To Automate or Not To Automate

Brian Carrier
VP of Digital Forensics

by BASIS TECHNOLOGY

Goals for this Session

- We’re going to talk about automation.
  - Where should it be used?
  - Do you need it?
  - Is it worth it?
- This is supposed to be interactive.
- Feel free to shout and argue.

Agenda

- Define the basics and an evaluation framework.
- Dive into a couple steps.
- Talk about what a typical company can do.

- More details on the blog. Not enough time today.
  - http://www.cybertriage.com/

What is Automation?

- There are a lot of definitions out there.
- Many are very academic.
  - See John Allspaw (Etsy) blog. Not security related, but a must read:

http://www.kitchensoap.com/2012/09/21/a-mature-role-for-automation-part-i/

- My simplistic definition:
  - When the computer does the next step without human intervention.

This Is Automated…

No human intervention…

Levels of Automation

- Expectations of automation change.
  - 15+ years ago automation in DFIR was only about reducing steps.
    - Pressing a button in a UI instead of several command line tools.
    - Running a script from a floppy that ran several tools.
  - Now we are looking beyond that.

	The computer…
1	Offers no assistance: human must take all decision and actions.
2	Offers a complete set of options.
3	Offers a narrowed set of options.
4	Suggests one of the options.
5	Executes the suggested option if human approves.
6	Executes the suggested option unless human vetos it in a given time.
7	Executes the suggested option automatically and informs human after.
8	Executes the suggested option automatically and informs human only if asked.
9	Executes the suggested option automatically and informs humans only if the computer decides to.
10	The computer decides everything and acts autonomously, ignoring the human.

From: Sheridan and Verplank

My Simple Levels

- Manual:
  - No computer assistance
- Partially automated:
  - Some human interaction
- Fully automated:
  - No human interaction

When Should We Automate

- Choose the highest level of automation where:
  - The benefit of automatically performing the next step is high.
  - You can afford the cost (software, hardware, and personnel) associated with its complexity.
  - The impact and likelihood of a mistake are acceptable

Automation in IR

- The traditional way to think about IR:
  - Identification
  - Containment
  - Investigation
  - Eradication
  - Recovery

IR Activities

- From a technical perspective, I organize IR into two types of work:
- Investigation: Collecting and analyzing data to answer questions.
  - Identification and Investigation phases
- Mitigation: Taking actions to reduce further damage.
  - Containment, eradication, and recovery phases

Investigation Activities

- Common Questions:
  - Triage: Is it compromised? How badly?
  - Deep Dive / Forensics: Who, When, Why?
  - Hunting: Which other computers have this file?
- General Process:
  - Define the questions
  - Collect data that can answer the questions
  - Analyze the collected data
  - Answer the question based on the results
- Each of these can be broken down further.

Investigation Breakdown

Data Collection			Data Analysis			Inference
Pick Host	Pick Data Types	Get It	Known Bads	Typically Bads	Anomalies	

- **Note:** Continuous monitoring systems are always collecting

Investigation Example

- **Alert:** C&C traffic from a laptop.
- **Question:** “What on the laptop is doing that?” (Triage)
- **Data Collection**:
  1. Identify Systems: The laptop
  2. Identify Data Types: Volatile data, startup items, scheduled tasks, user activity.
  3. Collect The Data
- **Data Analysis**
  1. Known Bad / IOC: Malware signatures, etc.
  2. Typically Bad:
  3. Abnormal: Suspicious processes, scheduled tasks, etc.

Apply the Framework

- Review each step with the framework to decide if it should be automated.
- Framework Review:
  - What are the benefits of automatically performing the next step instead of doing it manually?
  - What is the cost to automate?
  - How likely is a mistake and what is the impact?

Identify the Target System

- Idea: Automatically identify what remote system to collect data from based on:
  - The investigation type
  - What questions need to be answered
- Example Scenarios:
  - A NIPS alert about C&C traffic from a laptop.
  - During an investigation, hostnames are found associated with network shares and remote desktop activity.
- Would you automatically collect them? Who already automates this?

Identify System: Pros / Cons

- Benefit of automation: Speed
  - Reduce time because collection starts faster
- Mistakes:
  - Collect from a system that wasn’t needed.
    - Performance impact on target system and “server”
  - Don’t collect a system that was needed.
    - Can always start it manually.

Identify Systems: Cost

- Full: Computer decides based on incident & questions
  - Very high cost to support all incidents and questions. Needs to know how to answer any question.
  - Low cost for limited scenarios: Specific SIEM alerts
- Partial: Computer finds references to remote systems and human decides.
  - Low cost and low error rate
- Manual: Human finds host references and copies them.
- What should we do? What would you pay for?

Abnormal Analysis

- Idea: The collected data is automatically analyzed for abnormal things (not IOCs).
- Examples:
  - User accessing network shares in a different department
  - Suspicious startup programs
  - Suspicious local user accounts
- Who automates this?

Abnormal Analysis: Pros / Cons

- Benefit of automation:
  - Can answer questions faster (reduce dwell time).
  - Could be more accurate (humans can’t remember what is normal everywhere)
- Mistake: Wrong result
  - Miss evidence that would have been suspicious to a human. Humans can be better at new trends.
  - Generate false positives

Abnormal Analysis: Complexity

- Full: Computer knows everything (servers, user roles, how each role should behave, etc.). It knows what normal is.
  - Very high cost to support all scenarios with low error rate.
  - High cost for limited scenarios with low error rate.
- Partial: Human sees list of possible abnormal data.
  - Medium cost
- Manual: Human reviews all data
- What should we do?

The Other Steps

- We can continue this analysis for each investigation and mitigation step.
- But, we don’t have time.

What Can the Typical Company Do?

- What can you do today?
- I’m going to assume that the typical company:
  - Is most interested in triaging their alerts
  - Outsources deep diving / forensics
  - Does not have huge investment yet in a platform to learn what is normal and orchestrate mitigation.

- Who here fits into that category?

Automation for the 99%

Step 1: Fully automate steps that don’t require decisions.

- Low cost and low risk of mistakes.
- Doing a collection (manually identify target and data)
- Analyze for IOCs.

How many of you are doing this?

Automation for the 99% (2)

Step 2: Partially automate some of the “easier” decisions:

- Abnormal Analysis: Let computer help you identify suspicious data.
  - It will never be fully automated.
- Choosing target systems and types by integrating with your SIEM or platform.
  - As you learn, fully automate some alerts.

Automation for the 99% (3)

Step 3: Automate the execution of mitigation (block users / computers). Have a human make choice.

Step 4: Partially automate answering questions or choosing the mitigation approach.

Other

- There is so much more to talk about.
- How do you get people to trust automation?
- What are other risks of automation?
  - How do people learn the basics of IR if it is automated?
  - What happens when the automation doesn’t work and people are rusty and can’t do it manually anymore?

- More about this on the blog (www.cybertriage.com)

Questions?

Vote for talks at osdfcon.org

Brian Carrier
brianc@basistech.com

by BASIS TECHNOLOGY