International Journal of Computer Science Trends and Technology (IJCST) – Volume 13 Issue 2, Mar-Apr 2025

RESEARCH ARTICLE OPEN ACCESS

Designing Dynamic Honeypots for Cyber Threats Detection

Ms. PRADHIKSHA S*, [2]Ms. C.VISHNU PRIYA**,

*(Department of Computer Science and Engineering, Dr. M.G.R. Educational and Research Institute, Chennai, India)

**(Assistant Professor (IDE). Centre for Cyber Forensics and InformationSecurity, University of Madras, Chepauk Chennai)

ABSTRACT

In Today's interconnected world, smart devices are frequently connected to public networks for various operations. These devices are highly exposed to publicly accessible networks, increasing the probability of cyberattacks. The Existing systems utilize machine learning techniques, such as the Bayesian Boost model, to analyze data and protect smart home applications from spam. However, to enhance the security of publicly connected PCs and other IoT devices, the proposed system leverages the CIC dataset for spam detection. The system employs a Random Forest algorithm to detect abnormal activities in public networks. By analyzing anomalies within the given dataset, it provides notifications regarding potential threats, such as fake applications and fraudulent job offers. The system's performance is evaluated based on key metrics, including accuracy,precision, and F1-score, ensuring effective and reliable spam detection in smart environments.

KEYWORDS: Public Networks, Machine Learning, CIC Dataset, Random Forest Algorithm, Spam Detection, Cyberattacks.

I. INTRODUCTION

The Paper 'DESIGNING DYNAMIC HONEYPOT FOR CYBER THREATS DETECTION' aims to enhance smart device security by combining the principles of honeypots with the power of machine learning, ensuring reliable protection against modern cyber threats in dynamic network environments. The rapid growth of smart devices and their integration into everyday life have significantly increased reliance on public networks for communication and operations. However, this widespread connectivity has also exposed these devices to various cybersecurity threats, including malware injections, phishing attemps, and unauthorized access. Protecting the network from miscellaneous activity is mandatory and that will save a huge data to be hacked from suspicious activity. [1]. Honeypots are an important part of a comprehensive cybersecurity strategy. The main objective of honeypots is to expose vulnerabilities in the existing system and also to provide accurate information about the vulnerabilities in the given dataset and also alerts information about emerging threats and attack methods. [2].

Traditional spam detection systems have predominantly relied on machine learning techniques like the Bayesian classifier to identify and filter malicious content . While effective in certain scenarios, these models often encounter challenges such as high false positive rates and limited adaptability to emerging threats . To address these limitations, this system proposes a spam detection system that leverages the CIC dataset and employs the Random Forest algorithm, renowned for its robustness and efficiency in classification tasks . [3]. The Random Forest algorithm is particularly adept at handling large datasets and detecting anomalous patterns in network traffic, making it suitable for real-time identification of suspicious activities . By analyzing network behavior, our system aims to detect potential threats such as fraudulent applications and phishing attempts promptly. To ensure the system's reliability and effectiveness, we evaluate its performance using key metrics, including accuracy, precision, and F1-score. [4].

This research contributes to the field of cybersecurity by introducing a robust spam detection mechanism tailored for devices connected to public networks. By integrating machine learning techniques with comprehensive datasets, we aim to enhance the security of smart environments and provide users with effective protection against evolving cyber threats. [5].

II. LITERATURE REVIEW

Inadyuti Dutt and et al., [6] had proposed Immune System Based Intrusion Detection System (IS-IDS) This paper explores the natural immune system to detect network intrusions. Their model consists of two layers Statistical Modeling-based Anomaly Detection (SMAD) and Adaptive Immune-based Anomaly Detection (AIAD**).** The system captures initial network traffic vulnerabilities and utilizes machine learning techniques to classify anomalous patterns effectively. The researchers demonstrated high detection accuracy using datasets like KDD99 and UNSW-NB15. Results show significant true positive rate, closer to almost 99% of accurately detecting the file based and user based anomalies for both the real time traffic and standard data sets.

Mohammed Tarek Abdelaziz and et al., [7] had proposed Enhancing Network Threats Detection with Random Forest-Based NIDS and Permutation Feature Importance. This paper explores the effectiveness of the Random Forest classifier in improving NIDS capabilities through machine learning-based detection methods. Using the CICIDS-2017 dataset, data preprocessing techniques were applied to remove redundancies and enhance data quality. Results indicate that optimizing class weights, applying a customized prediction function, and leveraging 26 key features allow the Random Forest classifier to achieve an impressive 99.8% weighted F1-score and 93.31% macro F1-score across multiple attack types

J.Franco and et al., [8] had proposed A Survey of Honeypots and Honeynets for Internet of Things, Industrial Internet of Things, and Cyber-Physical Systems This paper presents a comprehensive survey on honeypots and honeynets for IoT, IIoT, and CPS environments. It provides a taxonomy and an in-depth analysis of an existing honeypot and honeynet implementations, it discusses the key design considerations for modern honeypot research, and highlights open challenges that need to be addressed for future developments in securing IoT, IIoT, and CPS infrastructures.

Keskin and et al., [9] had proposed Machine Learning Based Classification for Spam Detection This paper uses five different machine learning algorithms — Random Forest (RF), Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), and Artificial Neural Network (ANN) — were applied to classify spam and non-spam emails. A dataset of 5,558 emails was used, and the models were evaluated based on accuracy, precision, sensitivity, and F1-score. Among these, the Random Forest algorithm achieved the highest accuracy of 98.83%, making it the most effective model for this task.The results of this study show that machine learning algorithms can accurately classify spam emails, offering a strong defense against email-based threats.

Wang Jiasheng and et al., [10] had proposed Research on dynamic honeynet technology based on machine learning. The research introduces a dynamic honeynet system enhanced with machine learning technique that uses the Random Forest algorithm to build a data detection module for accurately identifying anomalous traffic, and incorporates the DQN (Deep Q-Network) reinforcement learning algorithm to dynamically adjust defense strategies based on different types of attacks. Using the CIC-IDS-2017 dataset, experiments confirm that this intelligent honeynet design significantly improves ICS network security capabilities.

Commey and et al., [11] had proposed Dynamic Honeypot Conversion for Enhanced IoT Security. This paper explores the blockchain-enabled honeypots IoT Conversion system, a smart solution that dynamically transforms IoT nodes into honeypots based on threat levels using a lightweigth machine learning model.The results shoe the BHICS provides a 76.5% attack prevention rate, close to dedicated honeypot system(79.6%), while reducing node compromise rates from 49.8% to 22.3%. This approach offers a scalable and efficient IoT security solution that balances performance, resource usage, and reliable logging.

Diandra Amiruddin Firmansyah and et al., [12] had proposed Honeypot based threat detection using Machine learning techniques. This paper investigates the use of machine learning (ML) techniques to improve honeypot-based threat detection in cybersecurity.While honeypots generate valuable security insights, they also produce large volumes of data, which can be difficult for human analysts to manage manually. the study applies machine learning algorithms to automate data analysis and improve detection accuracy. The findings confirm that machine learning enhances the speed and accuracy of honeypot-based threat detection, allowing cybersecurity teams to respond more effectively to potential threats.

III. PROPOSED METHODOLOGY

The Research aim is to enhance the security of smart devices that are frequently connected to public networks. These devices, including PCs and other IoT-enabled systems, are often vulnerable in such environments. To address this, the study utilizes the CIC-IDS2017 dataset for spam and anomaly detection. A Random Forest algorithm is implemented to identify abnormal activities in public networks. By analyzing anomalies within the dataset, the system is capable of providing notifications regarding potential threats such as fake applications, fake news, and Android-defender. The performance of the proposed system is evaluated using key metrics, including accuracy, precision, and F1-score.

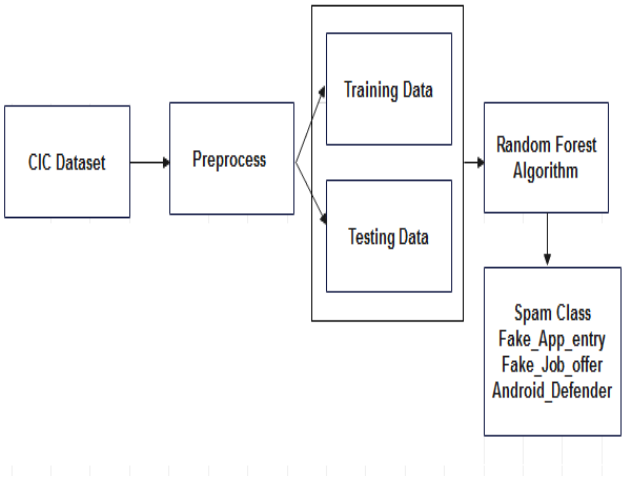

Fig 3.1.1 Research Design

3.1 METHODS

3.1.1 DATA COLLECTION

The Dataset utilized in this research is sourced from the Canadian Institute for Cybersecurity (CIC). This dataset serves as the foundation for training and evaluating the proposed intrusion detection system.

3.1.2 ABOUT DATASET

CIC DATASET:

- The CIC dataset is a comprehensive and widely adopted cybersecurity dataset curated by the Canadian Institute for Cybersecurity. [13] It is specifically designed for applications such as network intrusion detection and malware classification. The dataset incorporates a broad range of contemporary and realistic cyberattacks, closely simulating real-world scenarios through full packet capture (PCAP) files. [14].

- The CIC2017 Dataset includes detailed network traffic analysis generated using CICFlowMeter. Each network flow is labeled and enriched with metadata such as timestamps, source and destination IP addresses, source and destination ports, protocols, and attack categories. The structured format, typically in CSV files, makes it suitable for various machine learning tasks, including supervised classification mode.

	Destination Port	Protocol	Flow Duration	Total Fwd Packets	Total Backward Packets	Total Length of Fwd Packets	Total Length of Bwd Packets	Fwd Packet Length Max	Fwd Packet Length Min	Fwd Packet Length Mean	...
0	443	6	37198	1	1	31	0.0	31	31	31.0	...
1	443	6	36824	1	1	31	0.0	31	31	31.0	...
2	443	6	36728	1	2	31	0.0	31	31	31.0	...
3	443	6	37714	1	2	23	0.0	23	23	23.0	...
4	443	6	37598	1	1	23	0.0	23	23	23.0	...

5 rows × 80 columns

Fig 3.1.2 CIC Dataset

In the above Figure 3.1.2 each line in dataset represents a separate network flow and includes various features that describe traffic behavior. The Major characteristics include destination ports, protocols, and flow periods, which are required to identify the communication pattern between devices. Additional features such as the total FWD packets, total backward packets, and the total length of FWD/BWD packets capture the volume and size of the packets transmitted in each direction. These matrix are important to separate general traffic from malicious activity. In addition, FWD packet length maximum, FWD packet length minimum, and features such as FWD packet length provide insight into the variability of packet sizes, indicating unusual behavior such as DDOS attacks or data exfoliation. Overall, the rich set of features available in this dataset makes it highly suitable for machine learning-based classification tasks, allowing effective identification of suspected patterns in network traffic.

DATA PREPROCESSING:

The CIC-IDS2017 dataset is imported into the Google Colab environment. Preprocessing includes data cleaning, handling missing values, and normalization to prepare the dataset for training. Feature selection is performed to reduce dimensionality and improve model accuracy.

TRAINING AND TESTING DATA:

The Dataset is then split into , 80% of the data is allocated for training the machine learning model and 20% is reserved for testing its effectiveness. During the training phase, the Random Forest algorithm learns to identify patterns and relationships between input features and corresponding labels (normal or malicious).

RANDOM FOREST ALGORITHM:

The Random Forest algorithm is employed as the core classification technique to detect cyber threats within public networks. By Using the CIC-IDS2017 dataset, which includes labeled records of both normal and malicious traffic, the Random Forest model is trained to distinguish between legitimate network activity and various types of cyberattacks.

IV. FINDINGS:

The Research focuses on protecting smart devices that are frequently connected to public networks, making them vulnerable to various cyberattacks. The implemented system leverages machine learning, particularly the Random Forest algorithm, to identify and classify abnormal activities within network traffic. This approach is applied to a dataset collected from the CIC-IDS2017 dataset. The goal of the system is to analyze and monitor public networks accessed by smart devices, especially those vulnerable to attacks such as fake job offers, fake applications, or malicious downloads. These threats often target devices connected to publicly accessible networks like those used in smart environments. By Using Python on Google Colab, the dataset was first preprocessed and visualized to understand the pattern of network behavior. Two key packet features forward packets and backward packets were extracted and used to calculate the difference in flow rates, which is a critical indicator of potential anomalies. The system generates graphical visualizations to clearly highlight packet flow irregularities and help security analysts easily interpret the

behavior. For instance, spikes in the plot indicate a potential anomaly, allowing quick identification of threats.

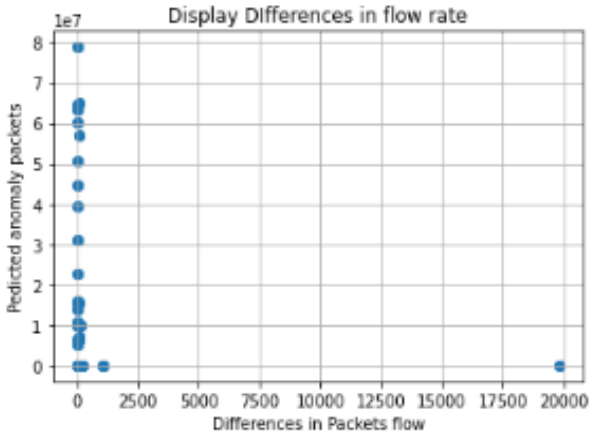

Fig 4.1.1 Anomaly Packet Data Flow Rate

To evaluate the model, predictions were compared against manually labeled outcomes using a classification report. The accuracy and performance of the model were considered satisfactory for detecting specific attack types. Furthermore, the system incorporates a simple decision mechanism. When the user enters a specific system ID, the model checks its prediction and outputs a corresponding alert, such as:

- Fake Job Offer – Social engineering attacks that lure users with fraudulent employment opportunities.
- Fake Application – Malicious apps that imitate legitimate software.
- Android Defender Alert – A general security warning for high-risk Android-based threats.

```
import numpy as np
from tkinter import messagebox
np.size([Y_rt_predict])
test_index = int(input("enter System ID to test(1:290)"))
ans=Y_rt_predict[1,test_index]
if test_index > 200 :
  print('Android_Defender');
else :
  if ans == 1.0:
    print("Fake_Job_offer")
  elif ans==2.0:
    print("Fake_App");
```

```
enter System ID to test(1:297)290
Android_Defender
```

Fig 4.1.2 System ID Based Threats Alert

V.CONCLUSION:

The Dynamic Honeypot System for Threat Detection developed in this research offers an effective and reliable way to improve the security of smart devices connected to public networks. By combining honeypot techniques with machine learning, particularly the Random Forest algorithm, the system is capable of detecting and preventing various cyber threats such as malware, and unauthorized access. [15] . The use of the CIC-IDS2017 dataset ensures that the model is trained on realistic and diverse traffic data, allowing it to recognize both known and emerging attack patterns with good accuracy. In summary, this project presents a scalable and adaptive security solution that can evolve alongside the threats it is designed to detect. It serves as a proactive defense mechanism for today's increasingly interconnected digital landscape, where smart devices are often targeted due to their exposure on public networks.

VI. REFERENCE:

[1] https://www.crowdstrike.com

[2] https://www.sophos.com

[3] A. Madhukar and K. Gautham, "Spam email detection using machine learning techniques," *International Journal of Scientific & Technology Research*, vol. 9, no. 3, pp. 653–657, 2020.

[4] T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," *IEEE Communications Surveys & Tutorials*, vol. 10, no. 4, pp. 56–76,

[5] I. Dutt, S. Borah and I.K. Maitra, "Immune System Based Intrusion Detection System (IS-IDS): A Proposed Model," in IEEE Access, Vol. 8, 2020.

[6] J. Franco, A. Aris, B. Canberk and A. S. Uluagac, "A Survey of Honeypots and Honeynets for Internet of Things, Industrial Internet of Things, and Cyber-Physical Systems," in IEEE Communications Surveys & Tutorials, vol. 23, 2021.

[7] W.Jiasheng, C. Ping, Y. Yutong, S. Zhihong and Y. Qing, "Research on Dynamic Honeynet Technology Based on Machine Learning", IEEE Smart World Congress, 2024.

[8] Abdelaziz, M.T., Radwan, A., Mamdouh, H. et al. " Enhancing Network Threat Detection with Random Forest-Based NIDS and Permutation Feature Importance". J Netw Syst Manage 33, 2 (2025).

[9] Keskin, Serkan, Sevli, Onur, "Machine Learning Based Classification For Spam Detection", vol-28, 2024.

[10] Commey, Daniel, Nkoom, Matilda, Hounsinou, Sena, Crosby, Garth, "Dynamic Honeypot Conversion for Enhanced IoT Security", 2025.

[11] Diandra Amiruddin Firmansyah and Amalia Zahra, "Honeypot-Based Thread Detection using Machine Learning Techniques",International Journal of Engineering Trends and technology, vol 71, 2023.

[12] https://www.unb.ca/cic/dataset.

[13] Hand, D.J., Christen, P. & Kirielle, N. F*: an interpretable transformation of the F-measure. Mach Learn 110, 451–456 (2021).

[14] M. Catillo, A. Del Vecchio, A. Pecchia, and U. Villano, "A Case Study with CICIDS2017 on the Robustness of Machine Learning against Adversarial Attacks in Intrusion Detection," in Proc. 18th Int. Conf. on Availability, Reliability and Security (ARES '23), Benevento, Italy, 2023, Art. No. 74. doi: 10.1145/3600160.3605031.

[15] Farooqi AH, Akhtar S, Rahman H, Sadiq T, Abbass W. Enhancing Network Intrusion Detection Using an Ensemble Voting Classifier for Internet of Things. Sensors (Basel). 2023. doi: 10.3390/s24010127.