Carnegie Mellon University
Software Engineering Institute

AI Evaluation Methodology for Defensive Cyber Operator Tools

NOVEMBER 14–16, 2022

Dr. Shing-hon Lau
Senior Cybersecurity Engineer

Document Markings

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM22-0857

Introduction: AI Tools Enhance the Cybersecurity Workforce

- Organizations within the DoD and DIB are turning to AI-powered defensive cyber operator tools to enhance their cybersecurity workforce.
- These organizations must assess whether these tools are appropriate for defending their networks.
- Our project aims to develop a methodology for evaluating these tools in a black-box manner.

Agenda

- **Motivation**
- Problem and Approach
- Methodology
- Wrapping up

Organizations Are Turning to AI-Powered Defenses

- There is a significant shortage of qualified cybersecurity staff.
  - The US has a shortfall of 700k+ cybersecurity staff. (https://www.cyberseek.org/heatmap.html)
  - AI can act as significant force multiplier.
  - AI can address “easy” alerts, freeing human analysts to handle harder problems.
  - AI may be able to catch complex threats that may elude analyst detection (e.g., SolarWinds).

- Cyber attacks can be so rapid that human response is impractical.
  - NotPetya attack took down an entire Ukrainian bank in 45 seconds.
  - Human reaction to the threat is slow; the damage can be irreversible.

Think Before You Buy

Suppose that you were responsible for deciding whether an AI-powered defense was appropriate for your organization. What questions might you ask?

- What kinds of cyberattacks can I expect the defense to protect against?
- What kinds of cyberattacks are beyond the detection capabilities of the defense?
- Does the installation of the defense create additional vulnerabilities that an adversary may exploit?
- How do I test the defensive capabilities of what I am buying?

Agenda

- Motivation
- **Problem and approach**
- Methodology
- Wrapping up

Test and Evaluation of AI Defenses

AI defenses pose a test and evaluation challenge unlike those posed by traditional cybersecurity defenses.

- Organizations might need to evaluate tools in a black-box or gray-box environment, without direct access to the innards of the defense.
- AI defense designers intend for systems to learn from their network environment, necessitating creation of a realistic testbed.
- Designers intend for defenses to learn and change over time, so a singular evaluation is insufficient.
- Adversarial manipulation can fool AI defenses, creating vulnerabilities an adversary may exploit.

Creating a Testing and Evaluation Methodology

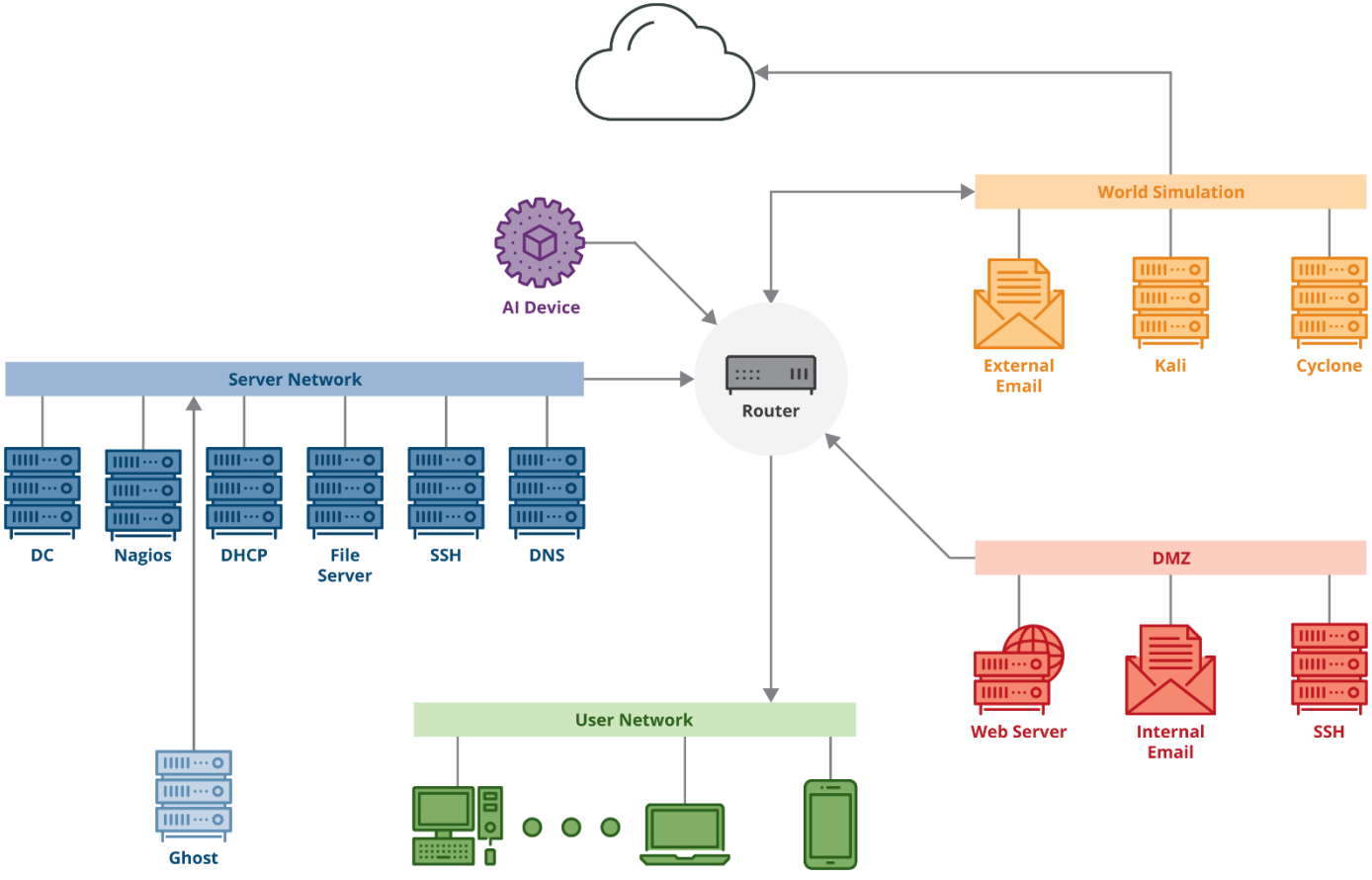
Based on the identified challenges, our methodological approach must

- create a realistic network environment where an AI defense can be deployed.
- populate that network environment with sufficiently realistic background traffic to allow the AI to learn.
- test AI defense performance against realistic cyber attacks.
- test AI defense performance when exposed to adversarial manipulation.

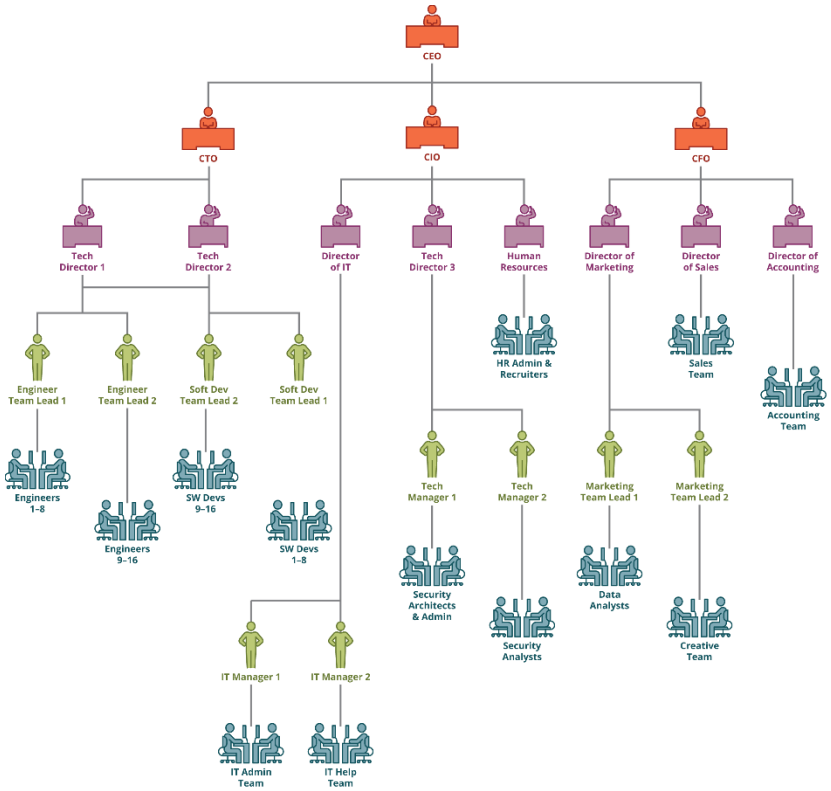
Agenda

- Motivation
- Problem and approach
- **Methodology**
- Wrapping up

Creating a Network Environment

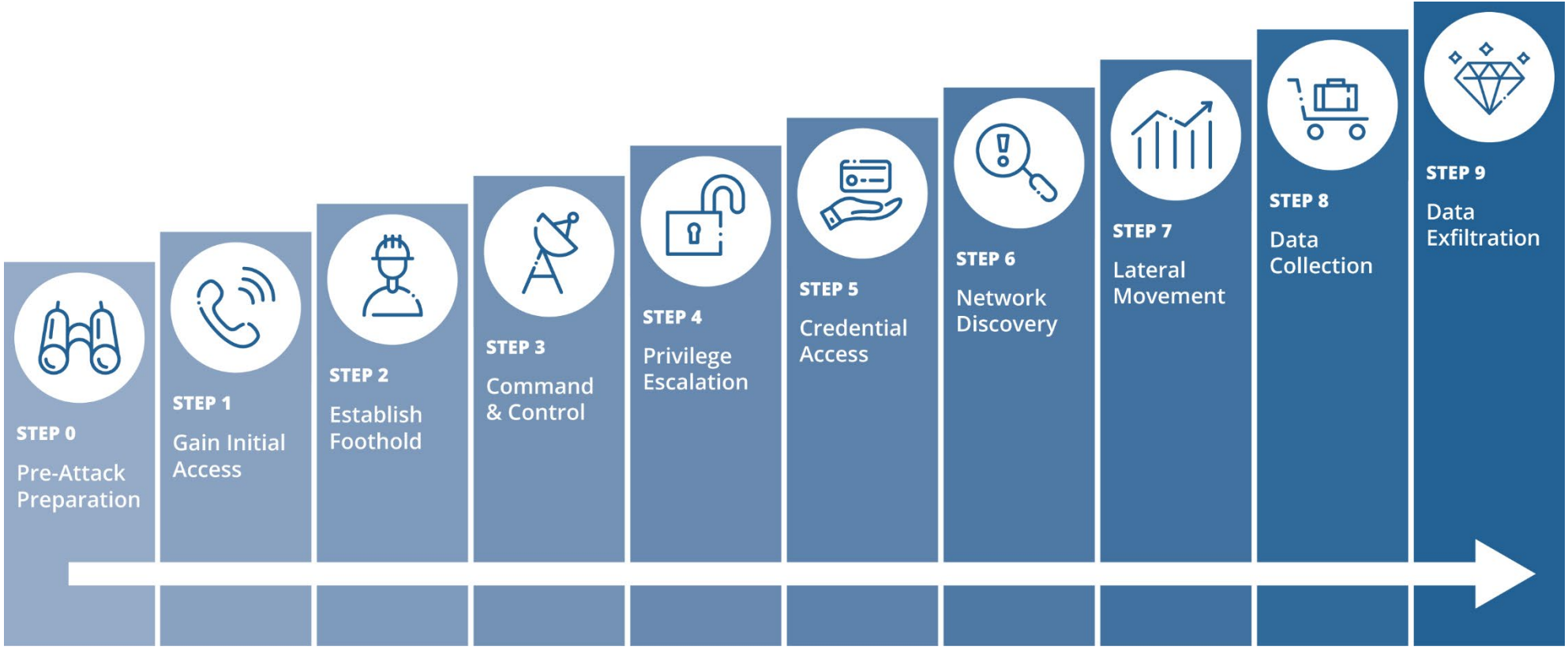


Simulating Realistic User Behavior

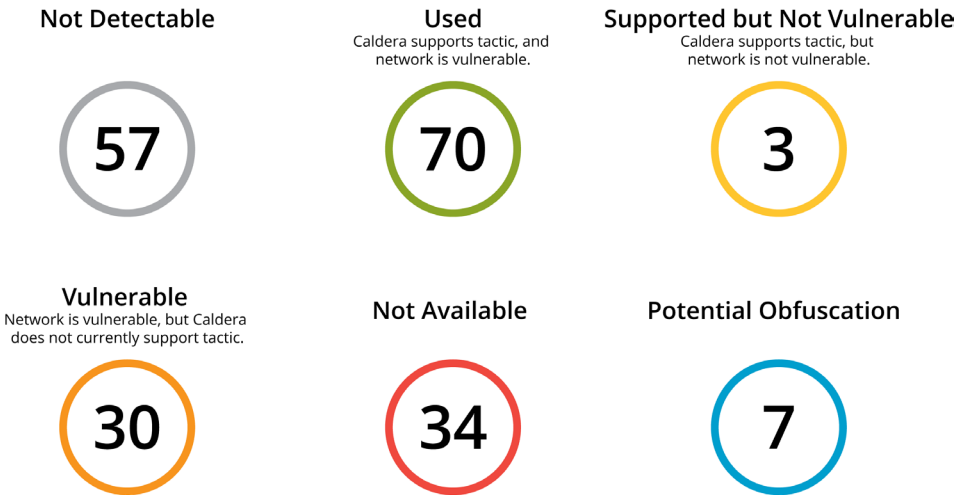


- 99 employees split across 5 divisions
- We provided a unique behavior for each user.
  - customized work schedules
  - role-specific work tasks
  - hobbies that influence personal use
- We set privileges and access by role.
- We used SEI GHOSTS software to simulate user behavior.
- Traffic results from the simulated behavior–it is not simulated directly!

Cyber Attack Exemplar



Cyber Attack Test Coverage



- We have mapped our cyber attack test suite to the MITRE ATT&CK framework.
- Not all attacks with the ATT&CK framework are detectable by the types of AI defenses we consider.
- A total of 70 techniques are covered so far in our methodology.

The Need for Continual Evaluation

- AI defense designers intend for defenses to constantly learn from the traffic they observe.
- As a result, the decisions that a defense makes will change over time.
- Therefore, evaluating a defense at a single point in time is insufficient.
- Our methodology calls for continual evaluation of the defense to gauge how capability may change over time.
- Note that tooling is required to support this continual evaluation.

Adversarial Manipulation: Obfuscation

- Adversaries may attempt to evade detection by obfuscating their attacks.
- changing tactics
  - use a different tool to perform the same tactic to avoid behavioral characterization
  - perform a different lateral movement path to gain access to the target
- modifying details
  - change order of events
  - change protocols used
  - change rate of tasks or time of day
  - employ encryption

Considerations:

- The adversary must still accomplish their original objective (i.e., the attack must still succeed).
- The attack may take longer to succeed due to rate limitations or circuitous path.
- There are a limited amount of resources for testing:
  - bandwidth
  - computing resources
  - testing window

Adversarial Manipulation: Data Poisoning

- An adversary causes the system to learn the wrong thing through data poisoning.
  - Adversaries provide valid and seemingly benign traffic to adjust the decision boundary of the AI.
  - Valid traffic is similar to expected attack traffic.
  - For example, suppose the AI is attentive to the volume of outbound traffic sent from an internal host.
    - Poisoning may consist of sending out a high volume of benign traffic from that host (e.g., e-mails with attachments, web traffic).
    - Exfiltration of a high volume of data from the host is more likely to succeed after poisoning.

Considerations:

- The addition of the poisoned traffic may itself set off alarms, so adversaries must take care to avoid detection.
- An adversary must have an idea of what a model uses for training to gauge what poisoning may be effective.

Agenda

- Motivation
- Problem and approach
- **Methodology**
- Wrapping up

Companion Project: Methodology Validation

- Adversaries carry out baseline attacks, attacks with obfuscation, and attacks with data poisoning against real systems.
- Our goals are to
  - show our methodology is effective.
  - measure how fragile systems are to various types of attacks.
  - determine what an assessment would reveal.
  - determine how much effort and length of time needed to perform an assessment.
  - identify limitations to our methodology.
  - develop a tooling suite to support practical testing.

Agenda

- Motivation
- Problem and approach
- Methodology
- **Wrapping up**

Conclusion: Testing Is Key for AI Defenses

- Effective use of AI-powered defensive operator tools requires an ability to evaluate their capabilities.
- Testing of AI defenses differs from testing of traditional cybersecurity devices.
  - Organizations must test AI defenses in a realistic network environment with realistic traffic.
  - AI defense designers intend for their defenses to change their behavior over time, necessitating continual evaluation.
  - Defenses are subject to adversarial manipulation via obfuscation and data poisoning.

- Our methodology tests the capability of AI defenses against actual cyberattacks, providing organizations with insight into defensive capability.
  - We map the cyber attacks we consider against the MITRE ATT&CK framework.

Future Directions

- consideration of different types of AI defenses, such as endpoint defenses and cloud-based defenses
- expansion of testbed infrastructure to include non-enterprise networks
- development of a methodology that employs the replay of traffic collected from a specific network of interest

We are actively seeking collaborators and transition partners for current and future work. Please contact us if you are interested!

The Team

Shing-hon Lau
Senior Cybersecurity Engineer

Grant Deffenbaugh
Senior Security Researcher

Lyndsi Hughes
Systems Engineer

Jarrett Booz
Associate Cybersecurity Engineer

Ken Brown
Assistant Cybersecurity Engineer

Brandon Marzik
Associate Cybersecurity Engineer

Derrick Spooner
MTS - Senior Engineer