6. Attacker behavioral profiling in ssh honeypots

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]