Log File Analyzer for Intrusion Detection

Abstract

This project focuses on analyzing system log files, particularly SSH authentication logs, to detect signs of brute-force attacks. Using Python, pandas, and regular expressions, the tool parses the log entries, extracts failed login attempts, identifies suspicious IP addresses, and visualizes the frequency of these attacks using bar graphs.

Introduction

Intrusion detection plays a vital role in cybersecurity. By analyzing log files generated by systems and services, security analysts can identify unauthorized or suspicious behavior. This project aims to automate the detection of brute-force login attempts by examining SSH logs.

Tools Used

- Python 3
- pandas
- matplotlib
- re (regex)
- Ubuntu Linux environment

Steps Involved in Building the Project

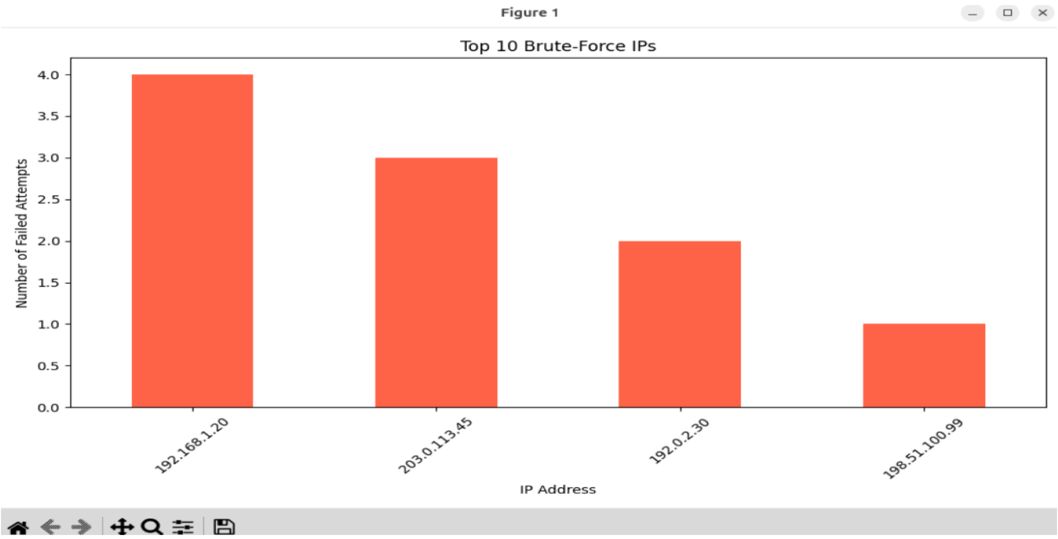
1. Setup the development environment using Python and virtual environment.
2. Collected SSH log data from /var/log/auth.log or a sample dataset.
3. Parsed log lines using regular expressions to extract failed login attempts and IPs.
4. Used pandas to analyze and count the number of failed attempts per IP.
5. Visualized the data using matplotlib by plotting the top 10 brute-force IPs.
6. Exported the report as a CSV file and generated a bar graph as PNG.
7. Parsed Apache logs (access.log) to identify suspicious request patterns such as repeated 404 errors and user-agent anomalies.

8. Detected potential scanning and DoS patterns based on high-frequency requests from single IPs within a short time.
9. Cross-referenced IP addresses with public blacklists (e.g., AbuseIPDB) to flag known malicious actors.

The analyzer also detected scanning and potential DoS behavior based on traffic patterns.
IP addresses were cross-checked with public blacklists to identify known malicious entities.
Apache access logs were analyzed to detect abnormal request spikes and suspicious activity.

Output & Observation

The tool successfully identified multiple IP addresses attempting unauthorized access via SSH. It generated a visual report of the top offenders based on failed login attempts. This kind of analysis is crucial for proactive threat monitoring in a SOC environment.



Conclusion

This project demonstrates how log analysis can be automated using simple Python tools. By analyzing authentication logs, we can detect brute-force attacks early and respond quickly. The solution is lightweight, extensible, and ideal for entry-level SOC use cases.