Institute

The longer an attack goes **UNDETECTED**, the more time it takes to **CONTAIN** it.

The longer it takes to **CONTAIN** it, the more it will **COST**.
Loud Infection → Fast Response

~70% of the infected machines are remediated within a week.
Silent Infection → Slow Response

~60% of the infected machines remediation takes more than a month.
# Loud vs. Silent: What is More Severe?

<table>
<thead>
<tr>
<th></th>
<th>Loud</th>
<th>Silent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Median Response Time</strong></td>
<td>1-3 days</td>
<td>More than 28 days</td>
</tr>
<tr>
<td><strong>Attack Vector</strong></td>
<td>Mostly Phishing &amp; Exploit kits</td>
<td>Mostly Phishing &amp; Exploit kits</td>
</tr>
<tr>
<td><strong>Attack Type</strong></td>
<td>Data corruption, Denial of Service, Ransom demanding</td>
<td>Espionage, Banking credentials, Data breach</td>
</tr>
<tr>
<td><strong>Discovery</strong></td>
<td>Easy</td>
<td>Hard</td>
</tr>
<tr>
<td><strong>Damage</strong></td>
<td>Temporal</td>
<td>Continual</td>
</tr>
</tbody>
</table>
Reasons For Slow Response

• Internal bureaucracy and politics
  Different teams with different agendas need to collaborate

• Network configuration issues
  Difficult or impossible to track the infected host

• Understaffed security teams
  “62% of organizations are receiving more alerts than they can feasibly investigate”

Source: 2015 Incident Detection & Response Survey, RAPID7
Threat Context

• Given one or more hosts access a “Malicious site”

• What should the security team do with such information?

• How should it be prioritized vs. other alerts?
Research Questions & Directions

• How to choose your battles
  Aggregate & summarize multiple alerts to a reasonable number of incidents to decrease workload

• How to fight the right war
  Adding a context layer to incidents to better prioritize their urgency
Algorithmic Overview
Aggregating Events to Incidents

- Discover similarity between compromised hosts
- Reduce overhead of security incidents
- Assist in prioritization & remediation
  One script to clean them all
Step 1 – Pre-processing

• Get all alerts from all available sensors’ events:
  ▪ FW & IDS
  ▪ End Point
  ▪ Domain Controller
  ▪ Proxy & DNS Servers
Step 2 – Feature Vector

• Create a list of all unique IoC
  ▪ Domains
  ▪ Destination IP for non HTTP/DNS addresses
  ▪ Destination port
  ▪ And any other forensics telemetry type you can get

• Not all features are equally weighted features
Step 3 – Host Matrix

• Create a matrix where the rows are for hosts and the columns are for the features

• Example:

  3 hosts – A, B, C
  4 IoCs – evil-1.com, evil-2.com, 1.2.3.4, TCP/6667
  Domain weight is 1, IP weight is 1.3, Port weight is 1.6

<table>
<thead>
<tr>
<th></th>
<th>evil-1.com</th>
<th>evil-2.com</th>
<th>1.2.3.4</th>
<th>TCP/6667</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host A</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Host B</td>
<td>0</td>
<td>0</td>
<td>1.3</td>
<td>1.6</td>
</tr>
<tr>
<td>Host C</td>
<td>0</td>
<td>1</td>
<td>1.3</td>
<td>1.6</td>
</tr>
</tbody>
</table>
Cosine Similarity

• a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them – number in range [0,1]

\[
similarity = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}
\]

```python
def cosine_similarity(x,y):
    numerator = sum(a*b for a,b in zip(x,y))
    denominator = square_rooted(x)*square_rooted(y)
    return round(numerator/float(denominator),3)
```

Step 4 – Similarity Matrix

• Create the Cosine Similarity matrix when we are comparing every 2 hosts’

• In the below example:
  - **Green** is for strong matches
  - **Yellow** is for weak matches
  - **Red** is for non-matches

<table>
<thead>
<tr>
<th></th>
<th>Host A</th>
<th>Host B</th>
<th>Host C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host A</td>
<td>1</td>
<td>0</td>
<td>0.3</td>
</tr>
<tr>
<td>Host B</td>
<td>-</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>Host C</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>
Step 5 – Noise Reduction

• Mask out weak matches for noise reduction

<table>
<thead>
<tr>
<th></th>
<th>Host A</th>
<th>Host B</th>
<th>Host C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host A</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Host B</td>
<td>-</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>Host C</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>
Step 6 - Extract Incidents

- Create a graph using the similarity matrix as a graph adjacency matrix

- Find the graph connected components which comprise the security incident that we looked for:
  
  \{\text{Host A}\}, \{\text{Host B, Host C}\}
## PoC at Customer sites (24 Hours)

<table>
<thead>
<tr>
<th></th>
<th>Organization A</th>
<th>Organization B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique Indicators</td>
<td>177</td>
<td>41</td>
</tr>
<tr>
<td>Compromised Hosts</td>
<td>29</td>
<td>19</td>
</tr>
<tr>
<td>Security Incidents</td>
<td>11 (-62%)</td>
<td>9 (-52%)</td>
</tr>
</tbody>
</table>

**Illustration**

[Diagram of network analysis]
## PoC at Customer sites (24 Hours)

<table>
<thead>
<tr>
<th></th>
<th>Organization C</th>
<th>Organization D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique Indicators</td>
<td>42</td>
<td>90</td>
</tr>
<tr>
<td>Compromised Hosts</td>
<td>35</td>
<td>44</td>
</tr>
<tr>
<td>Security Incidents</td>
<td>13 (-62%)</td>
<td>16 (-63%)</td>
</tr>
</tbody>
</table>

**Illustration**
Model Limitation

• This model has a hidden assumption that all indicators that are found on a given host are related
• We all know that’s not always the case
Differentiate Incidents

• To differentiate the incident we need to break it down to its components – indicators
• Define similarity between indicators
• Consider recurring occurrences of the same indicators on different hosts
URL Similarity

- Equal non-zero amount of dashes
- Equal non-zero amount of digits
- Digits/Dash are on the same index
- Subdomains under same domain
- Same exact registrant
- Same anonymized registrant service
- Different anonymized registrant
- Small domain/registrant edit distance
- Same exact domain name
- Same domain name length
- Same IP resolutions amount
- Both domains had never had IP allocated
- Shared ASN
- Shared IP addresses
- Same TLD which is not .com and not local
- Close registration date
- Close first detected date
- Close language ratio
- Shared URL path exactly
- Similar URL path
CryptoWall C2 Servers

• Are the URLs below related?
  ▪ abelindia.com/1LaXd8.php
  ▪ purposenowacademy.com/5_YQDI.php
  ▪ mycampusjuice.com/z9r0qh.php
  ▪ theGinGod.com/HS0ILJ.php
  ▪ yahoosupportaustralia.com/8gX7hN.php
  ▪ successafter60.com/iCqjno.php
  ▪ alltimefacts.com/EiFSId.php

• Other than the funny URL path pattern
  ▪ All the above URLs were first seen on 04-Nov-2015 which indicate they belong to the same campaign
Emotet Malware DGA

• Are the domains below similar?
  ▪ myjfqirgagnpboou.eu
  ▪ kgpaorkwqlgrfcre.eu
  ▪ pqxhpvpumynikjh.eu
  ▪ iddxbogywitoaddv.eu
  ▪ clgarxlbvxcraqht.eu
  ▪ ...

• Other than the simple pattern [a-z]{16}\.eu
  ▪ All domains had never had an IP allocated
  ▪ All domains were never registered
  ▪ Close linguistic ratio
  ▪ Same TLD which is not .com and not local
# Virus Total URL - Emotet DGA

<table>
<thead>
<tr>
<th>Domain</th>
<th>VT URL Detection*</th>
</tr>
</thead>
<tbody>
<tr>
<td>pqxhqvumylnikjh.eu</td>
<td>0/67</td>
</tr>
<tr>
<td>iddxbogywitoaddv.eu</td>
<td>0/67</td>
</tr>
<tr>
<td>idlueqkbkkclcdj.eu</td>
<td>0/67</td>
</tr>
<tr>
<td>jjnstqfppyclvonk.eu</td>
<td>0/67</td>
</tr>
<tr>
<td>clgarxlbvxcraqht.eu</td>
<td>1/67</td>
</tr>
<tr>
<td>kgpaorkwqqlgrfcre.eu</td>
<td>1/66</td>
</tr>
</tbody>
</table>

* Scanned on May-2016
Can you spot the Phishy one?
Expedia Phishing Campaign

expediapartnercentral.it
Legal Department- Domain Administrator
domains@expedia.com

danito alex
alexissisi@libero.it
More Like This…

• Under the name of “danito alex” two more domains were registered on the same day
  ▪ accessoclienti-expedia.it
  ▪ accessoclienti-expedia.com

List of domain names registered by Danito Alex

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Create Date</th>
<th>Registrar</th>
</tr>
</thead>
<tbody>
<tr>
<td>partner-expedia.com</td>
<td>2016-04-27</td>
<td>pop.it</td>
</tr>
<tr>
<td>accessoclienti-expedia.com</td>
<td>2016-04-27</td>
<td>ascio.com</td>
</tr>
<tr>
<td>accessoclienti-expedia.it</td>
<td>2016-04-27</td>
<td></td>
</tr>
</tbody>
</table>

Source: http://domainbigdata.com/name/danito%20alex
VT URL - Expedia Phishing Campaign

<table>
<thead>
<tr>
<th>Domain</th>
<th>VT URL Detection*</th>
</tr>
</thead>
<tbody>
<tr>
<td>accessoclienti-expedia.com</td>
<td>0/67</td>
</tr>
<tr>
<td>accessoclienti-expedia.it</td>
<td>2/67</td>
</tr>
<tr>
<td>partner-expedia.com</td>
<td>7/67</td>
</tr>
</tbody>
</table>

* Scanned on June-2016
Step 1 – Pre-processing

- Get all IoC from all available sensors’ events:
  - FW & IDS
  - End Point
  - Domain Controller
  - Proxy & DNS Servers
Step 2 – Similarity Graph

G \leftarrow \text{Init-Graph()}

For each pair of IoC of same type, do:

\begin{align*}
G &. \text{Add-Node}(\text{IoC-A}) \\
G &. \text{Add-Node}(\text{IoC-B}) \\
\text{If } G &. \text{Has-Path}(\text{IoC-A}, \text{IoC-B}) = \text{False} \\
\text{AND } \text{IoC-A} & \text{ is similar to IoC-B, then:} \\
G &. \text{Add-Edge}(\text{IoC-A}, \text{IoC-B}) \\
\text{Incidents} & \leftarrow G.\text{Connected-Components()}
\end{align*}
Phishing Actor

• Are the domains below similar?
  • settings-yahoo.com
  • linkedin.net
  • antiviruspc-update.com
  • google-japan2010.com
  • yahoo-japan2010.com
  • facebook-support.org
Phishing Actor

- settings-yahoo.com ⇔ 1linkedin.net
  - Same anonymized registrant service provider

- settings-yahoo.com ⇔ antiviruspc-update.com
  - Shared IP addresses
  - Same anonymized registrant service provider
  - Equal non-zero amount of dashes
  - Same IP resolutions amount

- settings-yahoo.com ⇔ google-japan2010.com
  - Same anonymized registrant service provider
  - Equal non-zero amount of dashes
  - Both contain popular domain name

- settings-yahoo.com ⇔ yahoo-japan2010.com
  - Shared IP addresses
  - Same anonymized registrant service provider
  - Equal non-zero amount of dashes
  - Both contain same popular domain name

- settings-yahoo.com ⇔ facebook-support.org
  - Shared IP addresses
  - Same IP resolutions amount
  - Equal non-zero amount of dashes
  - Both contain popular domain name
Phishing Actor

- 1linkedin.net ⇔ antiviruspc-update.com
  - Same anonymized registrant service provider
  - Close registration date
  - Shared IP addresses

- 1linkedin.net ⇔ google-japan2010.com
  - Same anonymized registrant service provider
  - Shared IP addresses

- 1linkedin.net ⇔ facebook-support.org
  - Shared IP addresses
  - Close registration date
Phishing Actor

- antiviruspc-update.com ⇔ google-japan2010.com
  - Same anonymized registrant service provider
  - Equal non-zero amount of dashes
  - Shared IP addresses

- antiviruspc-update.com ⇔ facebook-support.org
  - Close registration date
  - Same IP resolutions amount
  - Equal non-zero amount of dashes
  - Shared IP addresses

The graph is connected; therefore, all the domains are related.
# Virus Total URL - Phishing Actor

<table>
<thead>
<tr>
<th>Domain</th>
<th>VT URL Detection*</th>
</tr>
</thead>
<tbody>
<tr>
<td>google-japan2010.com</td>
<td>0/67</td>
</tr>
<tr>
<td>yahoo-japan2010.com</td>
<td>0/67</td>
</tr>
<tr>
<td>facebook-support.org</td>
<td>1/66</td>
</tr>
<tr>
<td>linkedin.net</td>
<td>1/67</td>
</tr>
<tr>
<td>antiviruspc-update.com</td>
<td>2/67</td>
</tr>
<tr>
<td>settings-yahoo.com</td>
<td>5/67</td>
</tr>
</tbody>
</table>

* Scanned on May-2016
There’s Always Room For More BL

- There are many Threat Intelligence (TI) feeds out there
- The overlap between them is surprisingly low
- Putting all the vendors together still gives a partial coverage of the evilness on the internet

Source: Data Driven Threat Intelligence: Metrics on Indicator Dissemination and Sharing, MLSec/Niddel
IoC Similarity as a TI Feed

- The idea is to leverage existing feeds to create an in-house TI feed

Source: https://www.threatcrowd.org/
investigate-domain(domain)

If domain is suspicious, then:

   For each domain’s ip resolution, do:
       ip-investigation-queue.enqueue(ip)

   For each file downloaded/communicated with the domain:
       file-investigation-queue.enqueue(file)

   For each registrant owned the domain:
       registrant-investigation-queue.enqueue(registrant)
In-House Feed Value

- Feed relevancy is crucial
- High hit rate of harvested indicators comparing to common TI feeds
- Proactively get as many indicators as possible of the current actor attacking the network
Share TI For Your Own Interest

• Organizations on same geo/industry/size are likely to get the same kind of attacks

• Sharing indicators between them could be the key differentiator between DETECT vs. PREVENT

• Actively sharing communities should be everyone’s interest
Threat Context

• Adding more IoC is great
• But more alerts are pointless if they are without the proper threat context
# Domain Classification Analysis #1

**settings-yahoo.com**

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain Contained popular domain string (by Alexa)</td>
<td><img src="image" alt="settings-yahoo.com" /></td>
</tr>
<tr>
<td>Anonymized domain registrations (by who.is)</td>
<td>Registrant Email: <a href="mailto:whoisproxy@value-domain.com">whoisproxy@value-domain.com</a></td>
</tr>
<tr>
<td>Website going up and down (by PassiveTotal)</td>
<td><img src="image" alt="Website Traffic Chart" /></td>
</tr>
</tbody>
</table>

**Verdict:** Evidence implies a phishing /infecting website – Pre-Intrusion
# Domain Classification Analysis #2

**kgpaorkwqlgrfcre.eu**

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain is available for registration (by who.is)</td>
<td><img src="image" alt="Domain is available" /></td>
</tr>
<tr>
<td>Domain was never assigned to an IP (by PassiveTotal)</td>
<td><img src="image" alt="Domain Attributes" /></td>
</tr>
<tr>
<td>Domain was seen with which many like him within several minutes</td>
<td>myjfqrimgagnpboou.eu, pqxhpvumynikjh.eu, iddxbogywitoaddv.eu, clgarxbvxcraqht.eu, jnnstqfppyclvonk.eu, idlueqkbkfkclcdj.eu</td>
</tr>
</tbody>
</table>

**Verdict: Evidence implies a CnC server – Post Intrusion**
Alerts Prioritization

• Host resolving a phishing/infecting domain indicates an infection attempt
• Host resolving a CnC server domain indicates an on-going infection
Staying a Step Ahead of Threats

Events to Incidents → Faster Remediation

In-House TI Feed → Faster Intrusion Containment

Sharing TI → Moving From Detect To Prevent
Choose Your Battles
How To Fight The Right Wars

Eyal Paz, Security Researcher

Check Point
SOFTWARE TECHNOLOGIES LTD

Thank You!
References

• 2015 Incident Detection & Response Survey, RAPID7
• 2015 Cost of data breach study: global analysis, Ponemon Institute
• Data Driven Threat Intelligence: Metrics on Indicator Dissemination and Sharing, MLSec/Niddel
• Similarity measures in Python, dataaspirant.com
• Cosine similarity, Wikipedia
References cont.

- http://blogs.checkpoint.com/
- https://threatcrowd.org/
- https://passivetotal.org/
- https://virustotal.com/
- http://who.is/
- https://google.com/
- http://malwarefor.me/
- http://domainbigdata.com/