Automating Incident Investigations: Sit Back and Relax, Bots are Taking Over.....

Elvis Hovor
Mohamed El-Sharkawi

November 3, 2016
### Automating Incident Investigations: Sit Back and Relax, Bots are Taking Over

<table>
<thead>
<tr>
<th>Content</th>
<th>Slide #</th>
</tr>
</thead>
<tbody>
<tr>
<td>CURRENT CHALLENGES WITH INCIDENT INVESTIGATION</td>
<td>3</td>
</tr>
<tr>
<td>COMPREHENSIVE CYBER ONTOLOGY</td>
<td>13</td>
</tr>
<tr>
<td>HUMAN-LIKE DECISION MAKING</td>
<td>17</td>
</tr>
<tr>
<td>SAMPLE INVESTIGATION ENGINE</td>
<td>31</td>
</tr>
<tr>
<td>DEMO PLUS Q&amp;A</td>
<td></td>
</tr>
</tbody>
</table>
CURRENT CHALLENGES WITH INCIDENT INVESTIGATION
The Hard Facts – Get ready to be “Breached”

Suffering a breach is almost certainly inevitable

In 2015 Information Security Breaches survey show that 90% of large organizations and 74% of small organizations suffered a data breach in 2014.

Breaches are getting more expensive for organizations

The average total cost of a data breach for the participating companies increased 23 percent over the past two years to $3.79 million.

- Department for Business, Innovation and Skills (BIS)

- Ponemon Institute
The Problem: Incident Investigation & Response

Incident response is **time consuming, resource intensive, and a costly process**. Ineffective triaging allows early warnings of compromises to **slip through the cracks**.

<table>
<thead>
<tr>
<th>Operational Metric</th>
<th>Time</th>
<th>Seconds</th>
<th>Minutes</th>
<th>Hours</th>
<th>Days</th>
<th>Weeks</th>
<th>Months</th>
<th>Can't Determine</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTTV</td>
<td></td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
<td>15%</td>
<td>56%</td>
<td>17%</td>
<td>10%</td>
</tr>
<tr>
<td>MTTF</td>
<td></td>
<td>0%</td>
<td>0%</td>
<td>11%</td>
<td>36%</td>
<td>35%</td>
<td>8%</td>
<td>10%</td>
</tr>
<tr>
<td>MTTK</td>
<td></td>
<td>0%</td>
<td>0%</td>
<td>9%</td>
<td>40%</td>
<td>29%</td>
<td>13%</td>
<td>9%</td>
</tr>
<tr>
<td>MTTI</td>
<td></td>
<td>11%</td>
<td>23%</td>
<td>29%</td>
<td>16%</td>
<td>8%</td>
<td>3%</td>
<td>10%</td>
</tr>
</tbody>
</table>

-Ponemon Institute
The Problem: Incident Investigation & Response

<table>
<thead>
<tr>
<th>Time</th>
<th>Seconds</th>
<th>Minutes</th>
<th>Hours</th>
<th>Days</th>
<th>Weeks</th>
<th>Months</th>
<th>Can't Determine</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTTK</td>
<td>0%</td>
<td>0%</td>
<td>9%</td>
<td>40%</td>
<td>29%</td>
<td>13%</td>
<td>9%</td>
</tr>
<tr>
<td>MTTF</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
<td>15%</td>
<td>56%</td>
<td>17%</td>
<td>10%</td>
</tr>
<tr>
<td>MTTI</td>
<td>11%</td>
<td>23%</td>
<td>29%</td>
<td>16%</td>
<td>8%</td>
<td>3%</td>
<td>10%</td>
</tr>
</tbody>
</table>

-Ponemon Institute
Detect, Understand, & Respond

Organizations are focused on **detecting compromises quicker** and **understanding the scope of the compromise** in order to adequately remediate the breach.

**WHERE** the attacker entered and **HOW** the network was compromised → the attackers TTPs

**WHEN** and how long has the adversary compromised my network

**WHY** you were compromised and what the motivations of the adversary is

**WHAT** damage was done by adversary

**WHO** compromised your network and what type of adversary it is

Asking the **RIGHT** Questions
The Solution: Process Orchestration & Task Automation in IR

Key Benefits

**Accelerated Response Times**
Automate the incident response processes to speed reaction and scale human skill.
- Respond to threats in real time and without human intervention

**Efficiency**
Streamlines repeated processes.
- Reduces the risk of human error in the investigation process.
- Saves time, money and resources.

**Risk Reduction**
Reduces risk of threats slipping through the cracks by providing the ability to investigate more alerts

**Simplified IR Process**
Reduces Complexity, and Disruption of the Incident Response process.

**Empower Users**
Bridge the skills gap to enable less skilled analysts tackle more complex threat investigations.

**Workforce Satisfaction**
Employee attrition from long hours of work will be reduced.
- Working on mundane uninteresting work will be eliminated
Automated Incident Investigation and Remediation

How elements interrelate in a computer network (ability to connect information)

How and where a specific cyber threat interacts with the computer system (threat related information)

Methods of investigating specific cyber threats

Pivots in technique/process due to additional indicators

Choosing investigative technique/method

Self Improvement/Continuous Learning

System and Data Interaction

Incident Investigations Data Bank

Decision Making

Human Analysis Capabilities

Capability Framework
Our Beginning - Unleashing the Bots

Our initial research started with one objective in mind, to build a platform agnostic engine to moderate various investigative bots that can carry out specific investigative tasks.

1. Sends alerts to automated incident engine to kick off investigative techniques
2. Enables investigative bots
3. Predefined Investigative Techniques
4. Bots request specific timeframe of network logs
5. Queries and provides access to data for various investigations
6. Findings from automated bot incident bot investigation
Enabling an Incident Investigation Capability

Effective Automation of Incident Investigations involves leveraging key capabilities

**Comprehensive Cyber Ontology**
A combination of multiple multi-focused cyber security ontologies that transforms the investigative task by connecting information

**System Interconnectivity and Data Sharing**
Data drives all incident response investigations. Access to the right type of data and having the ability to utilize data-on-demand from any system to enrich the investigation is critical

**Global Incident Investigation Corpus**
A dictionary of investigative techniques used for detecting and investigating threats and the data needed for investigation

**Cognitive Decision Making & Learning**
Automation in complex processes when a machine has ability to pivot and address unaccounted specified workflows
Modeling Manual Investigations

A framework to describe an investigation at different levels of detail. This includes independent vocabulary can be used to describe the researching process in more detail and the ability for different technologies to communicate and exchange data based on well-defined widely adopted standards.

<table>
<thead>
<tr>
<th>Layer 1</th>
<th>Ontology</th>
<th>Layer 2</th>
<th>Interoperability</th>
<th>Layer 3</th>
<th>Investigation Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Unified Cyber Ontology (UCO)</td>
<td></td>
<td></td>
<td></td>
<td>• Incident Investigation Reports</td>
</tr>
<tr>
<td></td>
<td>• Integrated Cyber Analysis System (ICAS)</td>
<td></td>
<td></td>
<td></td>
<td>• Incident Investigations Corpus</td>
</tr>
<tr>
<td></td>
<td>• CMU Insider Threat Ontology</td>
<td></td>
<td></td>
<td></td>
<td>• Digital Forensics Analysis Expression (DFAX)</td>
</tr>
<tr>
<td></td>
<td>• Malware Classes Ontology</td>
<td></td>
<td></td>
<td></td>
<td>• Common Remediation Enumeration (CRE)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Cognitive Decision Making</td>
</tr>
</tbody>
</table>
COMPREHENSIVE CYBER ONTOLOGY
HUMAN-LIKE DECISION MAKING
Making Complex Decisions
For automation in incident investigations to be as effective as human incident responders it to be able to:
• Make decision on how investigative an alert. Which path/ technique from the investigative corpus is most suitable for each alert and the reasons for taking that path
• Pivot and provide alternative investigative techniques if the codified path/technique for investigating an alert is not sufficient

Learning & Continuous Improvement
To keep automated incident investigations current and updated it needs to:
• Have the ability to capture human decision making and model it into new automated investigative path/techniques to be used in automated decision making.
• Learn new ways to investigate threats that are not currently captured in the incident investigation corpus and existing ontologies
Cognitive Modeling – Core and Alternate Path Selection

Initial/core path for an Incident Investigation

- We use the **Binary Tree Traversal algorithm (BTT)** – Preorder traversal to determine the most efficient path for each type of investigation. This decision making technique is used every time a new manual investigative path is added to the list of paths in the incident investigation engine.
- Currently we are manually scoring the relevance of a path
- With BTT we plot out all possible paths of an incident investigation, computer generated workers go through every path possible while relaying the efficiency and quality of the path. This allows future iterations to go down the more efficient paths.
Cognitive Modeling – Binary Tree Traversal

Sample paths checking for malicious process

- **A**: Check running process
  - **B**: Compare pid to network statistics
    - **C**: Compare external IP to Virus Total DB
  - **E**: Check startup for persistent process
    - **F**: Check registry for startup processes
  - **D**: Compare to bad process list
Cognitive Modeling – Binary Tree Traversal

Binary tree goes down the left most path for each node and takes the score of that path

Path 1: A, B, C = 10
**Cognitive Modeling – Binary Tree Traversal**

When going left is no longer an option the tree starts traversing down the right path, takes the score of that path.

**Diagram:**

- Path 2: A, D, E = 15

**Nodes:**
- A: Check running process
- B: Compare pid to network statistics
- C: Compare external IP to Virus Total DB
- D: Compare to bad process list 10
- E: Check startup for persistent process 5
- F: Check registry for startup processes
Cognitive Modeling – Binary Tree Traversal

Binary tree goes down the final path and obtains the score of that path

Path 3: A, D, F = 13
Cognitive Modeling – Binary Tree Traversal

Path A, D, E becomes the core path since it retrieved the best score, the rest become alternate paths.
Modelling Manual Investigation & Updating Workflows

- This process records manual incident investigations by analyst & replicates in automated workflows that can be used to update workflows in the automated investigation engine to provide new investigative paths and its associated scores.

- The algorithm that helps achieve this is a modified version of Critical Path Analysis.

- After each step of a manual investigation by an incident responder is recorded we see which paths have been taken the most by these incident responders. That path is added as an additional branch to the workflow to later be analyzed by our binary tree traversal model. The other paths are stored for later, and used when an investigation wasn’t resolved by the specified paths then we go back to these alternate paths for further investigation.
Cognitive Modeling – Modified Critical Path Analysis

**Manual Scenario:** Diagram shows an example of an incident investigation for the same event by multiple incident responders. Each path recorded shows the investigative path that the incident responder took to arrive at their conclusions (end).
Cognitive Modeling – Critical Path Analysis

Part 2: Check the first node of every path and it’s following node for similarities

<table>
<thead>
<tr>
<th>First Node</th>
<th>Second Node</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td>Process</td>
<td>1</td>
</tr>
<tr>
<td>Process</td>
<td>Network</td>
<td>1</td>
</tr>
<tr>
<td>Registry</td>
<td>Process</td>
<td>1</td>
</tr>
<tr>
<td>Memory</td>
<td>Registry</td>
<td>1</td>
</tr>
</tbody>
</table>
Cognitive Modeling – Critical Path Analysis

Part 3: Check the second node of every path and its following node for similarities

<table>
<thead>
<tr>
<th>First Node</th>
<th>Second Node</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td>Process</td>
<td>1</td>
</tr>
<tr>
<td>Process</td>
<td>Network</td>
<td>2</td>
</tr>
<tr>
<td>Registry</td>
<td>Process</td>
<td>1</td>
</tr>
<tr>
<td>Memory</td>
<td>Registry</td>
<td>1</td>
</tr>
<tr>
<td>Network</td>
<td>Software</td>
<td>1</td>
</tr>
<tr>
<td>Registry</td>
<td>End</td>
<td>1</td>
</tr>
</tbody>
</table>

Diagram:
- Malware
- Memory
- Process
- Registry
- Memory
- Process
- Network
- Registry
- Network
- Software
- Network
- End
Cognitive Modeling – Critical Path Analysis

Part 4: Check the third node of every path and its following node for similarities

<table>
<thead>
<tr>
<th>First Node</th>
<th>Second Node</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td>Process</td>
<td>1</td>
</tr>
<tr>
<td>Process</td>
<td>Network</td>
<td>2</td>
</tr>
<tr>
<td>Registry</td>
<td>Process</td>
<td>1</td>
</tr>
<tr>
<td>Memory</td>
<td>Registry</td>
<td>1</td>
</tr>
<tr>
<td>Network</td>
<td>Software</td>
<td>1</td>
</tr>
<tr>
<td>Registry</td>
<td>End</td>
<td>1</td>
</tr>
<tr>
<td>Software</td>
<td>End</td>
<td>1</td>
</tr>
<tr>
<td>Network</td>
<td>End</td>
<td>1</td>
</tr>
</tbody>
</table>
Cognitive Modeling – Critical Path Analysis

Part 5: Create new path based on previous nodes investigations

<table>
<thead>
<tr>
<th>First Node</th>
<th>Second Node</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td>Process</td>
<td>1</td>
</tr>
<tr>
<td>Process</td>
<td>Network</td>
<td>2</td>
</tr>
<tr>
<td>Registry</td>
<td>Process</td>
<td>1</td>
</tr>
<tr>
<td>Memory</td>
<td>Registry</td>
<td>1</td>
</tr>
<tr>
<td>Network</td>
<td>Software</td>
<td>1</td>
</tr>
<tr>
<td>Registry</td>
<td>End</td>
<td>1</td>
</tr>
<tr>
<td>Software</td>
<td>End</td>
<td>1</td>
</tr>
<tr>
<td>Network</td>
<td>End</td>
<td>1</td>
</tr>
</tbody>
</table>
Future Work & Improvements

Benchmarking
To compare the speed and accuracy of the automated system to the speed with which a human will perform the same investigation.

VS.

Improvements
• Incorporate the BBT algorithm for picking core and alternate paths as part of one automated system
• Build investigative bots directly from SPARQL queries from ontologies and automatically pick core and alternate paths without any scoring

Improved Cognitive Learning Techniques
– Markov Chain
• With more investigative data we can start using the Markov chain to improve the advanced decision making model
## Automating Incident Investigations

### Blockers

**Comprehensive Ontology**
- Ontologies available currently address piecemeal incident investigations concepts and not the incident investigation life cycle as a whole.

**Investigative Corpus**
- Most incident responders are not willing to freely submit their knowledge on incident investigations into a global database.

**Automated Investigations**
- We do not have enough data to build full cognitive models to drive automated investigations. It is also very reliant on organizations being able to connect to all systems and query data on demand.

### Existing Work

**ICAS**
- The DARPA funded Integrated Cyber Analysis System (ICAS) ontologies had a primary focus of incident response.
  

**CMU Insider Threat Ontology**
- Funded by DARPA and FBI and developed by CMU. Has an insider threat focus.
  

**Unified Cyber Ontology**
- Attempt at combining knowledge schemas from different cybersecurity systems and most commonly used cybersecurity standards
  

---

*Cyber Security Ontologies (user: Spriley) [http://cps-vo.org/node/26324](http://cps-vo.org/node/26324)*
Black Hat Sound Bites (3)