Threat Hunting: From Fudd to Terminators

Heather Adkins
Director, Information Security & Privacy @Google
Why Are We Here?
There is little question that contemporary commercially available systems do not provide an adequate defense against malicious threat.

— James Anderson, 1972

On James Anderson: http://www.cerias.purdue.edu/site/blog/post/passing-of-a-pioneer/
Behavioral Deviation
Log Centralization
Self-Learning Expert Systems
Static Rules
Manual Analysis
Behavioral Deviation
Statistical Analysis

James Anderson
Computer Security Threat Monitoring and Surveillance

Sources: See Bibliography Slide
James Anderson  

Computer Security Threat Monitoring and Surveillance

- Self-Learning Expert Systems
- Static Rules
- Manual Analysis
- Log Centralization
- Behavioral Deviation
- Statistical Analysis
- Artificial Intelligence
- Data Mining
- Analysis GUI
- Network Security Monitoring
- Neural Networks
- Genetic Algorithms

Sources: See Bibliography Slide
Behavioral Deviation
Log Centralization
Self-Learning Expert Systems
Static Rules
Manual Analysis
Behavioral Deviation
Statistical Analysis
Artificial Intelligence
Data Mining
Analysis GUI
Stats + Behaviors + Rules
Network Security Monitoring
Machine Learning
Neural Networks
Genetic Algorithms

Sources: See Bibliography Slide
Common Element: Humans
Analysts That Stare At Logs

Feb 17 00:31:58 marvin: ASL Sender Stats...
Feb 17 13:35:14 marvin: server Reconnect...
Feb 17 13:36:25 marvin: Error Antivirus...
Feb 17 14:55:19 marvin: Unknown IP Login...
Feb 18 16:50:19 marvin: Error Missing File...
Feb 19 01:22:56 marvin: ASL Sender Stats...
Feb 22 06:02:19 marvin: Unknown IP Login...
Feb 22 06:32:22 marvin: Kernel Panic!
James Anderson

Computer Security Threat Monitoring and Surveillance

Auditor
Security Manager
Security Analyst
Activity Security Officer
System Security Officer

Security Administrator

Security Operations Center
Security Engineer

Sources: See Bibliography Slide
James Anderson

Computer Security Threat Monitoring and Surveillance

Sources: See Bibliography Slide
Analysts Need Data

We Need Experience Data Overload

We Need Technology: Automation, ML, AI, etc

Analysts Do It Better

Technology Doesn’t Work

Analysts Need To Detect Malicious Activity

Every ~5 years we “re-invent” the Security Analyst & Build Old New Stuff.
Centralized Logs + Printers

Database-Backed GUIs + Pagers

Signatures, SIEMs
SOCs

Threat Hunting Tools, SIEMs, Data Science, Bears, Oh My
~4 years

Estimated length of career for a consistent Greyhat / Whitehat exploit developer.

The same for Security Surveillance? Causality: Technology Fatigue?

2016: Team Mayhem Wins DARPA Challenge. First Cyber-Reasoning System To Play CTF

Humans Program Surveillance Systems

Humans Build Self-Learning Systems

Humans Teach The Machines (Intelligent Assitants)

The Machines Defend Themselves

Terminators Can Spell

Source: http://www.darpa.mil/news-events/2016-08-05a
“Over time, the computer itself -- whatever its form factor -- will be an intelligent assistant helping you through your day.”

— Sundar Pichai, 2016
Google CEO
The Detection Pipeline

DATA STORE (PB)

- INTEL Malware IoC Feeds
- Fleet Data
- Log Ingestion 100 GBs / Day
- Real Time

- Normalize
- Static Rules
- Advanced Models
- Correlation
- Testing

- Enrich

EVENT (SIEM)

- Malware
- IoC Feeds
- Fleet Data
- Real Time

Log Ingestion 100 GBs / Day
The Detection Pipeline

DATA STORE (PB)

INTEL Malware IoC Feeds
Fleet Data
Log Ingestion 100 GBs / Day
Real Time

Testing
Static Rules
Correlation
Advanced Models

Enrich
EVENT (SIEM)

Normalize

Malware IoC Feeds
Fleet Data
Log Ingestion 100 GBs / Day
Real Time
A Bit Like Everyone Else

Threat Analysis

Security Surveillance (Detection)

Incident Management

Security Monitoring Tools

Digital Forensics
Golden Rule: Hunt Once

- Detector Dev
- Write Tests
- Research
- Build Hypotheses
- Hunting

Source:

Where Useful?

- Developing New Detection
- Assess New Environments
- Spot Checks
- Validating Tools
- “Train Like You Fight”
Hunting Grounds

Stored Raw Data
- DNS Logs, File Monitoring, User Behavior, etc.

Pre-Processed Signals
- Database store + easy search (Hunt) interface
- Unsifted High-FP Noisy Signals

Live Fleet
- GRR (agent-based)
- Memory, Registry, Startup Items, BIOS, Firmware, etc.
<table>
<thead>
<tr>
<th>TIME</th>
<th>ACTIVITIES</th>
</tr>
</thead>
</table>
| 40%  | Tactical Surveillance ("Ops")  
Event Triage + Hunting in Pre-Processed Data, Fleet Wide Checks |
| 40%  | Technology Specific Projects  
Threat Intel Integration, Hypothesis-Driven Hunting, Detector Development, Testing, Automation |
### Hypothesis

<table>
<thead>
<tr>
<th>Stored Raw Data</th>
<th>Lots Of SQL</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Attackers are using dyndns hostnames for C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attackers might be using DNS queries for hostnames never seen before.</td>
</tr>
</tbody>
</table>

### Method

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Method</th>
<th>Pros/Cons</th>
</tr>
</thead>
</table>
| Attackers are using dyndns hostnames for C2 | 1. Compare DNS queries with known dyndns provider domains.  
2. Triage results. | Pro: Scan a new environment (effective)  
Pro: Processing / triage can be scripted. |
| Attackers might be using DNS queries for hostnames never seen before. | 1. Collect DNS Query data for 60+ days,  
2. sort to find things ‘first seen’ recently, especially during incidents. | Con: Manual and complicated  
Tip: Filter out Ad networks |

DynDNS Hostnames: [https://gist.github.com/neu5ron/8dd695d4cb26b6dcd997](https://gist.github.com/neu5ron/8dd695d4cb26b6dcd997)
<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Method</th>
<th>Pros/Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attackers have moved laterally using pass-the-hash and ‘at’ jobs</td>
<td>1. Collect a copy of every C:\Windows\Tasks*job files in the fleet. 2. Extract what was executed, compare.</td>
<td>Pro: Scan a new environment (effective) Pro: Processing / triage can be scripted.</td>
</tr>
<tr>
<td>Attackers maintaining persistence via LaunchAgents, LaunchDemons, Plists (Mac)</td>
<td>1. Collect data from fleet 2. Triage.</td>
<td>Con: Long tail of results Con: Tons of clumsy adware</td>
</tr>
</tbody>
</table>

https://github.com/google/grr
<table>
<thead>
<tr>
<th>Live Fleet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis</td>
</tr>
<tr>
<td>Attackers will execute new binaries on our systems.</td>
</tr>
</tbody>
</table>
### Hypothesis

Attackers have backdoored our firmware

### Method

Use Chipsec (Integrated with GRR) to get a copy of BIOS / UEFI. Cluster, triage.

### Pros/Cons

- **Pro:** Mac environments cluster nicely
- **Con:** PC environments do not cluster nicely
- **Con:** Specialized skills required to triage
- **Con:** Difficult to find baselines

---

Talk: Firmware Biopsy by Thiebaud Weksteen @Ruxcon 2016
<table>
<thead>
<tr>
<th>Task</th>
<th>Toolset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefetch</td>
<td>suid files</td>
</tr>
<tr>
<td>LNK Files</td>
<td>BIOS / UEFI / EFI</td>
</tr>
<tr>
<td>At Jobs</td>
<td>MBR</td>
</tr>
<tr>
<td>autoruns</td>
<td>Memory + Yara</td>
</tr>
<tr>
<td>Recycler</td>
<td>Memory + Rekall Mods</td>
</tr>
</tbody>
</table>
1. Study Your History!
2. Prepare For The Future (4 years!)
3. Surveillance Time Allocation
4. Hunt Once; Then Automate
5. Find The Right Tools (GRR!)
Some Light Reading…

Stalking the Wily Hacker, Clifford Stoll, 1988
An Intrusion-Detection Model, Dorothy Denning, 1986
IDES: The Enhanced Prototype, 1988


http://www.cerias.purdue.edu/about/history/coast_resources/idcontent/ids.html
**Hypothesis**

Attacker have stored their binaries in the Windows Registry.

**Method**

Collect Registry contents where contents > XXX MB.

**Pros/Cons**

Con: Triage requires special knowledge
Con: Unexpected findings