Intrusion Detection System based on Raspberry Pi Honeypot and Machine Learning for Home Network

Phakonekham Phichit, ID 001041931

BSc H COMPUTER SCI (EXT-UGIC)

University of Greenwich Computing & Mathematical Sci.

Contact Information:
Computing & Mathematical Sci
University of Greenwich Old Royal Naval College
Park Row, Greenwich London SE10 9LS United Kingdom

Email: phakonekham.phichit@gmail.com

Abstract

This study presents a research honeypot intrusion detection system based on a Raspberry Pi 4B device, a cloud-based ELK stack system, and machine learning classification algorithms for SSH and Telnet services in a home network. The system aims to detect unauthorized access attempts by creating a decoy environment that emulates legitimate services. The project aims to improve the implementation by Jeremiah (2019).

Introduction

This project that aims to develop a honeypot intrusion detection system for home networks. The system utilizes a Raspberry Pi 4B device, a cloud-based ELK Stack system, and machine learning classification algorithms to detect potential threats and unauthorized access attempts. The honeypot technology creates a decoy environment to collect valuable data, and machine learning techniques improve intrusion detection capabilities by identifying patterns of malicious activity. The ELK Stack system provides real-time monitoring and data visualization, while Discord alerts notify users of potential security breaches. Overall, the project seeks to create a comprehensive solution for detecting and mitigating cyber-attacks and enhancing the security of home networks.

Project Idea & Objectives

This study presents a research honeypot intrusion detection system based on a Raspberry Pi 4B device, a cloud-based ELK stack system, and machine learning classification algorithms for SSH and Telnet services in a home network. The system aims to detect unauthorized access attempts by creating a decoy environment that emulates legitimate services. The project aims to improve the implementation of Jeremiah (2019) by creating data visualization, alert system and classification machine learning model solutions.

1. Home Network Raspberry Pi Cowrie Honeypot
2. Intrusion Detection System with an Alert System
3. Provide Real-Time Data Visualization to analyse Honeypot Data
4. Implement classification machine learning algorithm models and identify severe and non-severe attacks from the honeypot.

Methodology

The honeypot system is built on a Raspberry Pi running the Cowrie honeypot, integrated with a cloud-based ELK Stack server for data visualization. Deployed within a demilitarized zone (DMZ) on a home router, the system is designed to attract and monitor potential attackers, as per the network topology illustrated in Figure 1. The honeypot is deployed from April 7th to April 20th.

The combination of Raspberry Pi and selected software tools provide a cost-effective solution for home network security. The Cowrie honeypot can emulate SSH and Telnet services to capture malicious activity, while the ELK Stack offers comprehensive log management, data processing, and visualization capabilities. Filebeat and Packetbeat further ensure efficient log and network data collection, and Nginx serves as a secure web server for easy access to the Kibana dashboard.

ElastAlert is a tool that can monitor data streams and generate alerts based on custom rules, making it ideal for intrusion detection within the Elasticsearch environment. When integrated with Discord, it enables real-time notifications on multiple devices, facilitating immediate responses to potential threats.

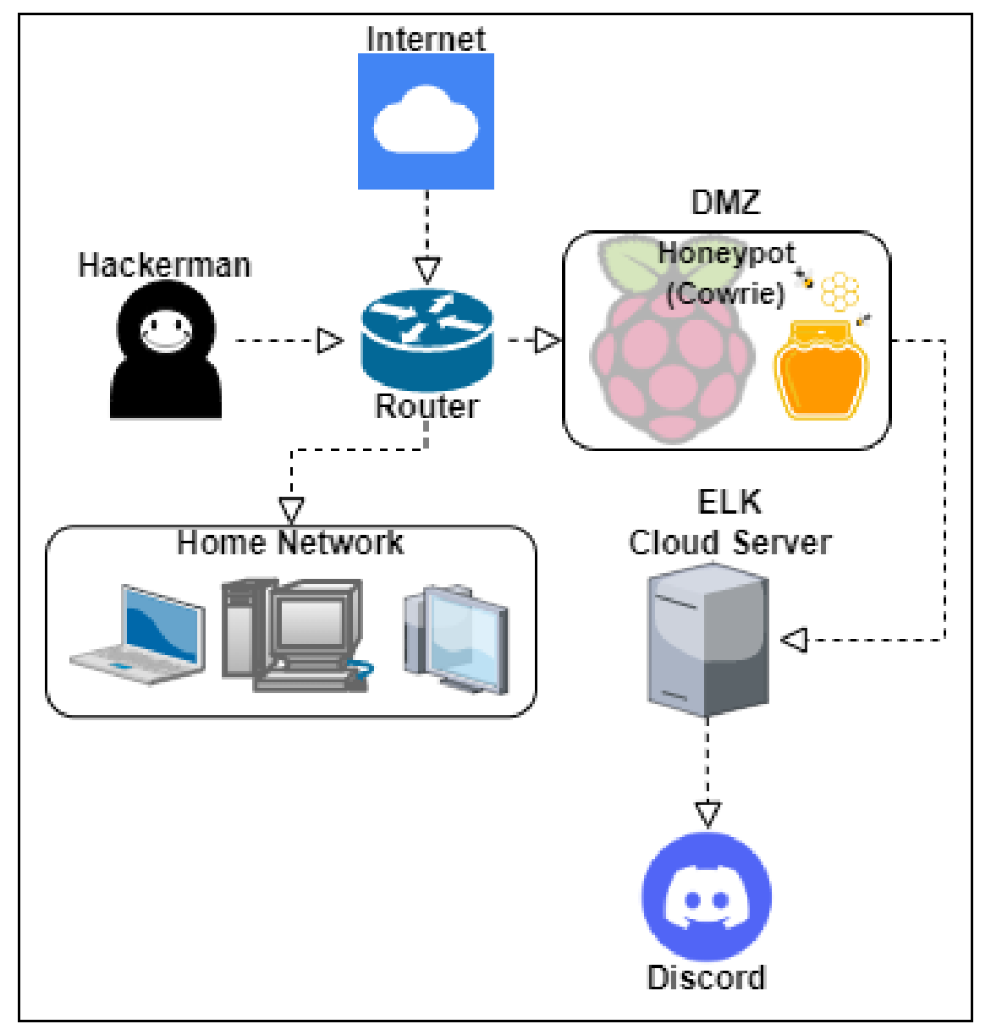

Figure 1: Honeypot Network Topology

The pre-processing pipeline for the honeypot system involves data collection, cleaning, feature selection, encoding, and partitioning. The system collects data on unauthorized access attempts, which is then cleaned to ensure its quality. Relevant features are selected for machine learning classification algorithms. Since the project's dataset includes categorical features like source IP and country, data encoding is performed to convert these into numerical values. The pre-processed data is then split into training and testing sets.

The classification target feature classifies potential malicious attacks into non-severe and severe attacks. Severe attacks are those where the attacker logs in and executes Unix commands. Non-severe attacks either involve a successful login without Unix command execution or a failed login attempt. The nature of the attack is determined by the behaviour following a successful login.

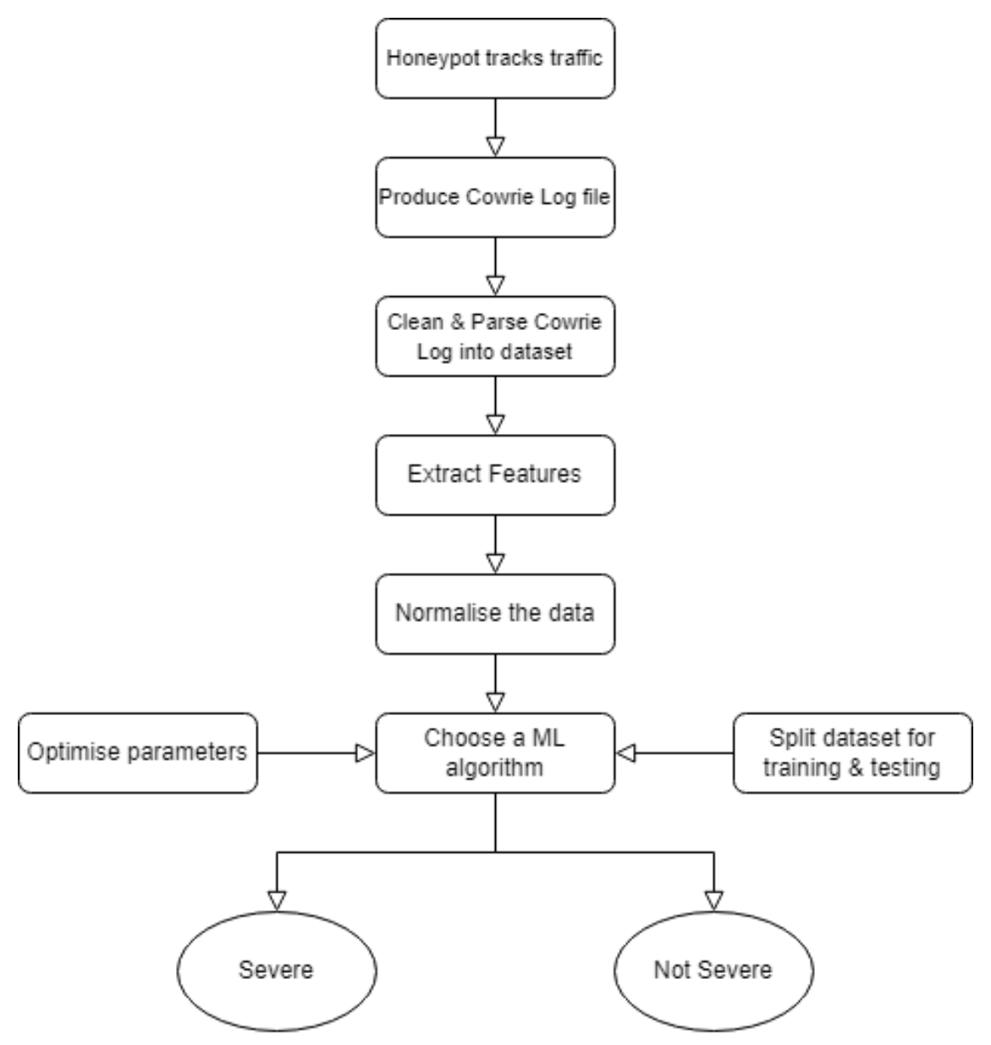

Figure 2: Classification Machine Learning Design and Development

Results

Over a two-week deployment period, the Raspberry Pi honeypot recorded 37,768 login attempts, with 19,163 successfully gaining unauthorized access, and 18,605 failing. These attempts originated from 74 different countries, underscoring the widespread issue of unauthorized access and the necessity for robust intrusion detection systems. Among these attempts, 54.74% targeted the Telnet protocol, while 45.26% were aimed at the SSH protocol.

Table 1 shows the most common username and password used in attempts to gain access to the Raspberry Pi honeypot.

Table 1: Top Usernames and Passwords

Rank	Username	Count	Password	Count
1	root	29,201	admin	578
2	aaliyah	3,013	1234	395
3	admin	1,802	password	380
4	guest	236	123456	372
5	user	183	root	253
6	supervisor	139	12345	232
7	ubnt	105	user	160
8	pi	102	123456789	150
9	support	99	7ujMko0admin	143

The Raspberry Pi honeypot system, as shown in Table 2, successfully detects SSH, Telnet, and DDOS attacks on ports 22 and 23, as well as port scanning from Nmap. Alerts for these attacks are efficiently relayed through a Discord alert system. However, the system displays delayed responses when faced with Brute Force attacks on both ports. This also includes the Raspberry Pi performance evaluation during each attack.

Attack Cases	Attack Detection	CPU	Memory	Temp(°C)
Port Scanning Nmap	Yes	2.30%	24.50%	58.4
Brute Force (Port 22)	Delay	3.3%	23.3%	58
Brute Force (Port 23)	Delay	4.7%	24.1%	58
DDOS Attack (Port 22)	Yes	45.60%	27.70%	60
DDOS Attack (Port 23)	Yes	40.60%	24.50%	60

Table 2: Raspberry Pi Test Cases & Performance Evaluation Results

Table 3 presents the classification performance metric results for three different machine learning models: K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM). The performance of each model is assessed using various metrics, such as accuracy, recall, precision, F1 score, and training duration.

Table 3: Classification Performance Metric Results

Model	Accuracy	Recall	Precision	F1 Score	Training Duration
KNN	0.9447	0.8838	0.9678	0.9239	23 mins
RF	0.9422	0.8839	0.9606	0.9207	3 mins
SVM	0.9459	0.8791	0.9760	0.9251	2h20 mins

Evaluation

The project successfully met fundamental requirements, implementing core features and enhancing the work of previous research, specifically by Jeremiah (2019). However, there were occasional delays in brute force notifications due to potential Filebeat configuration issues. The final product successfully visualized real-time data, collected and extracted data, and applied machine learning models to classify SSH and Telnet attacks. Limitations include a lack of real-time model implementation and the use of a single type of honeypot, limiting the diversity of attack data. External issues, such as ISP and power interruptions, affected the home network implementation. Further, while the classification models worked adequately, they could benefit from additional raw data features. Despite these challenges, the project provided a more comprehensive data visualization and alert system than comparable models, offering a solid foundation for future work.

Conclusions

This project aimed to develop an intrusion detection system using a Raspberry Pi honeypot, applying machine learning to identify cyberattacks. Despite some challenges with the alert system, the project successfully achieved most of its goals. It showcased the effectiveness of Raspberry Pi-based honeypot in detecting cyber threats on home networks and the potency of machine learning algorithms, like Support Vector Machines, K-Nearest Neighbors, and Random Forest, in classifying these threats. This work contributes to the field of honeypot intrusion detection systems, demonstrating their potential in-home network security. This also includes Real-time data visualisation of the honeypot traffic through ELK stack cloud server.

Limitations & Further Work

Several limitations were encountered during the development of this project. One primary issue was the inconsistent delay times in the brute force notifications provided by the Discord alert system. This might be due to challenges with the configuration of Filebeat, which may have failed to accurately identify specific elements from the Cowrie JSON log files. This flaw could potentially affect the real-time detection of repeated failed login attempts, a key characteristic of brute force attacks.

The project was also restricted in terms of the diversity of attack types and protocols it could analyze, as it only implemented a single type of honeypot. This limitation may have prevented the system from capturing a broader range of malicious activities, thereby reducing its overall effectiveness as an intrusion detection system. The machine learning models developed in this project weren't deployed in real-time, which meant they couldn't provide immediate threat detection and response.

Future directions include exploring other machine learning algorithms, expanding feature sets for detecting diverse attack types, and implementing real-time intrusion detection. Testing the system with larger, varied datasets and developing user-friendly interfaces for non-expert users could further enhance its effectiveness and usability. The project emphasized the importance of understanding attack techniques for better network protection and developing more efficient intrusion detection systems.