Anatomy of a Cyberattack: Standardizing Data Collection for Adversarial and Defensive Analyses

Jason Schlup, Vikram Kulkarni, Shawn Whetstone, Walter Dodson III

Institute for Defense Analyses
Operational Evaluation Division
DATAWorks 2019
Cybersecurity is **difficult**

- Technology changes rapidly
- Cybersecurity affects everyone
- Cyber activity generates tremendous amounts of data

**Problem:** How can we predict network resiliency and suggest improvements?

**Uniform terminology and easily digestible reporting enable effective cyber defense**

- Find trends in attackers
- Identify security weaknesses
- Improve cyber situational awareness
How hard is cybersecurity?

- Attackers need one success, you must defend your **entire** network *and* trash!

Cyberattacks in the news

- **Consumers/PII**
  - Yahoo (3 billion)
  - Marriott (500 million)
  - Equifax (143 million)
  - Target (110 million)
  - OPM (26 million)

- **Internet of Things (IoT)/Hardware**
  - Fish tanks
  - Spectre/Meltdown
  - Routers (VPNFilter)

- **Infrastructure**
  - Baltimore 911 services
  - Ukrainian power grid
  - German steel mill
  - British NHS
  - D.C. police cameras

- **Sensitive Data/IP Theft**
  - Oklahoma Dept. of Securities
  - US Navy contractors
  - Sony Pictures
Why is this hard to counteract?

- **Technology moves quickly**
  - New breaches daily, information/exploits traded
  - Minimal hardware/software requirements
  - Systems not designed with cybersecurity in mind

- **Defense evasion built into tools**
  - Hackers only need one success, even if it’s “trash”

- **“Big data” problem**
  - Local, remote, cloud-based data
  - Cross-domain (cyber-physical-social)


Defending a network and looking forward

- IDA looks at past cyber exercises for trends
  - Collect qualitative data (reports, logs, sensor data, emails, …)

- Requires common taxonomy and methods
  - MITRE ATT&CK™ framework[1]
  - Other choices available
    - NSA, NIST, or Lockheed Martin

- Develop quantitative measures

IDA primarily focuses on DOT&E operational tests and Cyber Assessment Program (DoD networks)

- Penetration testing, Red Teaming
  (Cooperative Assessment) (Adversary Simulation)

- Record attack threads
  - Attacker actions and defensive detections

- Interviews and follow-up
First analysis to achieve cyber picture

- Phishing email with malicious payload
- Router
- Firewall
- Attacker
- Internet

Enables attack

**Ingress:**
First system accessed within target network

**Lateral:**
Systems between Ingress and Objective

**Objective:**
Any system on which an information effect is created

Target Network
### Example attack thread – Notional data

- **IDA bins actions according to ATT&CK framework**
  - Enables analysis of attack threads

<table>
<thead>
<tr>
<th>Target IP</th>
<th>Tactic</th>
<th>Technique</th>
<th>Details</th>
<th>Tool Type</th>
<th>Detected?</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.10.1.4</td>
<td>Initial Access</td>
<td>Spearphishing Attachment</td>
<td>An email is sent with malicious executable</td>
<td>Email</td>
<td>No</td>
</tr>
<tr>
<td>10.10.1.4</td>
<td>Execution</td>
<td>User Execution</td>
<td>A legitimate user executed the payload</td>
<td>Cobalt Strike</td>
<td>No</td>
</tr>
<tr>
<td>10.10.1.4</td>
<td>Execution</td>
<td>Scripting</td>
<td>Batch file is launched from user interaction</td>
<td>Cobalt Strike</td>
<td>No</td>
</tr>
<tr>
<td>10.10.1.4</td>
<td>Execution</td>
<td>Rundll32</td>
<td>Batch file launches CS DLL payload via Rundll32</td>
<td>Cobalt Strike</td>
<td>No</td>
</tr>
<tr>
<td>10.10.1.4</td>
<td>Persistence</td>
<td>Registry Run Keys / Start Folder</td>
<td>Write new batch file to user's Startup folder</td>
<td>Cobalt Strike</td>
<td>Yes</td>
</tr>
<tr>
<td>10.10.1.4</td>
<td>Command and Control</td>
<td>Commonly Used Port</td>
<td>Uses DNS port 53</td>
<td>Cobalt Strike</td>
<td>No</td>
</tr>
<tr>
<td>10.10.1.4</td>
<td>Command and Control</td>
<td>Standard Application Layer Protocol</td>
<td>Operating over DNS</td>
<td>Cobalt Strike</td>
<td>No</td>
</tr>
</tbody>
</table>

**NOTE:** Notional data is used on this slide.

**Notional data set**
Data analysis – Notional data

- Look across attack threads
  - Was the attack detected?
  - What factors determine detection?

<table>
<thead>
<tr>
<th>Attack Thread</th>
<th>Foreign Tool Use</th>
<th>Thread Detected?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>88%</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>23%</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>80%</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>0%</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>20%</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>13%</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>50%</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>20%</td>
<td>No</td>
</tr>
</tbody>
</table>

Notional data set

NOTE: Notional data is used on this slide.
Can use data for future prediction!

<table>
<thead>
<tr>
<th>Attack Thread</th>
<th>Foreign Tool Use</th>
<th>Detected?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>88%</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>23%</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>80%</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>0%</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>20%</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>13%</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>50%</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>20%</td>
<td>No</td>
</tr>
</tbody>
</table>

Logistic regression of data set

Other controlling factors?

NOTE: Notional data is used on this slide.
### Other factors?

- **Other quantitative measures:**
  - Specific tool use
  - Time-to-objective, time-to-detection

<table>
<thead>
<tr>
<th>Attack Thread</th>
<th>Foreign Tool Use</th>
<th>PowerShell Use</th>
<th>Time-to-Objective (hr)</th>
<th>Factor 4</th>
<th>Factor 5</th>
<th>Detected?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>88%</td>
<td>30%</td>
<td>1</td>
<td>...</td>
<td>...</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>10%</td>
<td>4</td>
<td>...</td>
<td>...</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>23%</td>
<td>0%</td>
<td>0.2</td>
<td>...</td>
<td>...</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>80%</td>
<td>90%</td>
<td>0.2</td>
<td>...</td>
<td>...</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>0%</td>
<td>20%</td>
<td>0.1</td>
<td>...</td>
<td>...</td>
<td>No</td>
</tr>
</tbody>
</table>
Cybersecurity is asymmetric

Conclusions

Conclusions

- Cybersecurity is asymmetric

- If you can’t prevent the attack, at least detect it
  - What factors influence detection?
  - How can we increase detections?

- Use common taxonomy to categorize attacker behavior

- Inform defenses based on the findings
BACK-UP SLIDES
MITRE ATT&CK framework

- 11 tactics, 200+ techniques, and common knowledge
  - Initial Access
  - Execution
  - Persistence
  - Privilege Escalation
  - Defense Evasion
  - Credential Access
  - Discovery
  - Lateral Movement
  - Collection
  - Exfiltration
  - Command and Control

- Information on 78 known groups