HACK SECURE

Hack Secure 10 Days Campaign

Final Project/Task Submission

(Detailed explanation of your work, research, or findings.)

Title: AI-Driven Log Analysis & Threat Detection

1. Overview

This project focuses on detecting suspicious activity in web server logs using a combination of rule-based detection and AI-powered anomaly detection. The goal was to build a simple but effective threat detection pipeline that can be extended for real-world security monitoring.

2. Work & Research Done

Step 1 – Research

- Studied common log-based threats: directory brute-force, bots, and enumeration attacks.
- Reviewed Apache log formats and learned how to parse them using regex.
- Explored user-agent-based bot detection techniques.
- Researched AI methods for anomaly detection and selected Isolation Forest (an unsupervised machine learning model).

Step 2 – Implementation

I created three detection scripts as part of the project:

Part A – Directory Enumeration Detection

- Parsed Apache/Nginx logs to extract IPs and HTTP status codes.
- Counted the number of 404 errors per IP.
- Flagged IPs with ≥10 404s as possible enumeration attacks.

Part B – Bot vs Human Detection

- Parsed logs to extract IP, User-Agent, and Referrer.
- Classified traffic as bot if User-Agent contained terms like curl, wget, python-requests, or if Referrer was empty.
- Produced a table with IP, requests, and classification results.

Part C – AI-Based Anomaly Detection

- Extracted features: number of requests, errors, and unique URLs per IP.
- Used Isolation Forest to flag anomalous IPs as suspicious.
- Produced a table showing normal vs suspicious IPs.

Step 3 – Testing

- Created a sample access.log file with normal and suspicious entries.
- Ran all three scripts and confirmed that:
    - The directory enumeration script flagged high 404 IPs.
    - Bot detector classified curl/python-requests traffic as bot.
    - The AI model correctly detected anomalous IPs with unusually high requests or errors.

3. Findings

Part	Key Findings
Directory Enumeration	Multiple 404 errors from a single IP address indicate a likely directory brute-force attack.
Bot Detection	User-Agent analysis is effective in catching basic bots.
AI Anomaly Detection	Isolation Forest detects unusual IP behavior even if thresholds are not pre-defined, making it useful for unknown attack patterns.

4. Conclusion

This project demonstrates that log analysis combined with AI provides a powerful way to detect threats:

- Rule-based detection identifies known patterns, such as 404 floods and bots.
- AI-based detection identifies new, unseen anomalies.
- Together, they provide a strong early-warning system for suspicious activity.

5. Future Scope

- Real-time log monitoring and alerting.
- Integration with SIEM tools like Splunk or ELK Stack.
- Adding GeoIP lookup to identify attacker locations.
- Building dashboards for visualization.