SIRI FOR THE SOC
HOW AN INTELLIGENT ASSISTANT CAN AUGMENT THE SOC TEAM

Bobby Filar, Rich Seymour
SANS SOC Summit 2017
WHO ARE WE?

BOBBY FILAR
- Data Scientist
- Background in NLP
- @filar
- bfilar@endgame.com

RICH SEYMOUR
- Data Scientist
- Background in HPC
- @rseymour
- rseymour@endgame.com
Our Initial Problem

- Evented data on endpoints
- Event based data is Big Data problem
  - Big data == big problems
- How do we solve it?
### User-Centric Design Study

- **GOAL:** Capture team dynamics and worker roles within security organization to identify challenges common across security teams

<table>
<thead>
<tr>
<th>User Group</th>
<th>Team Type</th>
<th>Environment</th>
<th>Collection Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Traditional SOC</td>
<td>Day-to-day use</td>
<td>User interviews</td>
</tr>
<tr>
<td>B</td>
<td>Novice Training Team</td>
<td>Mock Scenario</td>
<td>Side-by-side monitoring &amp; interviews</td>
</tr>
<tr>
<td>C</td>
<td>Internal Red vs. Blue</td>
<td>Mock Scenario</td>
<td>Retrospective</td>
</tr>
<tr>
<td>D</td>
<td>Trad. SOC &amp; Consulting group</td>
<td>Day-to-day use</td>
<td>User testing</td>
</tr>
</tbody>
</table>
Findings: Security Worker Roles

**Tier 1 User**
- Have little to no prior experience (average of 1 year) in the cyber security space. First line of defense in a Security Operations Center.
- Main responsibility is to initially triage alerts and determine if escalation (to higher tiered) is required.
- Primarily rely on a platform’s GUI.

**Tier 3 User**
- Intimately understand network and platform architecture.
- Seen as domain experts on the SOC team and more comfortable working through the command line.
- Investigates escalated alerts, and determine root causes and extent to remediate problems.

**Forensic Investigator**
- Expert in EDR platforms and sophisticated investigation tools.
- Uses command line and scripting languages to bypass UI and collect large data feeds using 3rd party APIs.
Findings: Day in Life of a Security Analyst

- Data Deluge
- Lack of Context
- Repetitive Processes
- Searching not Analyzing
- Lack of Expertise
- Lack of Time
Example: EDR Alert

Alert Type: Suspicious Binary
Alert Created: Feb 11, 2017
Severity: High
Confidence: 73%
File Path: C:\Temp\aaa.exe
File Size: 45700
MD5: 5d41402abc4b2a76b9719d911017c592
File Created: Feb 11, 2017

What do you do when there are 100s of these each day?
Example: EDR Alert

- Lacks context
  - Is it actually bad?
  - Is it anywhere else?
  - Did it talk to the network?

- Lacks connectivity
  - Is this alert tied to any others?

- Pivot on single IOC
  - Hash
  - Filename
  - IP address

Alert Type: Suspicious Binary
Alert Created: Feb 11, 2017
Severity: High
Confidence: 73%
File Path: C:\Temp\aaa.exe
File Size: 45700
MD5: 5d41402abc4b2a76b9719d911017c592
File Created: Feb 11, 2017

What do you do when there are 100s of these each day?
Additional Findings

- More than search
- **AUTOMATE ALL THE THINGS!!1!1!!**
  - Attack Timeline
  - Process Lineage
  - Custom Workflows
  - Hunts
    - Malicious Powershell
    - C2 Beaconing
  - Netflow Activity
- Collaboration
Core Problems

Insufficient Resources
- Onboarding & training new hires
- Limited time to review alerts and incidents

Lack of easy-to-use automated tools
- Difficult for non-programmers to use
- Easy for programmers to mess up!

Security platforms are just difficult to use!
- Forces conformity
- Requires level of expertise to extract value
Solution

- A Bot is an application that assists in the automation of tasks
  - Mimics human conversation
  - Natural Language Understanding determines user intent

- Imagine an assistant that provides ability to:
  - Ask questions
  - Execute workflows
  - Educate users
  - Recommend next steps
Bot 101
Managing Expectations

VS.

Remember me?
A.I. → Almost Intelligent

- A user has a goal in mind, the bot needs to take that intention and turn it into action with the right variables assigned.

- At worst a bot is an interactive form

- At best a bot can predict what you need and help guide you
Click-Through vs. Conversational Interfaces

I’d like to order a cheese pizza.

Great! What size pizza would you like?

Medium

Okay. I have an order for a medium cheese pizza. Is that all?

Yes. That is all.

+ 6 more steps
Evolution of (Security) Bots

GO AWAY

OR I WILL REPLACE YOU WITH A VERY SMALL SHELL SCRIPT
Evolution of (Security) Bots

- Bash and Cron (and Perl) automating
- IRC bots (Nickserv & Chanserv)
- Hubot
- SecurityBot
- Demisto
- Twitter Bots
Bot Development Kits (BDKs)

- The rise of bots is largely due to an increase in BDKs
- Companies like Chatfuel, Wit.ai, and API.ai
  - Provide simple UI for development
  - *Closed-domain, Rule-based, goal-oriented bots*
  - No programming skills required
- How popular have these frameworks become?
  - Wit.ai → purchased by Facebook *(Jan 2015)*
  - API.ai → purchased by Google *(Sep 2016)*
- Major companies pushing out frameworks too!
  - Microsoft Bot Framework
  - Amazon Alexa Skills Kit
Bot Architecture

- Rule-based & machine learning
  - Dependent on amount of training data
  - NLP/ML is used behind the scenes, but not forced upon developer
- Construct a dialogue tree designed to complete a task
  - Each branch represents question the user must answer to progress
Bot Workflow

- Session
- Identity
- Dialogue Script
- Context
- Intents
- Entities
- Action
Context & Intent

- **Context (teal)**
  - Current state
  - Steers conversation
  - Ensures all req’d parameters are collected to perform **Action**

- **Intent (red)**
  - Maps **Utterance (blue)** → specific **Action**
  - **What does the user want?**
Entities

- Entities are parameters needed to satisfy **Intents** and execute **Actions**.

- Entities can be anything
  - File hash
  - IP address
  - Cities

- Permutations of an entity allow for **diversity of vocabulary**
  - Regular Expressions
  - Synonyms
Training Data

- Few bot developers have necessary training data
- BDKs provide access to Domains or pre-defined Knowledge packages:
  - Reservations
  - Small Talk
  - Weather
- Packages include common intents and entities to train your bot
- BDKs log user interactions to constantly update training set
NLP 101 → Direct Matching (No ML)

Utterance
Show processes for 1234123412341234 on Windows endpoints since Monday

Tokenization
Show processes for 1234123412341234 on Windows endpoints since Monday

Parts-of-Speech Tagging (POS)
Show processes for 1234123412341234 on Windows endpoints since Monday

Named Entity Recognition (NER)
Show processes for 1234123412341234 on Windows endpoints since Monday

Perform Action
if utterance contains show_vb AND processes_nns:
SHOW_PROCESS(md5, “windows endpoints”, “since Monday”)
NLP 101 → ML-Based

Utterance
Show processes for 1234123412341234 on Windows endpoints since Monday

Tokenization
Show processes for 1234123412341234 on Windows endpoints since Monday

Bag-of-Words (Feature Vector)
0 1 1 0 1 0 1 0 1 1 0 0 1

Intent Classification
SHOW_PROCESS

Named Entity Recognition (NER)
Show processes for 1234123412341234 on Windows endpoints since Monday

Perform Action
SHOW_PROCESS(md5, “windows endpoints”, “since Monday”)
**Actions**

- **Actions** are the outcome of an intent with the necessary parameters.

- Actions can be anything
  - A response
  - A knowledge base lookup
  - Script execution
  - Automation (do this daily)
Bots for Everyone

- Bots don’t have to be super complex AI
- Build them any way you like
- You can run them anywhere
- Automate custom scripts and workflows or tie into larger platforms
Case Study: Artemis
Design Goal

The challenge is building a feature that **augments** one part of the user base, but doesn’t **hinder** the other.

Steve Jobs (probably)
Architecture (Simplified)

Platform
Natural Language vs Query Language

Query Language
SELECT * FROM TABLE process_event WHERE process_name == “odinaff.exe”;

Natural Language
Show me process event data for odinaff.exe
User: Search processes

Artemis: Okay. Please provide a hash or filename
User: odinaff.exe

Artemis: Got it. Which endpoints would you like to target?

User: Windows 10 machines.

Artemis: Okay. Would you like to launch this search?

User: Yes
Interaction Types

Goal-oriented Conversation

**User:** Show me process event data for odinaff.exe on all Windows 10 endpoints

**Artemis:** Okay.
Interaction Types

API-Driven Investigations

```bash
curl 'api/v1/event_search' -H "Content-Type: application/json" -H 'authorization: <api_key> --insecure --data-binary
'{"intent":"search_process",
"parameters": {
    "process_name":"odinaff.exe",
    "filepath":"C:\Temp\*.exe"
}}
```
Search Space

Endgame Prevention & Detection Data

MalwareScore
Persistence
Credential Dumping
Permission Theft
Adversary Evasion

Endpoint Event Data

Process
File
Users
Network
Registry
Powershell Misuse Investigation

TRADITIONAL INVESTIGATION

1. Narrow scope to limited endpoints
2. Understand adversary TTPs
3. Gather events from limited endpoints
4. Analyze events from for signs of TTPs
5. Discover suspicious activity
6. Decode obfuscated commands
7. Pinpoint powershell activity
8. Expand scope to next set of endpoints
9. Repeat…

ENDGAME ARTEMIS

“Find powershell activity”

Automatically discovers and analyzes malicious activity across your global enterprise in minutes
Demo Time!
Thank You

Contact:
bfilar@endgame.com   @filar
rseymour@endgame.com @rseymour