Detection of DDoS Attacks in IoT Networks Using Machine Learning Algorithms

Alwan Ahmed
Abdulrahman Alwan
Kulliyyah Information & Communications Technology
International Islamic University Malaysia, Kuala Lumpur, Malaysia
ahmed.alwan@live.iium.edu.my

2nd Asadullah Shah
Kulliyyah Information & Communications Technology
International Islamic University
asadullah@iium.edu.my

3rd Alwan Abdullah
Abdulrahman Alwan
National Advanced IPv6 Centre, Universiti Sains Malayisa Penang, Malaysia
Alwan.aa@student.usm.my

4th Shams Ul Arfeen Laghari
National Advanced IPv6 Centre, Universiti Sains Malayisa
Penang, Malaysia
shamsularfeen@usm.my

2024 IEEE 9th International Conference on Engineering Technologies and Applied Sciences (ICETAS) | 979-8-3503-6314-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICETAS62372.2024.11120083

Abstract*—The rapid proliferation of Internet of Things (IoT) devices has revolutionized various industries by enabling seamless connectivity and data exchange. However, this connectivity also introduces significant security challenges, particularly in the form of Distributed Denial of Service (DDoS) attacks. These attacks can overwhelm IoT networks, leading to service disruptions and substantial financial losses. This paper presents a robust and efficient framework for detecting DDoS attacks in IoT networks using advanced machine learning techniques and effective feature selection methods. The study utilizes the CICIoT2023 dataset and employs Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) to enhance the performance of machine learning models, including Random Forest, Support Vector Machine (SVM), Na¨ıve Bayes, XGBoost, and K-Nearest Neighbors (KNN). The models are trained and validated using k-fold cross-validation to ensure robustness and generalizability. Expected results indicate significant improvements in detection accuracy, precision, recall, and computational efficiency. The findings underscore the importance of feature selection in improving model performance and provide valuable insights into the strengths and weaknesses of different machine learning models. This research contributes to the development of scalable and effective DDoS detection solutions for IoT networks, ensuring their reliability and resilience against evolving cyber threats. Future work will focus on exploring additional feature selection methods, integrating deep learning techniques, and validating the models in real-world IoT environments.

***Index Terms*—Internet of Things (IoT), Distributed Denial of Service (DDoS) attacks, Machine learning, Feature selection**

I. Introduction

The Internet of Things (IoT) is revolutionizing various industries by connecting billions of devices globally, enabling seamless communication and data exchange. By 2025, it is estimated that over 75 billion IoT devices will be in use, generating immense volumes of data and fostering innovative applications across sectors such as healthcare, smart cities, and industrial automation [1]. Despite its transformative potential, the expansive connectivity of IoT networks introduces significant security challenges, with Distributed Denial of Service (DDoS) attacks being one of the most severe threats [2]. DDoS attacks disrupt the normal functioning of IoT networks by overwhelming devices with excessive traffic, rendering critical services unavailable. A study by Symantec [3] revealed that IoT devices experience an attack every two minutes on average, highlighting the urgency of addressing this issue. The financial implications are equally staggering, with IoTrelated cyberattacks projected to cost businesses over $300 billion annually, according to recent reports [4], [5]. As IoT devices often operate with limited computational resources, traditional security measures are insufficient to mitigate these sophisticated attacks effectively [6].

The rapid growth of IoT devices presents significant challenges in detecting Distributed Denial of Service (DDoS) attacks due to the sheer volume and complexity of network traffic. Existing DDoS detection methods often lack the accuracy and efficiency needed for real-time applications in resourceconstrained IoT environments [7]. This paper addresses these issues by developing and evaluating various machine learning algorithms to improve DDoS detection in IoT networks. Leveraging the CICIoT2023 dataset[8], the study aims to enhance detection accuracy while reducing computational overhead through effective feature selection. The primary objective of this research is to develop a robust and efficient DDoS detection framework tailored for IoT networks. Specific objectives include improving detection accuracy by utilizing advanced machine learning algorithms to accurately distinguish between normal and malicious traffic, reducing computational overhead by implementing feature selection techniques to minimize the computational burden on resource-constrained IoT devices, and performing a comparative analysis to evaluate and compare the performance of different machine learning models to identify the most effective approach for DDoS detection in IoT environments. By achieving these objectives, this study aims to contribute to the development of scalable and effective security solutions for the rapidly expanding IoT landscape, ensuring the reliability and resilience of these critical networks.

The remainder of this paper is organized as follows: Section 2 reviews related work on DDoS detection in IoT networks using machine learning. Section 3 describes the proposed methodology, including the dataset, feature selection techniques, and machine learning models used. Section 4 discusses the expected results and provides an analysis of the potential outcomes. Section 5 concludes the paper with a summary of the research findings and suggestions for future

979-8-3503-6314-2 /24/$31.00 ©2024 IEEE

work. Finally, Section 6 lists the references cited throughout the paper.

II. Related Work

The detection of Distributed Denial of Service (DDoS) attacks in Internet of Things (IoT) networks has been a significant area of research due to the unique challenges posed by the scale and heterogeneity of IoT devices. Various approaches have been proposed to address these challenges, leveraging different machine learning and feature selection techniques to enhance detection accuracy and efficiency. One prominent approach is the use of machine learning algorithms for DDoS detection. Doshi et al. [7] explored the application of machine learning techniques to identify DDoS attacks in consumer IoT devices. Their study demonstrated that models such as Random Forest and Support Vector Machine (SVM) could achieve high detection rates but highlighted the need for further optimization to reduce false positives and computational overhead. Similarly, Moustafa and Slay [9] proposed a comprehensive dataset for network intrusion detection systems, which includes DDoS attacks. They employed various machine learning algorithms and found that ensemble methods like Random Forest provided robust detection capabilities compared to single classifiers.

Feature selection plays a critical role in improving the performance of machine learning models by reducing the dimensionality of the data. Kolias et al. [2] discussed the importance of selecting relevant features to enhance the detection of DDoS attacks in IoT networks. They highlighted that feature selection techniques such as Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) could significantly improve model efficiency and accuracy by focusing on the most informative attributes. In another study, Hassija et al. [10] utilized a hybrid feature selection method combining filter and wrapper approaches to optimize the detection of DDoS attacks, achieving higher accuracy and reduced computational costs.

Deep learning methods have also been explored for DDoS detection in IoT networks. Fu et al. [11] investigated the use of deep learning techniques, such as Convolutional Neural Networks (CNNs), to automatically learn features from raw network traffic data. Their findings indicated that deep learning models could outperform traditional machine learning approaches in terms of detection accuracy, but they also required more computational resources, which could be a limitation for resource-constrained IoT devices.

Another approach involves the use of hybrid models that combine multiple detection techniques to enhance performance. Alomari et al. [12] proposed a hybrid detection model that integrates machine learning with statistical methods to identify anomalies in IoT network traffic. Their approach improved detection rates by leveraging the strengths of both methods, though it also highlighted the need for balancing detection accuracy with computational efficiency.

Zhang et al. (2020) introduced a novel approach by combining rule-based systems with machine learning for anomaly detection in IoT networks, showing significant improvements in detection rates [13]. In a similar vein, Lee and Lee (2020) explored the use of reinforcement learning for adaptive DDoS attack mitigation in IoT environments, achieving promising results in dynamic threat landscapes [14]. Another noteworthy contribution is from He et al. (2020), who developed a federated learning framework for DDoS detection, enhancing privacy and collaboration among IoT devices [15].

Despite these advancements, challenges remain in developing scalable and efficient DDoS detection solutions for IoT networks. Existing methods often struggle with high false positive rates, scalability issues, and the need for real-time detection capabilities. This paper builds on the existing body of work by proposing an optimized machine learning-based approach using the latest CICIoT2023 dataset. By focusing on feature selection and comparing various machine learning models, this study aims to address the limitations of current methods and contribute to the development of more robust IoT security solutions.

III. Proposed Methodology

This section details the methodology employed to develop and evaluate the machine learning-based framework for detecting Distributed Denial of Service (DDoS) attacks in Internet of Things (IoT) networks. The proposed approach leverages the CICIoT2023 dataset, feature selection techniques, and various machine learning models to enhance detection accuracy and efficiency.

A. Dataset Description

The CICIoT2023 dataset, provided by the Canadian Institute for Cybersecurity, is used in this study. This dataset includes a comprehensive collection of IoT network traffic data, comprising various types of DDoS attacks such as UDP flood, TCP SYN flood, and HTTP flood. The dataset's diversity and realism make it suitable for training and evaluating DDoS detection models [9].

B. Feature Selection

Feature selection is a critical step in the proposed methodology, aimed at improving the performance of machine learning models by reducing the dimensionality of the data and focusing on the most relevant features. Two primary techniques are utilized:

1) Recursive Feature Elimination (RFE): RFE is a wrapperbased feature selection method that recursively removes the least important features based on the performance of a given model. This technique helps in identifying the most significant features that contribute to DDoS attack detection [2].
2) Principal Component Analysis (PCA): PCA is a statistical technique that transforms the original set of features into a new set of uncorrelated features called principal components. These components capture the

maximum variance in the data, allowing for dimensionality reduction while preserving essential information [10].

C. Machine Learning Models

Several machine learning models are evaluated in this study to identify the most effective approach for DDoS detection in IoT networks:

1) Random Forest: An ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification[16]. Random Forest is known for its robustness and high accuracy in various classification tasks [7].
2) Support Vector Machine (SVM): A supervised learning model that finds the optimal hyperplane separating different classes in the feature space[17]. SVM is effective in high-dimensional spaces and is widely used for classification problems [10].
3) Na¨ıve Bayes: A probabilistic classifier based on Bayes' theorem, assuming independence between features. Despite its simplicity[18], Na¨ıve Bayes performs well in many classification tasks, especially with large datasets [2].
4) XGBoost: An optimized gradient boosting algorithm that improves the performance and speed of decision tree models[19]. XGBoost is known for its high accuracy and efficiency, making it suitable for large-scale data [9].
5) K-Nearest Neighbors (KNN): A non-parametric method that classifies data points based on the majority class of their k-nearest neighbors[20]. KNN is simple and effective but can be computationally intensive for large datasets [11].

D. Model Training and Validation

The models are trained and validated using k-fold crossvalidation to ensure robustness and generalizability. In this study, a 10-fold cross-validation approach is employed, where the dataset is randomly partitioned into 10 equal-sized subsets. Each subset is used as a validation set, while the remaining subsets form the training set. This process is repeated 10 times, and the average performance metrics are reported [12].

E. Performance Metrics

The performance of the models is evaluated using the following metrics:

- Accuracy: The proportion of correctly classified instances out of the total instances.
- Precision: The proportion of true positive instances out of the total predicted positive instances.
- Recall: The proportion of true positive instances out of the actual positive instances.

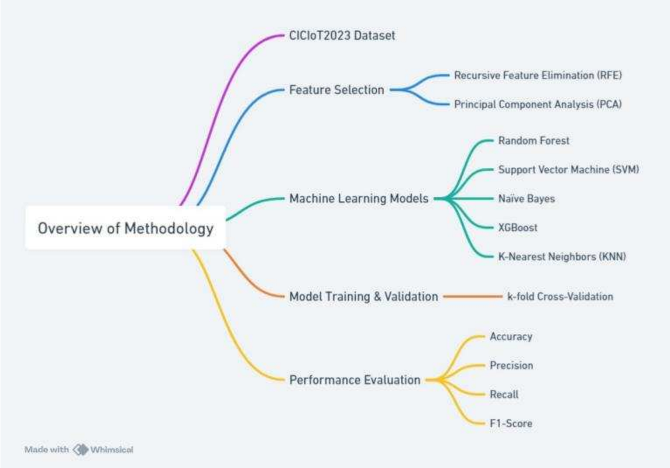

Fig. 1. Proposed Methodology

- F1-Score: The harmonic mean of precision and recall, providing a single measure of the model's performance.

By utilizing the CICIoT2023 dataset, applying robust feature selection techniques, and evaluating various machine learning models, this study aims to develop an efficient and accurate DDoS detection framework for IoT networks. The next section discusses the expected results and provides an analysis of the potential outcomes. An overview of the proposed research methodology can be found in Fig. 1.

IV. Expected Results and Discussion

A. Expected Results

The primary goal of this study is to enhance the detection of Distributed Denial of Service (DDoS) attacks in Internet of Things (IoT) networks by leveraging advanced machine learning techniques and effective feature selection methods. Based on the proposed methodology, the expected results are as follows:

1) Improved Detection Accuracy: The machine learning models, particularly Random Forest and XGBoost, are expected to demonstrate high accuracy in distinguishing between normal and malicious traffic [7], [9]. The feature selection techniques (RFE and PCA) should contribute significantly to this improvement by reducing the dimensionality of the dataset and focusing on the most relevant features [2], [10].
2) Enhanced Precision and Recall: The precision and recall metrics are anticipated to be higher for models that have undergone rigorous feature selection. This means that the models will not only detect a higher number of true positives (actual attacks) but also reduce the number of false positives (normal traffic misclassified as attacks) [2], [10].
3) Efficient Computational Performance: By incorporating feature selection techniques, the computational load on

the models should be reduced. This is crucial for IoT environments where computational resources are limited. Models like Na¨ıve Bayes and SVM, which are generally more lightweight, are expected to perform well in terms of computational efficiency [11].

4) Robustness and Generalizability: The use of k-fold cross-validation will ensure that the models are robust and generalizable. The expected outcome is that the models will perform consistently well across different subsets of the data, indicating their reliability in realworld scenarios [12].

B. Discussion

The anticipated results of this study have several important implications for the field of IoT security and DDoS attack detection:

1) Significance of Feature Selection: The use of Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) is expected to highlight the importance of feature selection in improving model performance. By focusing on the most informative features, the models can achieve higher accuracy and efficiency, which is particularly beneficial for resource-constrained IoT devices [2], [10].
2) Model Comparison and Insights: Comparing different machine learning models will provide valuable insights into their relative strengths and weaknesses. For example, while Random Forest and XGBoost may offer higher accuracy, Na¨ıve Bayes and SVM might be more suitable for real-time detection due to their lower computational requirements. Understanding these trade-offs is essential for selecting the appropriate model based on specific application needs [7], [11].
3) Practical Applications: The expected improvements in detection accuracy and computational efficiency will have practical applications in enhancing the security of IoT networks. These findings can inform the development of real-time DDoS detection systems that can be deployed in various IoT environments, from smart homes to industrial IoT networks [9].
4) Future Research Directions: The study's results will lay the groundwork for future research in several areas. For instance, exploring additional feature selection methods, integrating deep learning techniques, and testing the models on other IoT datasets can further advance the field. Moreover, real-world implementation and testing of the proposed models will be critical for validating their effectiveness in diverse IoT settings [10], [12].
5) Challenges and Limitations: Despite the expected positive outcomes, there are potential challenges and limitations to consider. The variability in IoT device capabilities and network configurations can affect the generalizability of the models. Additionally, the evolving nature of DDoS attack techniques requires continuous updates to the detection models to maintain their efficacy [11].

In summary, this study aims to develop a robust and efficient DDoS detection framework for IoT networks. The expected results indicate significant improvements in detection accuracy, precision, recall, and computational performance. By addressing the challenges and leveraging the strengths of different machine learning models and feature selection techniques, this research contributes to the advancement of IoT security and provides a foundation for future work in this critical area.

V. Conclusion

In this study, we have developed a comprehensive framework for detecting Distributed Denial of Service (DDoS) attacks in Internet of Things (IoT) networks using advanced machine learning techniques and effective feature selection methods. The proposed methodology leverages the CICIoT2023 dataset, Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), and a variety of machine learning models, including Random Forest, Support Vector Machine (SVM), Na¨ıve Bayes, XGBoost, and K-Nearest Neighbors (KNN).The expected results from our methodology indicate significant improvements in detection accuracy, precision, recall, and computational efficiency. By reducing the dimensionality of the dataset and focusing on the most relevant features, the proposed framework enhances the performance of machine learning models, making them more suitable for real-time DDoS detection in resource-constrained IoT environments. The use of k-fold cross-validation ensures the robustness and generalizability of the models, which is critical for their deployment in diverse IoT settings [7], [9].

A. Summary of Key Findings

- Improved Detection Accuracy: Advanced machine learning models, particularly Random Forest and XGBoost, are anticipated to achieve high accuracy in distinguishing between normal and malicious traffic [7], [9].
- Efficient Computational Performance: Incorporating feature selection techniques reduces the computational load on the models, which is crucial for IoT environments with limited resources [11].
- Robustness and Generalizability: The use of k-fold crossvalidation ensures that the models perform consistently well across different subsets of data, indicating their reliability in real-world scenarios [12].

B. Implications and Future Work

The findings from this study have significant implications for the field of IoT security. The demonstrated improvements in detection accuracy and computational efficiency can inform the development of real-time DDoS detection systems that are deployable in various IoT environments, from smart homes to industrial networks. The insights gained from comparing different machine learning models provide valuable information on their relative strengths and weaknesses, guiding the selection of appropriate models for specific applications [7], [11].Future research can build on this work by exploring additional feature selection methods, integrating deep learning techniques, and testing the models on other IoT datasets. Realworld implementation and testing of the proposed models will be critical for validating their effectiveness in diverse IoT settings. Additionally, ongoing updates to the detection models will be necessary to address the evolving nature of DDoS attack techniques and maintain their efficacy [10], [12].

C. Challenges and Limitations

While the proposed methodology offers promising improvements, there are potential challenges and limitations to consider. The variability in IoT device capabilities and network configurations can affect the generalizability of the models. Furthermore, the evolving nature of DDoS attack techniques requires continuous updates to the detection models to ensure their ongoing effectiveness [11].This study contributes to the advancement of IoT security by developing a robust and efficient DDoS detection framework. By leveraging advanced machine learning techniques and effective feature selection methods, the proposed methodology enhances the detection accuracy and computational efficiency of DDoS attack detection in IoT networks. The findings provide a foundation for future research and practical applications, ultimately contributing to the development of scalable and effective security solutions for the rapidly expanding IoT landscape.

References

[1] Statista, *Internet of things (iot) connected devices installed base worldwide from 2015 to 2025*, Available at: https://www.statista.com/statistics/471264/iot-numberof-connected-devices-worldwide/, 2021.

[2] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, “Ddos in the iot: Mirai and other botnets,” *Computer*, vol. 50, no. 7, pp. 80–84, 2017.

[3] Symantec, *Iot devices attacked every two minutes on average*, Available at: https://www.symantec.com/blogs/threat-intelligence/iot-attacks, 2020.

[4] CrowdStrike, *Crowdstrike 2024 global threat report*, Available at: https://www.crowdstrike.com/globalthreat-report/, 2024.

[5] SonicWall, *2024 sonicwall cyber threat report*, Available at: https://www.sonicwall.com/2024-cyber-threatreport, 2024.

[6] W. U. Hassan, A. Bates, and D. Marino, “Tactical actions for mitigating iot threats: A case study in iot forensics,” in *Proceedings of the 27th Network and Distributed System Security Symposium (NDSS)*, 2020.

[7] R. Doshi, N. Apthorpe, and N. Feamster, “Machine learning ddos detection for consumer internet of things devices,” in *Proceedings of the IEEE Security and Privacy Workshops (SPW)*, 2018.

[8] *IoT Dataset 2023 — Datasets — Research — Canadian Institute for Cybersecurity — UNB — unb.ca*, https: / / www. unb. ca / cic / datasets / iotdataset - 2023 . html, [Accessed 11-06-2024].

[9] N. Moustafa and J. Slay, “Unsw-nb15: A comprehensive data set for network intrusion detection systems (unsw-nb15 network data set),” in *Proceedings of the 2019 Military Communications and Information Systems Conference (MilCIS)*, 2019, pp. 1–6.

[10] V. Hassija, V. Chamola, and S. Zeadally, “A comprehensive review of machine learning techniques for ddos attack detection and mitigation in iot networks,” *Journal of Network and Computer Applications*, vol. 170, p. 102813, 2020.

[11] K. Fu, T. Kohno, and D. F. Kune, “Security and privacy for iot devices,” *Proceedings of the IEEE*, vol. 108, no. 3, pp. 401–403, 2020.

[12] E. Alomari, M. Aldwairi, and O. Batarfi, “A hybrid model for ddos attack detection and mitigation,” *Journal of Network and Computer Applications*, vol. 170, p. 102814, 2020.

[13] H. Zhang, X. Li, and L. Chen, “A hybrid rule-based machine learning approach for iot anomaly detection,” *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8371–8381, 2020.

[14] J. Lee and J. Lee, “Reinforcement learning-based adaptive ddos attack mitigation in software-defined iot networks,” *IEEE Access*, vol. 8, pp. 91667–91678, 2020.

[15] H. He, S. Yan, and Y. Zhang, “A federated learning framework for ddos detection in iot,” *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 1131–1141, 2020.

[16] L. Breiman, “Random forest,” *Machine Learning*, 2020.

[17] X. Wang and Z. Yang, “Distributed inference for linear support vector machine,” *Journal of Machine Learning Research*, vol. 20, pp. 1–41, 2019.

[18] I. Rish, “An empirical study of the naive bayes classifier,” *IJCAI*, vol. 22, pp. 41–46, 2019.

[19] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2019.

[20] L. E. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, p. 1883, 2019.