Defensive Deception in Enterprise Network

Mu Zhu

North Carolina State University
mzhu5@ncsu.edu

Oral Prelim; August 20, 2021

Contents

Introduction

Defensive Deception Technologies

Defensive Deception leverages **false information** to confuse, mislead, or lure the attacker.

Defensive Deception VS. Traditional Defensive Technologies

- Traditional cybersecurity: focuses on attacker actions
- Defensive deception: focuses on anticipating such actions

Conceptual Deception Categories

Mimicking, inventing, decoying . . .

Objectives: Asset protection; Attack detection

Benefits and Limitations of Deception

Advantages:

- Cost-effective security scheme
- In-depth understanding threats by participating attack processing
- High deployability

Disadvantages:

- Overhead
- Disturbing legitimate user

Main Concerns

Defensive Deception Techniques

Honeyfile: Crafted decoy documents

- Benefits:
  - Simple deployment and maintaining
  - Effective detecting stealthy attack (e.g., insider attack)
- Limitations:
  - Unnecessary overhead of storage
  - Confusing legitimate user
  - Generating false positive alarm, which disturbs the defender

High-interaction honeypot: Fake host for luring attackers

- Benefits:
  - Sophisticated and difficult to be detected by attackers
  - Can include false information (e.g., honeyfile)
- Limitation:
  - High cost

Main Concerns

Threat: Insider Attacks

- Traitors, who misuse their legitimate credentials; know a lot about the victim's information
- Masqueraders, who impersonate a legitimate user: know little about where the victim's valuable information reside

Difference: Knowledge about victim, such as file space

Main Concerns

Threat: Advanced Persistent Threats (APTs)

Meaning: Well-trained attackers who perform multiple-year threats to exfiltrate valuable and sensitive economic, proprietary, or national security information

Cyber-kill chain: Reconnaissance, Delivery, Initial intrusion, Command and control, Lateral movement, Data exfiltration

Considered action space in proposed work:

Reconnaissance: Gather information about the victim to decide whether attack or not.

Compromise: Penetrate a target device

Data Exfiltration: Harvest sensitive data and transfer them to outside (e.g., masqueraders)

Research Questions and Motivation

Improving Defensive Deception Techniques

Caring about legitimate users

- How should the defender increase the deception attraction to the attacker?
- How should the defender effectively allocate resources?
- How should the defender reduce the impact from deception methods on the legitimate users?

Mee: Adaptive Honeyfile System

How to Enhance the Current Honeyfile System

The defender can:

- Adjust the number of honeyfiles by risk assessment
- Differentiate honeyfile alarms

Why not:

- Analyze suspicious behaviors across the network
- Make decision based on risk level

Mee:

- Decentralized deployment: deploys honeyfiles as a way to detect suspected behaviors by any user
- Centralized control: analyzes suspicious behavior across the network to determine the number and placement of honeyfiles for each device

Threat Model: Masquerader

Assumptions about the attacker:

- Has knowledge of the users' roles, e.g., via reconnaissance
- Has ability to infiltrate any connected device
- Is unfamiliar with the file system on a compromised device
- Knows of the existence of honeyfile system, but cannot distinguish between honeyfiles and real files
- Has clear target device to search for valuable files

In one compromised device, the attacker may obtain three results:

Success: Viewing or transferring the valuable files

Failure: Not finding valuable files, i.e., wasted effort

Loss: The defender cleans or replaces the compromised device

Legitimate Users and Insider Attacker

Users:

- Familiar with file system, e.g., lower probability to touch honeyfiles
- Open, but no transfer or modify

Attackers:

- Unfamiliar with file system, e.g., higher probability to touch honeyfiles
- Open, modify, transfer honeyfiles
- Attacking devices with tendency

Sensitivity, Seriousness, and Risk

To assist Mee to choose actions

File sensitivity: How valuable a honeyfile looks like for both the adversary and a legitimate user

Action seriousness: How much of a security threat the action is

- Weak: Open or close a honeyfile
- Strong: Edit, transfer, or zip or tar

Group of hosts:

- Groups: Based on organizational roles
- Group risk level: Represents a group's security situation
- Update risk estimate: Proportional to file sensitivity and action seriousness

Mee Architecture

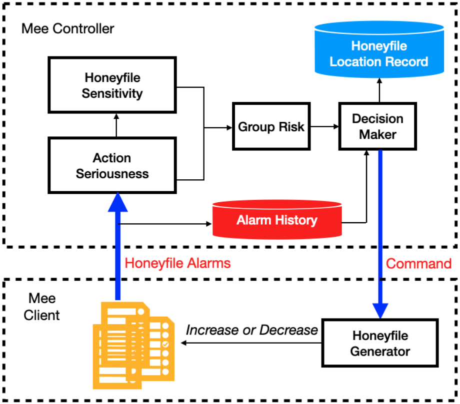
Decentralized deployment with centralized control

Mee Client:

- Generate and remove honeyfiles
- Detect file access on honeyfile and send alarms to Mee controller

Mee Controller:

- Analyze honeyfile alarms from Mee clients
- Instruct a Mee client to adjust the number of honeyfiles in its device



Group Risk Update and Classification

Group Risk Update

$$\triangle \text{risk}_{\text{group}}(\text{honeyfile}, \text{action}) = \frac{\text{sensitivity}_{\text{honeyfile}} * \text{seriousness}_{\text{action}}}{\text{number}_{\text{honeyfiles}}(\text{group})}$$

Group Classification

$$R_{-i} = \frac{\sum_{j \neq i} R_j}{Number\ of\ Groups - 1}$$

where R_{-i} represents the average group risk level except group i

$$Classification = \begin{cases} \text{Dangerous} & \text{if} \quad R_i > R_{-i} * 2 \\ \text{Medium} & \text{if} \quad R_{-i} < R_i < R_{-i} * 2 \\ \text{Safe} & \text{if} \quad R_i < R_{-i} \end{cases}$$

Scenario and Model

Defender Model

- Defender Action:
  - Check Device: Inform the Mee client to check the existed backdoor or update OS and application to avoid vulnerabilities
  - Increase Honeyfiles: Increase the number of honeyfiles in a device
  - Decrease Honeyfiles: Decrease the number of honeyfiles in a device
  - No Change: Idle to maintain current defensive strategy and save resources
- Defender Payoff:
  - Defence Cost: Cost to the defender when deploys an action
  - Fail in Protecting a Real File: Punishment of the defender when it fails in protecting real files
  - Impact to Regular User: Punishment of the defender when a legitimate user acts on a honeyfile

Scenario and Model

Attacker Model

- Attacker Action:
  - Infiltrate a Device
  - Read a file
  - Transfer/modify a file
  - Search, such as access a folder
- Attacker Payoff:
  - Effectiveness, such as the reward of accessing a real file
  - Action Cost
  - Impact of Failure

User Model

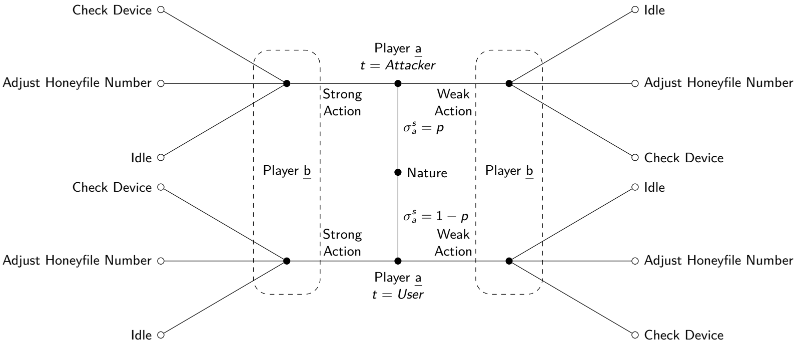
- User Action:
  - Login a device
  - Read a file
  - Transfer/modify a file
  - Search, such as access a folder

Honeyfile Game with Mee

1: From nature, Player a obtains type (attacker or user) as its private information

2: A honeyfile alarm represents an observation of the player b

3: The player b chooses an action based on a received message and its beliefs



Simulation and Evaluation

Test 1: Mee's performance

- Group risk level updating
- Number of honeyfiles in each group

Test 2: Comparison between Mee and traditional honeyfile system

- Tradition Honeyfile System: With different fixed number of honeyfiles in each device
- Mee: Dynamic number of honeyfiles in each device

Test 3: Comparison between Mee and traditional honeyfile system

- With different number of attackers

Metrics of Measurement:

- Defender Payoff
- Attacker Payoff
- Accuracy: True/false positive rate (ROC)

Mee v.2: Extension of the Honeyfile Research with DQL

Motivation

Limitation of Mee

- Simple scenario
- Game theory: Only two players at one time slot

Continue to have:

- Mee Structure: Controller and client
- Group and group risk level
- File sensitivity, action seriousness

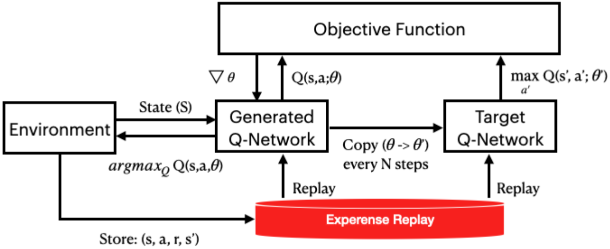
What is New?

- Complete scenario: More devices, active users and insider threats
- Deep reinforcement learning: Model multiple users at one time slot

Introduction: Deep Q Learning VS. Q learning

Agent (state, action space, observation); Environment; Reward function

- Neural Network: Using neural networks to approximate the Q-function
- Target Network: Employing a target network that delays the update of target values to increase learning stability
- Experience Replay: Sampling a random minibatch of transitions from experience replay buffer as training data

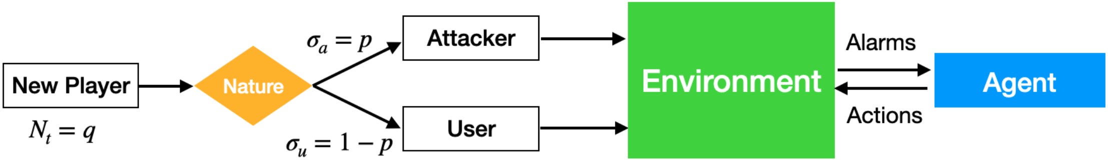


Environment

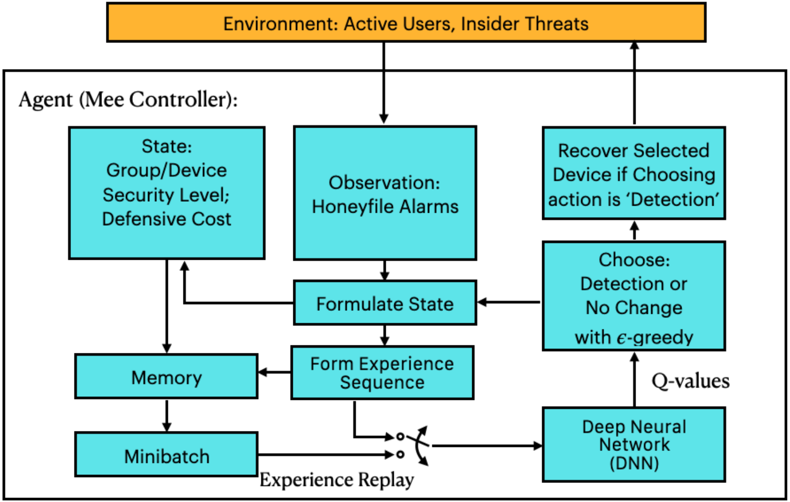
- Device: $\langle$Condition, Security Level, Importance, Groups$\rangle$
- Active User:
  - Action Space: Login, Search, Open a File, Edit a File
  - Being Familiar with File System
- Insider Attacker:
  - Action Space: Infiltration, Search, Open a File, Edit a File
  - No Knowledge of File System

Agent: Mee v.2 Controller

- Observation: Honeyfile alarms
- Action Space: ⟨Detection; No change⟩
- Utility: Effectiveness, Defence cost; Failure in protecting real file, Wrong detection



Mee v.2: DQN-Based Honeyfile System



GAN Based Honey Traffic Generation for Passive Monitoring

Threat Model

Passive Monitoring

Assumption

- Vantage point, such as compromised switches
- Scanning traffic and analyzing packets, e.g., packets sniffing and banner grabbing
- Searching for potential targets

Objective:

- Collect information through passive monitoring
- Compromise valuable devices
- Avoid to attack honeypots

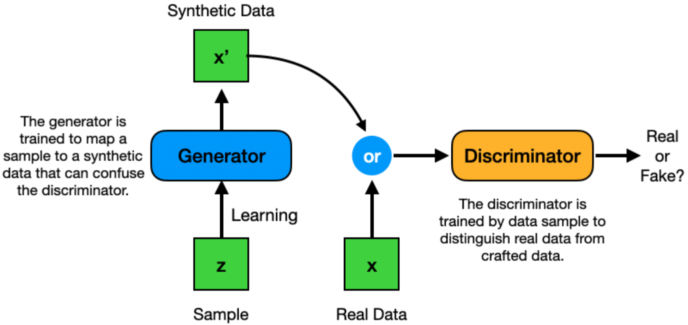
Deception Scheme

- High-interaction honeypot:
  - Includes vulnerable OS and applications
  - Mimics actual hosts
- Honey traffic:
  - Crafted TCP-based network flows
  - Transfer between honeypots

Introduction: Generative Adversary Networks (GANs)

Generator is trained to map from a latent space to a data distribution

Discriminator distinguishes candidates produced by the generator from the true data distribution



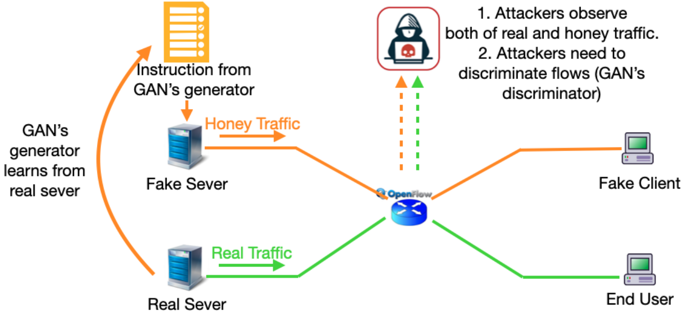
GAN-Based Honey Traffic Generation

Generator (Honeypot)

- Learning from actual server
- Generating craft fake traffic

Discriminator (Attacker)

- Learning from actual server
- Distinguishing real data from fake traffic
- Selecting device to compromise



Data Set and Features

Dataset: CIDDS-001 (includes flow-based network packets represented with network attributes)

Attribute	Type	Example
data first seen	timestamp	2018-03-13
duration	continuous	0.12
transport protocol	categorical	TCP
source IP address	categorical	192.168.100.5
source port	categorical	52128
destination IP address	categorical	8.8.8.8
destination IP port	categorical	80
bytes	numeric	2391
packets	numeric	12
TCP flags	binar/categorical	.A..S.

Conclusion

Summary of Works

Table: Time line for research approach

Schedule	Project
Complete	A Survey of Defensive Deception: Approaches Using Game Theory and Machine Learning
Complete	Mee: Adaptable Honeyfile System Based on Bayesian game
November 2021	Mee v.2: Deep Reinforcement Learning-Based Adaptive Honeyfile System
January 2022	GAN Based Honey Traffic Generation for Passive Monitoring

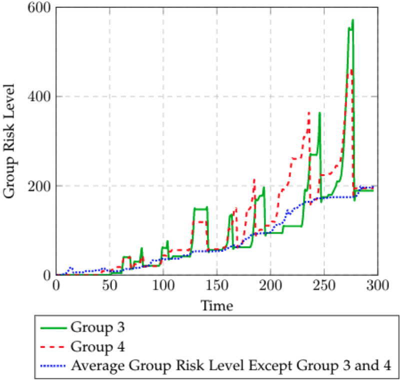
Plan of Works

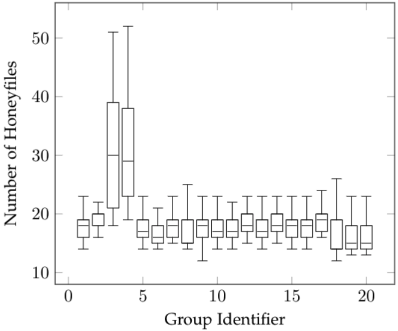
- Mee v2.: Implement simulation and finish evaluation (e.g., DQL and testbed)
- Honey traffic: Increse the complexity of testbed (e.g., involve more hosts)
- Hypergame based honeypot selection problem

Appendix

Test 1: Mee's Performances

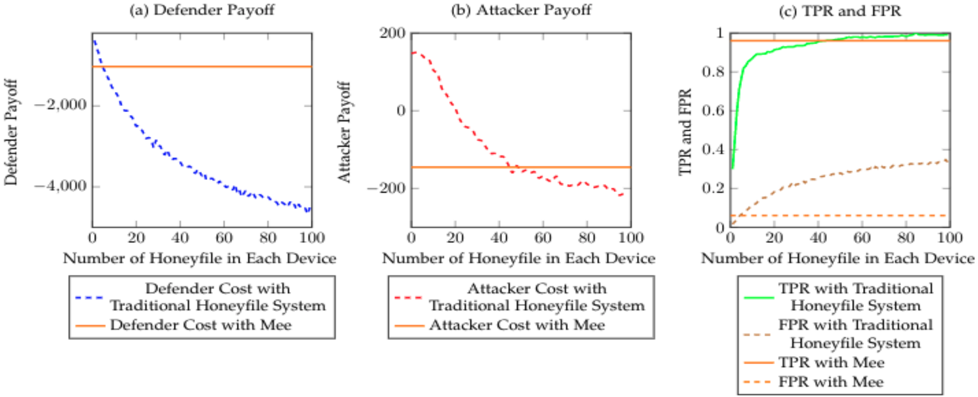
- Mee seeks to optimize resources while reducing false positives
  - Maintains group risk level
  - Adjusts the numbers of honeyfiles in various devices accordingly





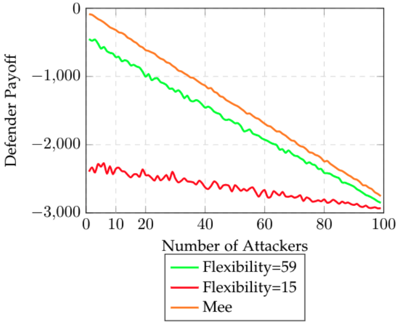
Test 2: Comparison between Mee and traditional honeyfile system

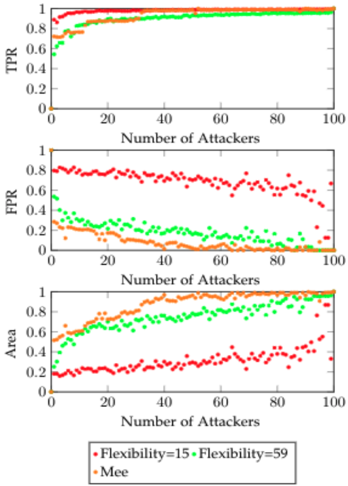
- Traditional Honeyfile System: the number of honeyfiles in one device is change from 0 to 100



Test 3: Comparison between Mee and traditional honeyfile system

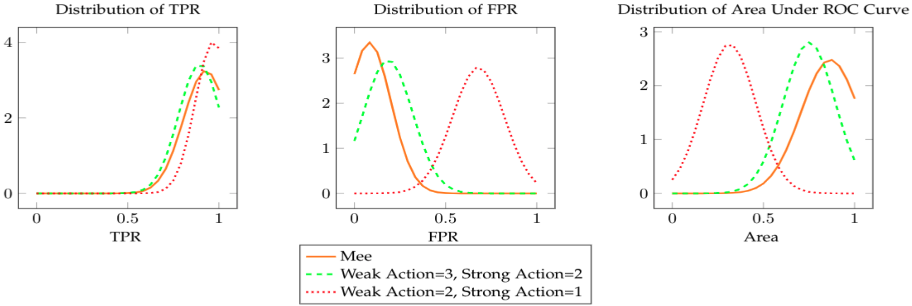
- Number of attackers is changed from 1 to 100
- Area under ROC Curve $= \text{TPR} * (1 - \text{FPR})$





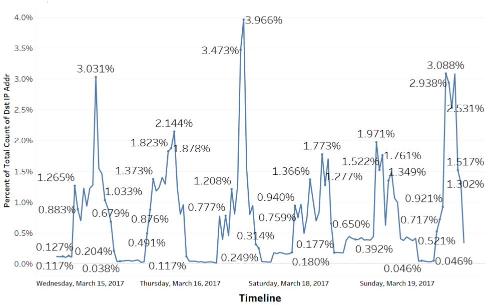
Detection Improvement: Effect Size

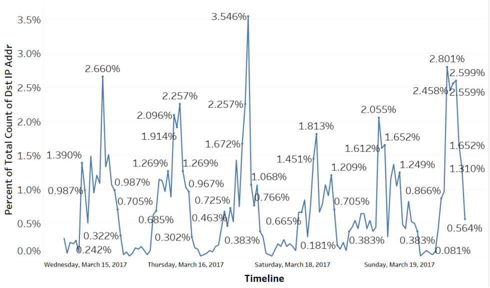
Cohen's d values for stated pairs	True Positive Rate	False Positive Rate	Area
(Weak Action = 2, Strong Action = 1) and (Weak Action = 3, Strong Action = 2)	0.28	3.57	3.05
(Weak Action = 2, Strong Action = 1) and Mee	0.38	4.55	3.62
(Weak Action = 3, Strong Action = 2) and Mee	0.70	0.80	0.76



GAN-based Honey Traffic Generation

Real (Right) and Generated (left) Network Flow





Acknowledge

The Army Research Office supported this work under award W911NF-17-1-0370. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes without standing any copyright notation.

Thank You

Mu Zhu (mzhu5@ncsu.edu)