Assessing Targeted Attacks in Incident Response Threat Correlation

Jan 2017

www.lookingglasscyber.com

PRESENTER:

Allan Thomson, CTO
Dr Jamison Day, Principal Data Scientist
What... threats are targeting?

Who... is impacted by targeted threats?
Why automation is critical to success…

Security data is not intelligence. Intelligence is data that has been refined, analyzed or processed such that it is relevant, actionable and valuable.
Choosing Threat Intelligence Feeds

- Ensure **rich context**: Vulnerabilities, TTPs, Indicators, Actors
- Ensure **broad coverage**: Surface web, Dark web, Social media, Human & Automated
- Ensure **Timely**: Real-time is important; Hourly and frequent updates
Choosing Threat Correlation Telemetry - Flows

- Provides network session context
- Typically done as a non-inline correlation process to enable identification of behaviors and patterns over time
- Often uses automated techniques defined later in the presentation

**Recommendations**

- Should include both northbound and east-west traffic flows to detect external and cross-domain traffic behaviors
- **If possible include payload extraction and correlation across packets**
- IPFIX (Netflow v10) supports much context beyond traditional 5-tuple
- Gather unsampled flow rather than sampled flow especially if you are doing behavioral analysis
Choosing Threat Correlation Telemetry - Packets

- Provides ability to identify content in every packet that matches specific patterns
- Typically network inspection devices are programmed with rules to identify regex, signatures and payload that may be malicious

**Recommendations**
- Must focus on inline data rate inspection
- Ability to correlate at line rate
Assess Organizational Threat Posture

Identify Potential Compromised Assets

Understand the full context of communication between the compromised asset and internet

Workflow Supporting Correlation Steps: 1 of 2
Workflow Supporting Correlation Steps 2 of 2

Identify any data exfiltration or impact on compromised asset

Identify the spread of any threat within the perimeter
Threat Correlation in Your Cyber Security Ecosystem

New Threat Context

Known Threat Context

Network Activity

Asset Risk Factors

Vulnerability

Attacks

Anomalies

New Attacks

Analyst

Mitigate Action
Threat Correlation Approaches
Threat Correlation Approaches

**Threat Correlation**
Identifies new cyber threat insights by associating events from multiple data sources

**Statistical Correlation**
Measures the similarity in fluctuations between two variables.

### Approaches

- Manual Threat Correlation
- Field Comparison
- Rules-Based Matching
- Fuzzy Matching
- Machine Learning
Manual Threat Correlation

- Human comparison of data from multiple sources to identify threat-related events

- **Advantages**
  - Pattern Recognition
  - Language Abilities
  - Creative Thinking
  - Flexible Inference
  - Intuition/Guessing

- **Drawbacks**
  - Slow step-by-step instruction execution
  - Imprecise, Unpredictable, Reproducibility Issues
  - Bias/Prejudice
Real World Example: Data Processing Reduction

- **Per Asset Collection**
  - In a typical organization a single networked asset may initiate between 3 to 4 flows/second.
  - When averaged, this is 115,000 flows for a typical 8-hour work day.

- **All Assets Collection**
  - If the same organization has 1000 networked assets, then their aggregate flow count is ~115 million.

- **Internet Connect Correlation**
  - The amount of flows crossing the perimeter is highly dependent on cloud services and the business model of the organization.
  - If we assume that 30% of all traffic for an organization is traffic to the Internet, then this provides us with 35.5 million flows to consider for an 8-hour work day.

- **Threat Intelligence Correlation**
  - If we then assume 5% of these flows are connecting to Internet assets that have any Threat Intelligence associated with them, the number of flows is 1.8 million flows for a work day.

- **Threat Scoring Correlation**
  - Finally, if we consider out of that number how many Internet sites have a higher Threat Score than elevated score and assume 10% of the remaining flows require investigation, this would be 180K flows.
Field Comparison

Identical features seen in fields of different datasets

• **Advantages**
  – Simple to Implement & Update
  – Very Fast
  – Very Scalable

• **Drawbacks**
  – Naïve Approach
  – Misses Sophisticated Attacks

### IP Blacklist

<table>
<thead>
<tr>
<th>IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.1.1</td>
</tr>
<tr>
<td>1.2.4.6</td>
</tr>
<tr>
<td><strong>5.1.1.1</strong></td>
</tr>
<tr>
<td>1.3.5.7</td>
</tr>
</tbody>
</table>

### URL Blacklist

<table>
<thead>
<tr>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>u.a.com</td>
</tr>
<tr>
<td>v.b.org</td>
</tr>
<tr>
<td>y.e.com</td>
</tr>
<tr>
<td>z.f.com</td>
</tr>
</tbody>
</table>

### Netflow Activity

<table>
<thead>
<tr>
<th>IP</th>
<th>Port</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1.1</td>
<td>80</td>
<td><strong>w.a.com</strong></td>
</tr>
<tr>
<td>2.1.1.1</td>
<td>21</td>
<td>v.b.org</td>
</tr>
<tr>
<td>1.1.1.1</td>
<td>80</td>
<td><strong>w.a.com</strong></td>
</tr>
<tr>
<td><strong>3.1.1.1</strong></td>
<td>80</td>
<td><strong>w.c.com</strong></td>
</tr>
<tr>
<td>1.1.1.1</td>
<td>443</td>
<td><strong>w.a.com</strong></td>
</tr>
<tr>
<td>4.1.1.1</td>
<td>1025</td>
<td>x.d.edu</td>
</tr>
<tr>
<td>1.1.1.1</td>
<td>80</td>
<td><strong>w.a.com</strong></td>
</tr>
<tr>
<td><strong>5.1.1.1</strong></td>
<td>123</td>
<td>y.e.com</td>
</tr>
<tr>
<td>1.1.1.1</td>
<td>80</td>
<td><strong>w.a.com</strong></td>
</tr>
<tr>
<td>6.1.1.1</td>
<td>753</td>
<td>z.f.org</td>
</tr>
</tbody>
</table>
Rules-Based Matching

Specific features seen in combination across datasets

- **Advantages**
  - Identifies complex interactions
  - Scalable

- **Drawbacks**
  - Requires managing a large number of pre-defined rules
  - New threats require new rules

### Threat Intelligence Feed Records & Signatures

<table>
<thead>
<tr>
<th>IP</th>
<th>Port</th>
<th>Protocol</th>
<th>Regex</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1.1</td>
<td>53</td>
<td>UDP</td>
<td>^\w+@[a-zA-Z_]+.[a-zA-Z]{2,3}$</td>
</tr>
<tr>
<td>2.1.1.1</td>
<td>80</td>
<td>TCP</td>
<td>((\d{3}\ ?)</td>
</tr>
</tbody>
</table>

### Netflow Activity

<table>
<thead>
<tr>
<th>IP</th>
<th>Port</th>
<th>Protocol</th>
<th>Regex</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1.1</td>
<td>53</td>
<td>UDP</td>
<td><a href="mailto:bad@malware.net">bad@malware.net</a></td>
</tr>
<tr>
<td>2.1.1.1</td>
<td>80</td>
<td>TCP</td>
<td>(800) 800-1337</td>
</tr>
<tr>
<td>2.1.1.1</td>
<td>53</td>
<td>TCP</td>
<td><a href="mailto:really.bad@malware.net">really.bad@malware.net</a></td>
</tr>
</tbody>
</table>
Fuzzy Matching

Approximate features seen in combination across datasets

- **Advantages**
  - Helps identify new tactics in complex interactions
  - Captures issues with minor changes

- **Drawbacks**
  - Fuzzier → more false positives
  - Requires feedback for refinement
  - Computationally expensive
Machine Learning

Program computers to learn which dataset features are relevant

• **Advantages**
  – Identifies correlations humans haven’t yet made
  – Can learn new tactics

• **Drawbacks**
  – Slow(ish)
  – Some ML approaches are not very scalable
  – Does not help build intuition
  – Tough to tune false positives/negatives

Classification

Clustering

Neural Networks
How Can Hackers Evade Threat Correlation Detection?

<table>
<thead>
<tr>
<th>Threat Correlation Approach</th>
<th>Common Evasion Tactics</th>
<th>Level of Effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual Threat Correlation</td>
<td>• Increase amount of traffic to overwhelm humans</td>
<td>Low</td>
</tr>
<tr>
<td>Field Comparison</td>
<td>• Rotate use of unique identifiers (such as IP addresses &amp; domains)</td>
<td>Low</td>
</tr>
<tr>
<td>Rules-Based Matching</td>
<td>• Rotate use of unique identifiers</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>• Slight modifications to tools</td>
<td></td>
</tr>
<tr>
<td>Fuzzy Matching</td>
<td>• Rotate use of unique identifiers</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>• Significant modifications to tools</td>
<td></td>
</tr>
<tr>
<td>Machine Learning</td>
<td>• Rotate use of unique identifiers</td>
<td>Very High</td>
</tr>
<tr>
<td></td>
<td>• Significant modification to tools</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Continuously change tactics</td>
<td></td>
</tr>
</tbody>
</table>
Assessing Targeted Attacks

• Automating correlation of threat & network information can help your organization:
  
  – Identify active attacks
  
  – Assess attack severity
  
  – Prioritize response and mitigation activity
  
  – Identify important new threats & anomalies
Recommendations

**Determine** which threat intelligence feeds are best for your organization

**Integrate** threat intelligence into your automated threat management

**Capture & analyze** your network activity

**Automate** correlation of network activity with threat intelligence

**Maximize** impact with feedback loops within your threat management activities to continuously improve your organization’s abilities
Thank you
Web: www.lookingglasscyber.com
Twitter: @LG_Cyber