

## 6. Attacker behavioral profiling in ssh honeypots

[silvio.russo3@unibo.it](mailto:silvio.russo3@unibo.it)

### Research Question

Can machine learning models accurately classify attackers into distinct behavioral profiles (automated bots, script kiddies, skilled operators) based on their command sequences and interaction patterns in SSH honeypots?

### Introduction

SSH honeypots collect vast amounts of attack data, but most analysis focuses on aggregate statistics rather than individual attacker behavior. This project's goal is to develop a classification system to automatically profile attackers based on interaction patterns: command diversity, timing, tool signatures, and reconnaissance depth. By distinguishing automated bots from skilled operators, defenders can prioritize responses appropriately.

The study extracts behavioral features from Cowrie honeypot logs and trains multiple classifiers on labeled datasets. Research investigates which features best discriminate against attacker types and whether automated classification matches expert analysis. Results provide actionable intelligence for security operations: automated bot attacks may warrant IP blocking, while sophisticated attacker sessions might trigger incident response escalation.

### Implementation Guidance

- Extract temporal features: inter-command timing, session duration, time-of-day patterns
- Command-based features: unique commands ratio, command diversity, tool signatures
- Behavioral patterns: reconnaissance vs. exploitation ratio, error rate, command correction attempts

### Public Datasets

1. CyberLab Honeynet Dataset
  - a. Link: <https://zenodo.org/records/3687527>
  - b. Description: Large-scale dataset (9 months, 50 nodes) with diverse attacker behaviors
  - c. Format: JSON logs with timestamps, commands, session metadata
  - d. Best for: Training classifiers on comprehensive feature sets
2. IEEE DataPort - SIHD Dataset
  - a. Link: <https://ieee-dataport.org/documents/sihd-smart-industrial-honeypot-dataset>
  - b. Description: Multi-region honeypot logs from 6 geographic locations
  - c. Format: .log files with full session details
  - d. Best for: Geographic analysis of attacker profiles

### Key References

- Nawrocki, M., et al. (2016). "A Survey on Honeypot Software and Data Analysis." *arXiv:1608.06249*. [Honeypot data analysis fundamentals]
- Alata, E., et al. (2006). "Lessons Learned from High-Interaction Honeypot Deployment." *EDCC*. [Real-world attack pattern analysis]
- Owens, J., & Matthews, J. (2008). "A Study of Passwords and Methods Used in Brute-Force SSH Attacks." *USENIX LISA*. [SSH attack characterization]

## Comprensione e analisi delle consegne:

TITOLO: profilazione comportamentale degli attaccanti in honeypot SSH

è possibile usare modelli di machine learning per classificare automaticamente diversi tipi di attacki in base ai loro comportamenti nei log di un honeypot SSH?

Un SSH HONEY POT è un sistema "trappola" che simula un server vulnerabile SSH per raccogliere dati sugli atacchi.

Solitamente si analizzano solo statistiche generali (quanti Stechi, da dove, ecc..)

Noi dobbiamo studiare il comportamento individuale dell'attaccante per creare un sistema di profilazione automatica dell'attaccante così da pensare una strategia di difesa migliore.

Bisogna osservare:

- varie tipi di stacco
  - temporistiche
  - forme dei tool usati in stacco
  - quando l'attaccante esplora il sistema

Guide all'implementazione fornite:

1. estrazione delle features da estrarre: - intervallo tra i comandi  
- durata della sessione  
- orari di attività

- rapporto tra comandi unici e totali  
- diversità dei comandi  
- forme di tool moti

- tempo di riconoscimento / tempo esplorazione  
- tasso di errore o tentativi falliti  
- quante volte corregge comandi sbagliati

temporali } basati su comandi } basati sul comportamento

Cose ci hanno fornito?

2 dataset per l'addestramento AI.

- Zemodo: 9 mesi di log de 50 sistemi honeypot.

Sono in formato JSON con timestamp, comandi e metadati

- IEEE: log di honeypot in 6 località diverse

utimo per analisi geografiche dei comportamenti

## Cose dobbiamo fare?

→ raccolte log di attacchi SSH

→ estrazione features comportamentali

→ ricorda o classifica i tipi di attacco → 3 profili principali:

1. automated bots: attacchi automatici e ripetitivi

2. Script Kiddies: attaccanti imprevedibili che usano script o tool già pr

3. Skilled operators: attaccanti esperti che agiscono manualmente e

4. altri...

4. addestrare e testare modelli ML

5. analisi risultati

6. relazione sui risultati ottenuti

## L'AI? La dobbiamo fare noi?

sí.

Creare e addestrare, ma Python ha librerie che ci aiutano di creare una da 0.

1. scikit-learn

2. XGBoost o LightGBM

3. TensorFlow / PyTorch (rete neurale)

Io direi di contattare il tutor: silvio.russo3@unibo.it

