Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 1 (2024)

Guardians of IoT: Malware Analysis of IoT Devices Using Machine Learning

Deepshika Vijayanand[1], Dr. Rabindra Kumar Singh[2]
[1,2] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India

Abstract:- The proliferation of Internet of Things (IoT) devices has introduced a new frontier for cyber threats, with malware targeting these devices becoming increasingly prevalent. This research paper presents an in-depth analysis of IoT malware using machine learning algorithms. We leverage the IoT-23 dataset, a comprehensive collection of network traffic data from both malicious and benign IoT devices, to develop and evaluate machine learning models for malware detection. Our study begins with data preprocessing, including data cleaning and feature engineering, to prepare the dataset for analysis. We explore the characteristics of the IoT-23 dataset, revealing insights into the protocols and behaviors of IoT malware. To enhance the predictive capabilities of our models, we employ techniques such as one-hot encoding to handle categorical variables effectively. We experiment with several machine learning algorithms, including Random Forest, Logistic Regression, K-Nearest Neighbors, and Naive Bayes, to classify network traffic into either benign or malicious categories. We evaluate the performance of these models using metrics such as accuracy, precision, recall, and F1-score. Additionally, we investigate the feature importance and correlations among different attributes to better understand the dataset. Our research findings shed light on the effectiveness of machine learning in detecting IoT malware, with implications for enhancing the security of IoT ecosystems. Employing machine learning models makes it possible to detect and mitigate IoT malware threats, ultimately safeguarding the integrity and privacy of IoT devices and networks. This paper contributes to the growing body of knowledge in IoT security and provides a foundation for further research in this critical domain.

Keywords*: IoT Devices, Malware Analysis, Machine Learning, Network Security, IoT Ecosystems*

1. Introduction

The advent of the Internet of Things (IoT) has ushered in a new era of connectivity and convenience, seamlessly integrating smart devices into our everyday lives. However, this increased interconnectivity has also opened the door to new security challenges, with IoT devices becoming prime targets for malicious actors. In this research paper, we delve into the critical realm of IoT security, focusing on the analysis of malware in IoT devices using advanced machine learning algorithms.

The IoT-23 dataset, a valuable resource for this study, offers a collection of network traffic captures from IoT devices. Developed by Sebastian Garcia, Agustin Parmisano, and Maria Jose Erquiaga in collaboration with Avast Software, this dataset contains real-world instances of both benign and malicious IoT network traffic. It comprises twenty captures of malicious scenarios executed on infected IoT devices, each associated with the specific malware sample employed. Additionally, three captures represent benign IoT devices, which include a Philips HUE smart LED lamp, an Amazon Echo home intelligent personal assistant, and a Somfy smart door lock. These

devices are not mere simulations; they are actual hardware, allowing us to examine and analyze real network behavior in controlled environments.

This research aims to address the pressing need for effective IoT malware detection and classification. With the proliferation of IoT devices in homes, industries, and critical infrastructure, understanding and combating IoT-based malware has become a paramount concern for security experts and researchers alike.

Our study begins by exploring the IoT-23 dataset, providing insights into its structure and content. We perform data preprocessing and analysis to gain a deeper understanding of the dataset's characteristics, such as protocol distributions and class imbalances. This preliminary analysis sets the stage for the subsequent steps in our research.

Machine learning algorithms play a pivotal role in this study. We employ a range of machine learning models, including Random Forest, Logistic Regression, K-Nearest Neighbors (KNN), and Gaussian Naive Bayes (NB), to build and evaluate classification models for detecting IoT malware. These algorithms are trained and tested on the dataset, and their performance is assessed in terms of accuracy, precision, recall, and F1-score. The goal is to identify the most effective algorithm for accurately classifying IoT network traffic as either benign or malicious.

Furthermore, we investigate the key features contributing to the accuracy of our models. By analyzing the correlation matrix, we gain insights into feature relationships and their impact on classification outcomes. This analysis aids in identifying important features and optimizing model performance.

In conclusion, this research paper offers a comprehensive exploration of IoT malware analysis using machine learning. By leveraging the IoT-23 dataset, we seek to enhance our understanding of the threats posed by malware in IoT devices and develop effective machine learning-based solutions for detection and classification. The findings presented herein contribute to the ongoing efforts to secure the ever-expanding IoT ecosystem and protect the privacy and safety of IoT users worldwide.

2. Related Work

[1] In the realm of IoT, where devices often lack intelligence and resources, the vulnerability to cyber threats looms large, posing risks such as device infections, network disruptions, and service denials to legitimate users. To counter these dangers, advanced artificial intelligence and machine learning techniques are deployed for network security. In this specific research endeavor, a support vector machine (SVM) was employed to discern normal from abnormal network traffic, enabling an in-depth analysis of network data to detect and thwart malicious activities. The study encompassed both static and dynamic malware analysis and utilized a network setup involving a Mininet emulator, VMware Fusion, Ubuntu Linux, and a tree-based network topology. Wireshark was used for scrutinizing network traffic, and the SVM classifier emerged as the top performer, boasting an impressive 99% accuracy rate.

[12] As the proliferation of Internet of Things (IoT) devices continues, the need for robust cybersecurity measures to counter cyberattacks stemming from these devices becomes increasingly crucial. While preventing malware infections in IoT devices is essential, it is challenging due to the complexity of infection techniques and the limited computational resources available for security software on these devices. Therefore, an equally important aspect is detecting malware infections to contain their spread. With the growing diversity of IoT devices and malware types, advanced anomaly detection technologies like machine learning are essential. However, IoT devices often lack the computational capacity to perform self-analysis using machine learning, making it necessary to execute such analysis at gateway devices connected to the internet. This paper presents an architecture for detecting malware traffic using summarized statistical data from packets, instead of analyzing the entire packet content. By utilizing only information on traffic volume and destination addresses for each IoT device, this approach conserves storage space and enables analysis of a large number of IoT devices with minimal computational resources. The study conducted malware traffic detection using machine learning algorithms, such as Isolation Forest and K-means clustering, on the proposed architecture and demonstrated that high accuracy can be achieved with the summarized statistical data. In their evaluation, data from 26 IoT devices across nine categories were collected, revealing that this approach reduced the required data size for analysis by over 90% while maintaining high accuracy.

[14] In addressing the security challenges posed by the extensive and diverse landscape of IoT devices, efficient data collection, and lightweight threat detection are of paramount importance. This paper introduces an architecture for malware detection, presenting methods to identify malware through flow information analysis and to minimize data transmission between servers within this framework. The study assesses both the effectiveness of malware detection and the reduction in data volume achieved through these methods. Notably, the research demonstrates that despite the reduction in data volume, the malware detection performance remains consistently robust.

[18] The widespread use of encryption in network traffic poses a significant challenge for anomaly detection systems. Current supervised and semi-supervised solutions encounter issues related to noisy data labeling, non-stationary traffic patterns, and high resource consumption for offline training. In response, this paper introduces an unsupervised, robust, and online anomaly detection approach for encrypted traffic called D2LAD. D2LAD leverages deep dictionary learning, employing an LSTM-based autoencoder to extract sequential features from raw encrypted traffic data. Using these sequential features, D2LAD iteratively explores hidden normal traffic patterns through deep dictionary learning, ultimately calculating an anomaly score for raw data based on its relevance to the deep dictionary. The paper includes a prototype implementation of D2LAD and evaluates its effectiveness and performance using real-world datasets. The results of the experiments demonstrate that D2LAD achieves high accuracy with minimal resource usage, surpassing state-of-the-art methods in real-world scenarios.

[20] Anomaly detection is emerging as a promising method for ensuring quality control in wireless and telecommunication networks, especially with the advancement of modern network architectures that support robust computing and communication at the network's edge, facilitating large-scale IoT applications. Detecting attacks in IoT infrastructure is crucial due to the growing adoption of IoT technologies, which are becoming attractive targets for various types of threats, including Denial of Service, Data Type Probing, Malicious Control, Malicious Operation, Scanning, Spying, and Wrong Setup. The paper recognizes the increasing need for information network security as internet applications continue to proliferate. It emphasizes the importance of dynamic adaptability in anomaly detection systems, given the evolving nature of network operations. The paper's primary objective is to identify attacks on IoT sensor networks and develop a generalized anomaly detection model using machine learning techniques to determine if an IoT sensor network exhibits anomalies or operates normally. Kaggle and NSL-KDD datasets were utilized for this study, with the model trained using Levenberg–Marquardt optimization and demonstrated high performance. Additionally, the paper creates a menu-based environment in Jupyter for ease of use.

[24] This paper presents a novel methodology to address the challenge of efficiently detecting security threats in IoT and CPS devices. By leveraging the power consumption patterns of wireless devices and employing Restricted Boltzmann Machine Autoencoders (RBM AE), the approach achieves robustness against various security threats. The method involves feature extraction using stacked RBM AE and Principal Component Analysis (PCA) to create feature vectors based on AE's reconstruction errors, followed by training a One-Class Support Vector Machine (OC-SVM) classifier for threat detection. Real-world dataset validation demonstrates impressive results, with detection accuracy reaching up to approximately 98%, surpassing existing techniques and affirming the practicality and effectiveness of this approach in enhancing security for IoT and CPS devices.

[25] In this paper by Zhang and Green, the focus is on addressing the critical security challenges in Internet of Things (IoT) networks, particularly the vulnerability to Distributed Denial of Service (DDoS) attacks, which can severely disrupt IoT services. The paper introduces a lightweight defensive algorithm designed to mitigate DDoS attacks in IoT environments. The proposed algorithm is tested across various scenarios to analyze communication patterns among different network nodes. Overall, the paper highlights the importance of securing IoT networks against DDoS attacks and presents a practical solution to enhance their resilience in the face of such threats.

[26] The paper authored by Džaferović, Sokol, Abd Almisreb, and Norzeli provides a review of the vulnerabilities related to Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks in Internet of Things (IoT) architecture. It acknowledges the widespread adoption of IoT devices for enhancing convenience and data collection but highlights the associated challenges and security vulnerabilities. The primary focus of the paper is

on examining the susceptibility of current IoT architectures to DDoS and DoS attacks, underscoring the need to address these issues within the IoT paradigm.

[33] This paper by Sain, Kang, and Lee provides a comprehensive overview of security concerns in the Internet of Things (IoT) domain. It highlights the diverse applications of IoT across various sectors and emphasizes the critical need for security and privacy measures, given the exchange of smart object services over potentially insecure channels. The paper conducts an analysis of IoT security in communication, application interfaces, and data security. It also reviews existing IoT technologies and models, identifying security gaps within these domains. Furthermore, the paper presents a summary of related research in IoT security and outlines open challenges and future research directions, underscoring the ongoing importance of addressing security issues in IoT ecosystems.

[42] This paper by Miranda, Kaddoum, Boukhtouta, Madi, and Alameddine addresses the vulnerability of 6LoWPAN networks, which are used in low-power IoT devices, to Rank attacks that disrupt packet routing efficiency. The paper proposes a novel intrusion prevention scheme using a combination of Software-Defined Networks (SDNs) and Reinforcement Learning (RL). The SDN controller is complemented by an RL agent, which helps optimize routing, Quality of Service (QoS) provisioning, and packet forwarding to counter Rank attacks. Experimental results validate the effectiveness of this approach, ensuring protection against Rank attacks while maintaining low latency, radio duty cycle, and maximizing packet delivery ratio. This solution offers a practical and efficient means to secure software-defined low-power IoT networks against Rank attacks.

[43] This paper by Ren, Wu, Ning, Hussain, and Chen addresses the escalating threat of Android malware in the context of the Internet of Things (IoT). It introduces two novel end-to-end Android malware detection methods based on deep learning, which do not require human expert intervention for feature engineering. Instead, these methods utilize raw bytecodes from Android applications and employ deep learning models for detection. The experiments demonstrate impressive detection accuracies of 93.4% and 95.8% for the proposed methods, emphasizing their suitability for application on Android IoT devices due to their efficiency, low resource consumption, and scalability beyond input file sizes, offering a promising solution to combat IoT-related malware threats.

[46] This paper by Bokka and Sadasivam addresses the escalating threat landscape in the Internet of Things (IoT) domain by developing a Deep Learning-based Deep Neural Network (DNN) for the detection of various attacks within IoT-based smart homes, including Denial of Service (DoS), spying, malicious control, and others. The model is evaluated using the DS2OS dataset and achieves an impressive accuracy of 99.42%, along with assessments of precision, F1-score, and recall. The research highlights the effectiveness of deep learning methods in enhancing security within IoT environments and compares the proposed model with other DL-based approaches, emphasizing the importance of robust security measures in the face of growing IoT-related attacks.

3. Methodology

The methodology of this research is divided in the following steps:

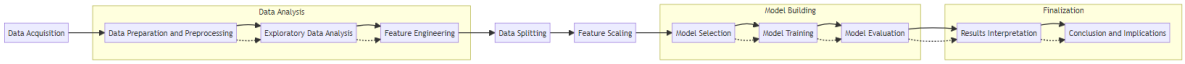

Figure 1: Methodology

1. Data Acquisition

Data Source: The research begins by acquiring the IoT-23 dataset, a comprehensive collection of 10,48,575 network traffic captures from various Internet of Things (IoT) devices. This dataset includes both benign and malicious network traffic, making it suitable for malware analysis.

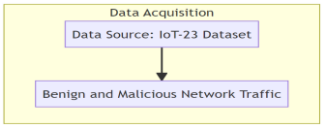

Figure 2: Data Acquisition

. Data Preprocessing

Preliminary Data Exploration: An initial examination of the dataset is conducted to understand its structure, contents, and characteristics. This phase involves gaining insights into the types of IoT devices, network protocols, and the distribution of classes (benign and malicious) within the dataset.

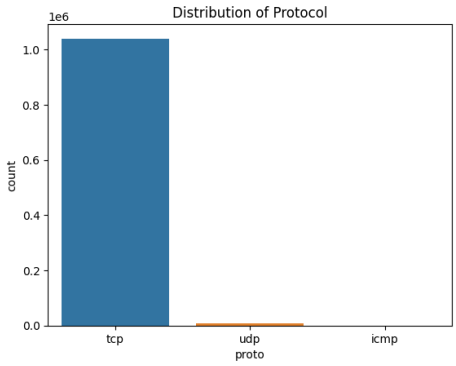

Figure 3: Distribution of Protocol

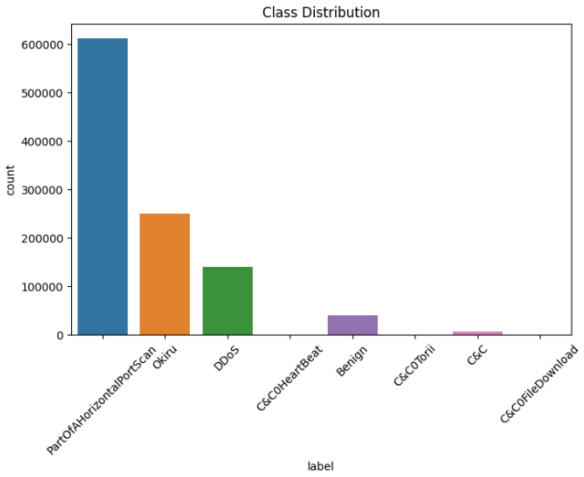

Figure 4: Class Distribution

Data Cleaning: The dataset is subjected to data cleaning procedures to handle any inconsistencies, missing values, or anomalies that may affect the quality of the analysis.

Feature Selection and Engineering: Relevant features are selected, and additional features may be engineered to enhance the dataset's suitability for machine learning algorithms. Feature scaling and transformation are applied as needed.

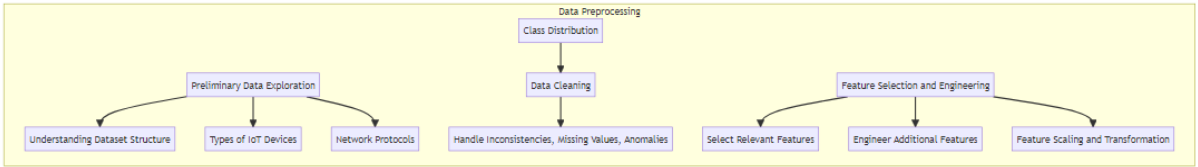

Figure 5: Data Preprocessing

3. Machine Learning Algorithms Selection

Algorithm Selection: Several machine learning algorithms are chosen for the analysis, including Random Forest, Logistic Regression, K-Nearest Neighbors (KNN), and Gaussian Naive Bayes (NB). These algorithms are selected for their appropriateness in classifying network traffic as benign or malicious.

Hyperparameter Tuning: Hyperparameter tuning is performed to optimize the selected algorithms, ensuring they are well-suited for the specific task of IoT malware detection.

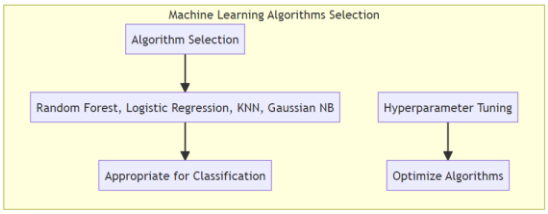

Figure 6: Machine Learning Algorithm Selection

4. Model Building and Evaluation

Training Phase: The selected machine learning algorithms are trained using the preprocessed dataset. The training process involves iteratively updating model parameters to improve classification performance.

Evaluation Metrics: Model performance is assessed using standard evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive view of the models' ability to classify IoT network traffic accurately.

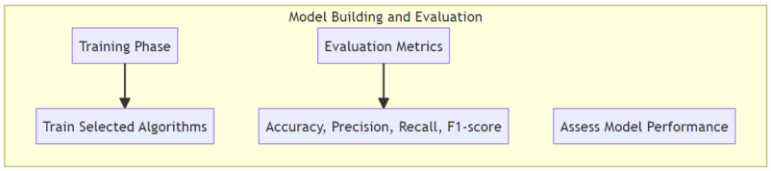

Figure 7: Model Building and Evaluation

4. Observation

The subplots provide insights into the distribution and characteristics of the 'duration,' 'orig_pkts,' 'resp_pkts,' and 'missed_bytes' variables in the dataset, which can be valuable for understanding the data and identifying any potential outliers or patterns.

Figure 8 displays a histogram of the 'duration' variable from the dataset df. The data is divided into 20 bins, and a kernel density estimate (KDE) curve is overlaid to visualize the data's distribution.

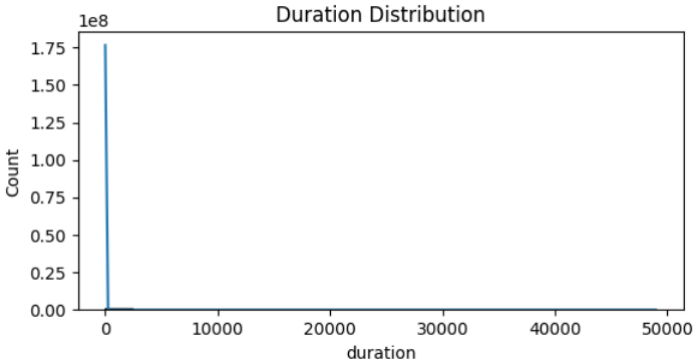

Figure 8: Duration Distribution

Figure 9 presents a box plot of the 'orig_pkts' variable from the dataset df. A box plot shows the distribution of data and helps identify outliers.

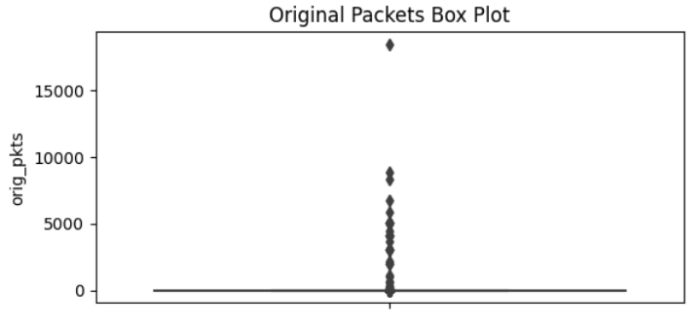

Figure 9: Original Packets Box Plot

Figure 10 displays another box plot, but this time for the 'resp_pkts' variable from the dataset df. Like the previous subplot, Figure 9, it helps visualize the distribution of data and identify potential outliers.

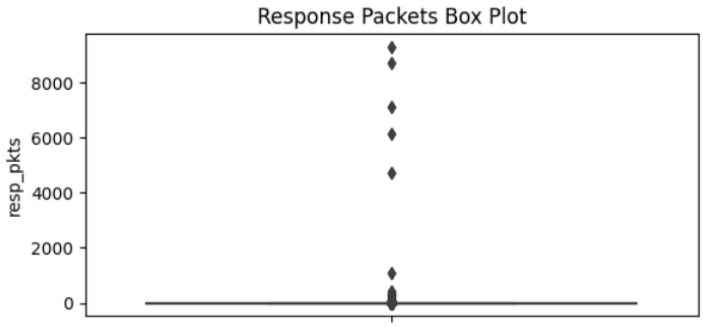

Figure 10: Response Packets Box Plot

Figure 11 visualizes the distribution of missed bytes. Box plots are useful for understanding the spread and central tendency of data.

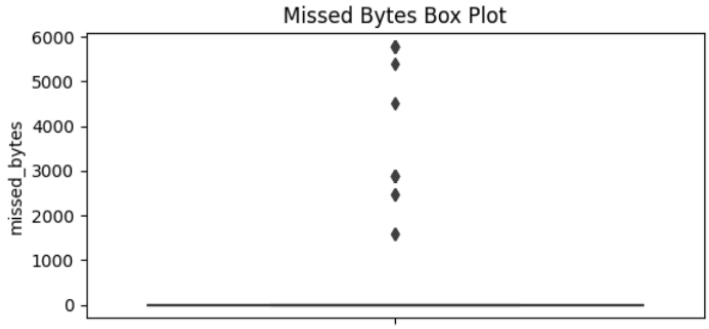

Figure 11: Missed Bytes Box Plot

The heatmap in Figure 12 provides an overview of the correlations between different attributes or features in the dataset. Each cell in the heatmap represents the correlation coefficient between two variables. In the context of IoT malware analysis, it helps identify whether there are strong positive or negative correlations between attributes. The annot=True parameter displays the actual correlation values within the cells, which can be insightful for understanding the relationships.

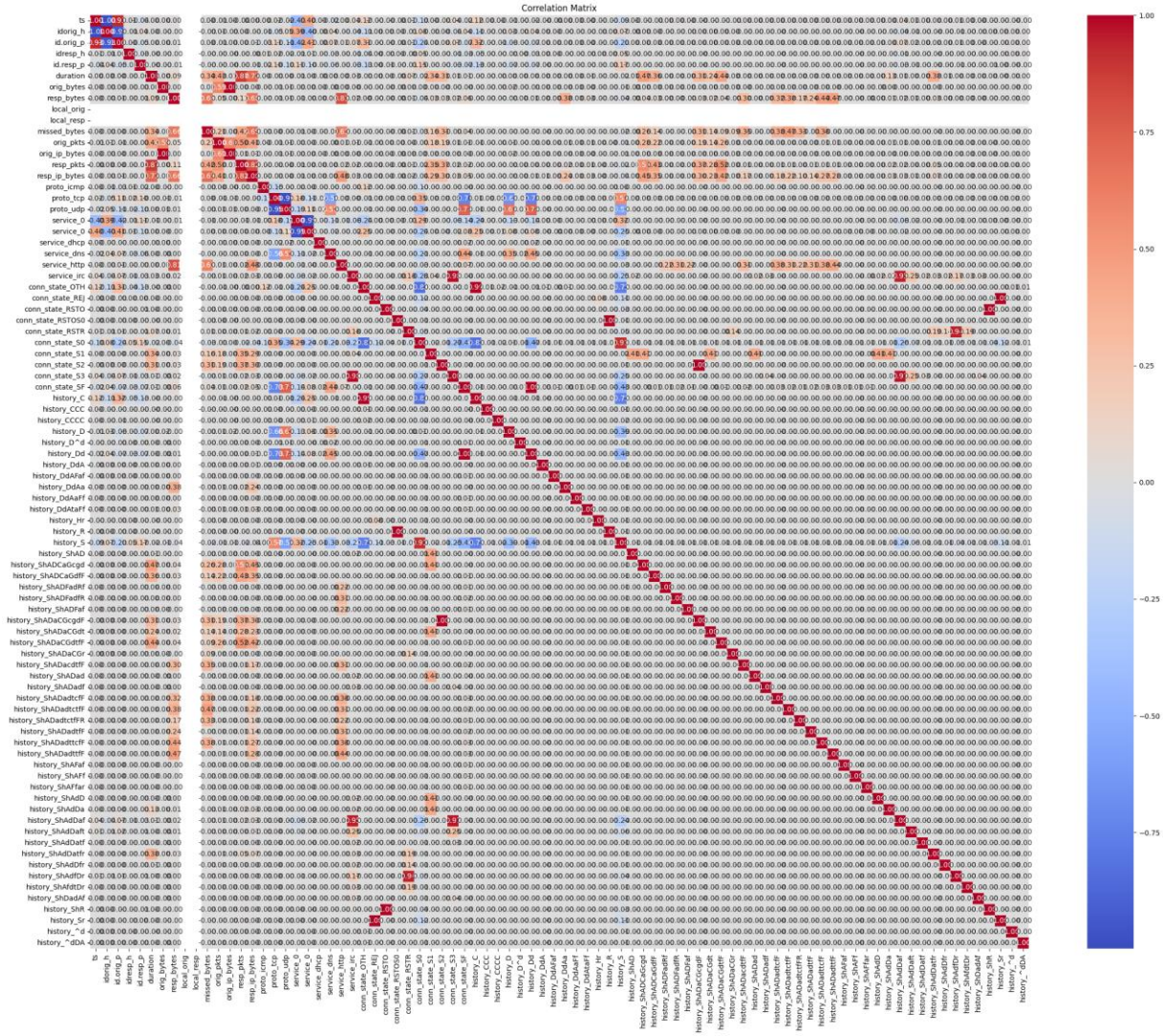

Figure 12: Correlation Heatmap

5. Result:

In the evaluation of machine learning models for IoT malware detection, four classifiers, namely Random Forest Classifier, Logistic Regression, K Neighbors Classifier, and Gaussian NB, were applied to a comprehensive dataset containing both benign and malicious network traffic. The results are summarized as follows:

Random Forest Classifier:

Table 1: Random Forest Classification Report

	Precision	Recall	F1-Score	Support
Benign	1.00	1.00	1.00	8171
C&C	1.00	1.00	1.00	1337
C&C0FileDownload	1.00	1.00	1.00	2
C&C0HeartBeat	1.00	1.00	1.00	22
C&C0Torii	1.00	1.00	1.00	6
DDoS	1.00	1.00	1.00	27800
Okiru	1.00	1.00	1.00	49927
PartOfAHorizontalPortScan	1.00	1.00	1.00	122450
Accuracy			1.00	209715
Macro Avg	1.00	1.00	1.00	209715
Weighted Avg	1.00	1.00	1.00	209715

- Achieved an exceptional accuracy of approximately 99.99%.
- Demonstrated high precision, recall, and F1-scores for various classes, including C&C, DDoS, Okiru, and others.
- The confusion matrix showed successful classification for most classes, with minimal misclassifications.

```
Confusion Matrix:
 [[  8171      0      0      0      0      0      0      0]
 [     0   1337      0      0      0      0      0      0]
 [     0      0      2      0      0      0      0      0]
 [     0      0      0     22      0      0      0      0]
 [     0      0      0      0      6      0      0      0]
 [     0      0      0      0      0  27800      0      0]
 [     0      0      0      0      0      0  49927      0]
 [     2      0      0      0      0      0      0 122448]]
```

Figure 13: Random Forest Classifier- Confusion Matrix

Logistic Regression:

Table 2: Logistic Regression Classification Report

	precision	recall	f1-score	support
Benign	0.00	0.00	0.00	8171
C&C	0.00	0.00	0.00	1337
C&C0FileDownload	0.00	0.00	0.00	2
C&C0HeartBeat	0.00	0.00	0.00	22
C&C0Torii	0.00	0.00	0.00	6
DDoS	0.00	0.00	0.00	27800
Okiru	0.00	0.00	0.00	49927
PartOfAHorizontalPortScan	0.58	1.00	0.74	122450
accuracy			0.58	209715
macro avg	0.07	0.12	0.09	209715
weighted avg	0.34	0.58	0.43	209715

- Exhibited a considerably lower accuracy of around 58.39%.
- Notably underperformed in several classes, with precision, recall, and F1-scores close to zero for some.

- The model struggled to effectively classify network traffic into the various classes.

```
Confusion Matrix:
 [[     0      0      0      0      0      0      0   8171]
 [     0      0      0      0      0      0      0   1337]
 [     0      0      0      0      0      0      0      2]
 [     0      0      0      0      0      0      0     22]
 [     0      0      0      0      0      0      0      6]
 [     0      0      0      0      0      0      0  27800]
 [     0      0      0      0      0      0      0  49927]
 [     0      0      0      0      0      0      0 122450]]
```

Figure 14: Logistic Regression- Confusion Matrix

K Neighbors Classifier:

Table 3: K Neighbors Classification Report

	precision	recall	f1-score	support
Benign	0.76	0.39	0.52	8171
C&C	1.00	1.00	1.00	1337
C&C0FileDownload	1.00	1.00	1.00	2
C&C0HeartBeat	0.92	1.00	0.96	22
C&C0Torii	1.00	1.00	1.00	6
DDoS	1.00	1.00	1.00	27800
Okiru	0.98	1.00	0.99	49927
PartOfAHorizontalPortScan	0.96	0.98	0.97	122450
accuracy			0.97	209715
macro avg	0.95	0.92	0.93	209715
weighted avg	0.96	0.97	0.96	209715

- Achieved a commendable accuracy of approximately 96.63%.
- Displayed strong performance in terms of precision, recall, and F1-scores, particularly for C&C, C&C0FileDownload, C&C0HeartBeat, and DDoS classes.
- The confusion matrix revealed successful classification across most classes, with some misclassifications, especially in the 'Benign' class.

```
Confusion Matrix:
 [[  3220      0      0      0      0      0     63   4888]
 [     0   1337      0      0      0      0      0      0]
 [     0      0      2      0      0      0      0      0]
 [     0      0      0     22      0      0      0      0]
 [     0      0      0      0      6      0      0      0]
 [     0      0      0      0      0  27701     99      0]
 [     0      0      0      0      0     61  49816     50]
 [   992      0      0      2      0      0    920 120536]]
```

Figure 15: K neighbours Classifier- Confusion Matrix

Gaussian NB:

Table 4: Gaussian NB Classification Report

	precision	recall	f1-score	support
Benign	0.50	0.02	0.04	8171
C&C	0.06	0.99	0.10	1337
C&C0FileDownload	0.03	1.00	0.06	2
C&C0HeartBeat	0.19	1.00	0.32	22
C&C0Torii	0.05	1.00	0.09	6
DDoS	0.25	0.83	0.39	27800
Okiru	0.00	0.00	0.00	49927
PartOfAHorizontalPortScan	0.67	0.51	0.58	122450
accuracy			0.42	209715
macro avg	0.22	0.67	0.20	209715
weighted avg	0.44	0.42	0.39	209715

- Attained an accuracy of approximately 41.50%.
- Demonstrated varying precision, recall, and F1-scores across classes, with substantial performance in classes like C&C, C&C0FileDownload, C&C0HeartBeat, and DDoS.
- Struggled in accurately classifying 'Benign' traffic and showed limitations in certain classes.

```
Confusion Matrix:
 [[  170   985    11     3     0  1238     0  5764]
 [    0  1329     8     0     0     0     0     0]
 [    0     0     2     0     0     0     0     0]
 [    0     0     0    22     0     0     0     0]
 [    0     0     0     0     6     0     0     0]
 [    0     0     0     0     0 23060     0  4740]
 [    0  6436     0     0     0 23064     0 20427]
 [  169 15229    47    90   127 44337     0 62451]]
```

Figure 16: Gaussian NB- Confusion Matrix

These results underscore the importance of selecting an appropriate machine learning algorithm for IoT malware detection.

Table 5: Machine Learning Model and its Accuracy

	Machine Learning Model	Accuracy
1	Random Forest Classifier	0.999
2	Logistic Regression	0.583
3	K Neighbours Classifier	0.966
4	Gaussian NB	0.415

RandomForestClassifier and KNeighborsClassifier exhibited strong performance, while LogisticRegression and GaussianNB faced challenges in achieving accurate classifications. The findings contribute valuable insights into the effectiveness of different classifiers in mitigating IoT malware threats, with implications for enhancing the security of IoT ecosystems.

6. Conclusion

In conclusion, this research paper delves into the critical realm of IoT security by conducting an in-depth analysis of IoT malware using a variety of machine learning algorithms. Leveraging the IoT-23 dataset, which encompasses a vast collection of network traffic data from both benign and malicious IoT devices, our study aims to develop and evaluate machine learning models for the detection of IoT malware. The findings reveal that certain algorithms, such as Random Forest Classifier and K Neighbors Classifier, exhibit promising capabilities in accurately detecting IoT malware, offering potential solutions to the escalating threats faced by IoT ecosystems. These insights underscore the significance of harnessing machine learning for IoT security and pave the way for further advancements in safeguarding IoT devices and networks, setting the stage for continued exploration and innovation in this critical domain.

7. Future Work

As the IoT landscape evolves, future research in IoT security should adopt a multi-faceted approach. This includes refining machine learning models, exploring ensemble techniques, and focusing on real-time threat detection. Collaborative data sharing efforts and compliance with emerging regulations are crucial. Additionally, advanced behavioral analysis, blockchain integration, quantum-resistant security, and user awareness initiatives are key areas for development. A holistic approach that combines technological innovation, regulatory compliance, ethical considerations, and user education will ensure a secure and resilient IoT ecosystem.

Refrences

[1] Mishra, S. (2021). Network Traffic Analysis Using Machine Learning Techniques in IoT Networks. *International Journal of Software Innovation (IJSI)*, *9*(4), 107-123.

[2] Hachim, E. A. W., Abbas, T., & Gaata, M. T. (2022, November). Modified RC4 Algorithm for Improve Data Protection in Cloud Environment. In *2022 International Conference on Information Technology Systems and Innovation (ICITSI)* (pp. 295-299). IEEE.

[3] Muñoz Castañeda, Á. L., Mata, J. A. A., & Aláiz-Moretón, H. (2023). Characterization of threats in IoT from an MQTT protocol-oriented dataset. *Complex & Intelligent Systems*, 1-16.

[4] Zeng, J. Y., Chang, L. E., Cho, H. H., Chen, C. Y., Chao, H. C., & Yeh, K. H. (2022, June). Using Poisson Distribution to Enhance CNN-based NB-IoT LDoS Attack Detection. In *2022 IEEE Conference on Dependable and Secure Computing (DSC)* (pp. 1-7). IEEE.

[5] Mishra, A. K., Sinha, M., & Tripathy, A. K. (2020). A sinkhole prevention mechanism for RPL in IoT. *International Journal of Computational Science and Engineering*, *23*(3), 262-270.

[6] Braghin, C., Lilli, M., & Riccobene, E. (2023). A model-based approach for vulnerability analysis of IoT security protocols: The Z-Wave case study. *Computers & Security*, *127*, 103037.

[7] Ahmad, M. Z., Adenan, A. R., Rohmad, M. S., & Yussoff, Y. M. (2023, March). Performance Analysis of Secure MQTT Communication Protocol. In *2023 19th IEEE International Colloquium on Signal Processing & Its Applications (CSPA)* (pp. 225-229). IEEE.

[8] De Vita, F., Bruneo, D., & Das, S. K. (2020, April). A novel data collection framework for telemetry and anomaly detection in industrial iot systems. In *2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI)* (pp. 245-251). IEEE.

[9] Wang, R., Gu, C., He, S., Shi, Z., & Meng, W. (2022). An interoperable and flat Industrial Internet of Things architecture for low latency data collection in manufacturing systems. *Journal of Systems Architecture*, *129*, 102631.

[10] Bin Mofidul, R., Alam, M. M., Rahman, M. H., & Jang, Y. M. (2022). Real-time energy data acquisition, anomaly detection, and monitoring system: Implementation of a secured, robust, and integrated global IIoT infrastructure with edge and cloud AI. *Sensors*, *22*(22), 8980.

[11] Wani, A. R., Gupta, S. K., Khanam, Z., Rashid, M., Alshamrani, S. S., & Baz, M. (2022). A novel approach for securing data against adversary attacks in UAV embedded HetNet using identity based authentication scheme. *IET Intelligent Transport Systems*.

[12] Nakahara, M., Okui, N., Kobayashi, Y., & Miyake, Y. (2020). Machine Learning based Malware Traffic Detection on IoT Devices using Summarized Packet Data. In *IoTBDS* (pp. 78-87).

[13] Nakahara, M., Okui, N., Kobayashi, Y., & Miyake, Y. (2021). Malware Detection for IoT Devices using Automatically Generated White List and Isolation Forest. In *IoTBDS* (pp. 38-47).

[14] Nakahara, M., Okui, N., Kobayashi, Y., Miyake, Y., & Kubota, A. (2022). Malware detection for IoT devices using hybrid system of whitelist and machine learning based on lightweight flow data. *Enterprise Information Systems*, 2142854.

[15] Liu, S., Han, Y., Hu, Y., & Tan, Q. (2021, September). Fa-net: Attention-based fusion network for malware https traffic classification. In *2021 IEEE Symposium on Computers and Communications (ISCC)* (pp. 1-7). IEEE.

[16] Liu, J., Tian, Z., Zheng, R., & Liu, L. (2019). A distance-based method for building an encrypted malware traffic identification framework. *IEEE Access*, *7*, 100014-100028.

[17] Yu, T., Zou, F., Li, L., & Yi, P. (2019, October). An encrypted malicious traffic detection system based on neural network. In *2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)* (pp. 62-70). IEEE.

[18] Xing, J., & Wu, C. (2020, July). Detecting anomalies in encrypted traffic via deep dictionary learning. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 734-739). IEEE.

[19] Chen, L., Gao, S., Liu, B., Lu, Z., & Jiang, Z. (2020). THS-IDPC: A three-stage hierarchical sampling method based on improved density peaks clustering algorithm for encrypted malicious traffic detection. *The Journal of Supercomputing*, *76*, 7489-7518.

[20] Mithran, K., & Gopi, C. (2022, June). Anomaly Detection in IoT Sensor Networks using Machine Learning. In *2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)* (pp. 1-7). IEEE.

[21] Feraudo, A., Yadav, P., Mortier, R., Bellavista, P., & Crowcroft, J. (2020). SoK: Beyond IoT MUD Deployments--Challenges and Future Directions. *arXiv preprint arXiv:2004.08003*.

[22] McKinney, S. (2019). Graph-Based Analysis for IoT Devices With Manufacturer Usage Descriptions—An ScM Research Project.

[23] Afek, Y., Bremler-Barr, A., Hay, D., Shafir, L., & Zhaika, I. (2020, April). NFV-based IoT security at the ISP level. In *NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium* (pp. 1-2). IEEE.

[24] Albasir, A., Hu, Q., Naik, K., & Naik, N. (2021). Unsupervised detection of security threats in cyberphysical system and IoT devices based on power fingerprints and RBM autoencoders. *Journal of Surveillance, Security and Safety*, *2*(1), 1-25.

[25] Zhang, C., & Green, R. (2015, April). Communication security in internet of thing: preventive measure and avoid DDoS attack over IoT network. In *Proceedings of the 18th symposium on communications & networking* (pp. 8-15).

[26] Džaferović, E., Sokol, A., Abd Almisreb, A., & Norzeli, S. M. (2019). DoS and DDoS vulnerability of IoT: a review. *Sustainable Engineering and Innovation*, *1*(1), 43-48.

[27] Vishwakarma, R., & Jain, A. K. (2020). A survey of DDoS attacking techniques and defence mechanisms in the IoT network. *Telecommunication systems*, *73*(1), 3-25.

[28] Misra, S., Krishna, P. V., Agarwal, H., Saxena, A., & Obaidat, M. S. (2011, October). A learning automata based solution for preventing distributed denial of service in internet of things. In *2011 international conference on internet of things and 4th international conference on cyber, physical and social computing* (pp. 114-122). IEEE.

[29] Malik, S., & Chauhan, R. (2020, February). Securing the internet of things using machine learning: A review. In *2020 International Conference on Convergence to Digital World-Quo Vadis (ICCDW)* (pp. 1-4). IEEE.

[30] Lee, W., & Jin, S. (2021, January). Encrypted malware traffic detection using tls features and random forest. In *International Conference on Computational & Experimental Engineering and Sciences* (pp. 85-100). Cham: Springer International Publishing.

[31] Li, R., Song, Z., Xie, W., Zhang, C., Zhong, G., & Pei, X. (2021, October). Halnet: A hybrid deep learning model for encrypted c&c malware traffic detection. In *International Conference on Network and System Security* (pp. 326-339). Cham: Springer International Publishing.

[32] Deore, B., Kyatham, A., & Narkhede, S. (2020). A novel approach to ensemble MLP and random forest for network security. In *ITM Web of Conferences* (Vol. 32, p. 03003). EDP Sciences.

[33] Sain, M., Kang, Y. J., & Lee, H. J. (2017, February). Survey on security in Internet of Things: State of the art and challenges. In *2017 19th International conference on advanced communication technology (ICACT)* (pp. 699-704). IEEE.

[34] Yang, Y., Wu, L., Yin, G., Li, L., & Zhao, H. (2017). A survey on security and privacy issues in Internet-of-Things. *IEEE Internet of things Journal*, *4*(5), 1250-1258.

[35] Ammar, M., Russello, G., & Crispo, B. (2018). Internet of Things: A survey on the security of IoT frameworks. *Journal of Information Security and Applications*, *38*, 8-27.

[36] Liu, Z., Zhang, L., Ni, Q., Chen, J., Wang, R., Li, Y., & He, Y. (2019). An integrated architecture for IoT malware analysis and detection. In *IoT as a Service: 4th EAI International Conference, IoTaaS 2018, Xi'an, China, November 17–18, 2018, Proceedings 4* (pp. 127-137). Springer International Publishing.

[37] Fujita, H., & Selamat, A. (2019, September). Clustering botnet behavior using k-means with uncertain data. In *IOS Press* (Vol. 318, p. 244).

[38] Zhang, S., Bu, Y., Chen, B., & Lu, X. (2021, April). Transfer learning for encrypted malicious traffic detection based on efficientnet. In *2021 3rd International Conference on Advances in Computer Technology, Information Science and Communication (CTISC)* (pp. 72-76). IEEE.

[39] Hamza, A., Ranathunga, D., Gharakheili, H. H., Benson, T. A., Roughan, M., & Sivaraman, V. (2020). Verifying and monitoring iots network behavior using mud profiles. *IEEE Transactions on Dependable and Secure Computing*, *19*(1), 1-18.

[40] Lade, V., Mohan, A., & Patil, S. (2018). 802.11 ax for Internet Of Things-Machine Learning Assisted Optimized Power Save Techniques for Iot Devices Using 802.11 ax Target Wake Time.

[41] Tu, J., Ogola, W., Xu, D., & Xie, W. (2022). Intrusion Detection Based on Generative Adversarial Network of Reinforcement Learning Strategy for Wireless Sensor Networks. *International Journal of Circuits, Systems and Signal Processing*, *16*, 478-482.

[42] Miranda, C., Kaddoum, G., Boukhtouta, A., Madi, T., & Alameddine, H. A. (2022). Intrusion prevention scheme against rank attacks for software-defined low power IoT networks. *IEEE Access*, *10*, 129970-129984.

[43] Ren, Z., Wu, H., Ning, Q., Hussain, I., & Chen, B. (2020). End-to-end malware detection for android IoT devices using deep learning. *Ad Hoc Networks*, *101*, 102098.

[44] Kumar, R., Zhang, X., Wang, W., Khan, R. U., Kumar, J., & Sharif, A. (2019). A multimodal malware detection technique for Android IoT devices using various features. *IEEE access*, *7*, 64411-64430.

[45] Ngo, Q. D., Nguyen, H. T., Le, V. H., & Nguyen, D. H. (2020). A survey of IoT malware and detection methods based on static features. *ICT Express*, *6*(4), 280-286.

[46] Bokka, R., & Sadasivam, T. (2021). Deep learning model for detection of attacks in the Internet of Things based smart home environment. In *Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2020* (pp. 725-735). Springer Singapore.

[47] Feng, Y., Cai, W., Yue, H., Xu, J., Lin, Y., Chen, J., & Hu, Z. (2022). An improved X-means and isolation forest based methodology for network traffic anomaly detection. *Plos one*, *17*(1), e0263423.

[48] Doshi, K., Mozaffari, M., & Yilmaz, Y. (2019, May). RAPID: Real-time anomaly-based preventive intrusion detection. In *Proceedings of the ACM Workshop on Wireless Security and Machine Learning* (pp. 49-54).