Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 10s (2025)

A Two Fold Machine Learning Approach to Prevent and Detect IoT Botnet Attacks

Miss. Usha Ravsaheb Dongare[1], Prof. Dr. Monika Rokade[2], Prof. Dr. Sunil Khatal[3]

[1]Department of Computer Engineering, Sharadchandra Pawar College of Engineering, Dumbarwadi, Otur, Pune, India

ushadongaremd@gmail.com

[2]Department of Computer Engineering, Sharadchandra Pawar College of Engineering, Dumbarwadi, Otur, Pune, India

monikarokade4@gmail.com

[3]HOD, Department of Computer Engineering, Sharadchandra Pawar College of Engineering, Dumbarwadi, Otur, Pune, India

khatalsunils88@gmail.com

Article History:

Received: 12-01-2025

Revised: 15-02-2025

Accepted: 01-03-2025

Abstract:

The frequency and diversity of cyberattacks are escalating owing to the ongoing proliferation of the Internet; ransom ware incidents are rising, % 0 day vulnerabilities are gaining substantial media coverage. A robust network needs numerous tiers of security measures, since antivirus software and firewalls alone are inadequate. An IDS is a crucial layer that continuously monitors the system to protect its target from potential attacks. The rapid expansion of the Internet of Things (IoT) has yielded considerable advantages, enabling improved connection and automation across several industries. The proliferation of IoT devices has rendered them more susceptible to assaults, particularly IoT botnet attacks, which use insecure devices to create extensive disruptions. This study introduces a dual machine learning approach aimed at the prevention and real-time detection of IoT botnet assaults. During the preventative phase, methods for finding anomalies are used to look for strange patterns in network traffic that could mean there is an active botnet. This protects devices against possible threats. In the detection phase, classification techniques, such as machine learning, analyze real-time data to identify active botnet assaults and implement prompt countermeasures. We evaluated the proposed model with benchmark IoT botnet datasets and found it to be very successful in detecting and mitigating botnet attacks. Our research shows that using a two-phase machine learning method makes IoT networks much safer by constantly watching them and changing their defenses to deal with new botnet threats. This study gives us important information for making IoT security solutions that are strong and scalable so that we can protect important infrastructures from the growing botnet threat.

Keywords: IDS, NIDS, Machine learning, Convolutional Neural Network, Deep Learning.

I. INTRODUCTION

The rapid expansion of network usage over recent decades has introduced significant security challenges for the Internet and computer systems. Malicious intrusions or attacks on networks can result in severe consequences. Intrusions are harmful entities responsible for network attacks that

compromise system integrity, confidentiality, and availability. In such scenarios, systems often fail to respond effectively to data theft or loss. To address this, intrusion detection systems (IDSs) are vital for minimizing the adverse effects of these attacks. An IDS is a system or software tool designed to identify unauthorized access to networks or computer systems. Malicious exploits, vulnerabilities, data-driven threats, and host-based intrusions like privilege violations, sensitive file access, unauthorized logins, and malware are just some of the types of attacks it can find.

To safeguard IoT networks from botnet attacks, we propose a dual-phase machine learning-based strategy comprising prevention and detection. During the prevention phase, algorithms that look for anomalies find strange network traffic or strange device behavior that could be signs of threats. This proactive approach aims to reduce vulnerabilities before attacks occur. In the detection phase, machine learning classification algorithms analyze network activity in real time, categorizing it as either benign or malicious. This two-phase model uses the best features of both anomaly detection and classification techniques to protect IoT systems in a strong and complete way.

II. LITERATURE SURVEY

In [1] Analyze these datasets to showcase the suggested model's capacity to accurately detect and classify harmful nodes. The looks into deep learning architecture for finding unauthorized access and shows how well the system model works on three important datasets for network attacks. The findings showed that the model worked well with the CSE-CIC-IDS2018, Newcastle University, and NSL-KDD datasets. We also looked at the outcomes for NSL-KDD, UNSW-NB15, & CSE-CIC-IDS2018 to see how well the model worked with different datasets and how stable it was in general.

According to [2] they contrast both machine learning and deep learning models for detecting network intrusions based on anomalies. It starts by looking at past work on ML and DL intrusion detection systems (IDS). Next, it examines the datasets utilized in those previous studies. Additionally, we tested ML and DL models using the KDD-99 dataset, presenting, analyzing, and discussing their performance results. The paper concludes by highlighting critical areas for future research and exploration.

According to [3] an in-deep learning-based intrusion detection system. The system has the ability to withstand adverse attacks. The system uses the maximum method to produce adversarial examples that generate high loss and attack deep neural networks. A design investigates and assesses the effectiveness of harmful attack methods and the resistance of combatively trained models to such attacks. The system suggests that we can use adversarial attack methods created for binary domains in continuous environments, which could lead to different levels of misclassification.

According to [4] a thorough experimental research based on different binary is given to improve detection frequency and reduce the error. Furthermore, numerous studies of intrusion detection systems have been conducted using old datasets such as the Kddcup'99 dataset. Because older datasets didn't reflect how attacks are done now, most models weren't able to find modern intrusions as accurately as they could have. The goal of this project is to create a hybrid anomaly-based intrusion detection system (IDS) that uses the newest dataset, "CICIDS2017," to test intrusion detection and either deep learning (DL) or binary algorithms (BA) as optimizers. Cost error, confusion matrix, sensitivity, specificity, recall, and accuracy are some of the key outcomes that the DNN and its hybrid implementation have been looked at and tested.

According to [5] present a new architecture for deep neural networks to develop adaptable and efficient intrusion detection models. This method has two parts: an unsupervised phase for extracting features from multiple channels, and a supervised phase for looking at how those features relate to each other across channels. This test's goal is to see if class-specific network flow characteristics can be added to basic features to make the model more accurate. Each day, we train two autoencoders separately and execute attacks on the flows during the unsupervised phase. This makes it easier to make two extra feature vectors that turn each network flow into a multi-channel sample. The suggested neural network architecture surpasses current intrusion detection frameworks regarding prediction accuracy when assessed on three benchmark datasets.

According to [6] multiple sensor nodes, randomly dispersed throughout specified regions for data collection, comprise Wireless Sensor Networks (WSNs). Given their location in exposed and rough areas, these nodes are particularly susceptible to assaults, underscoring the need for a powerful intrusion detection system (IDS). Machine learning (ML) techniques have become more important in improving intrusion detection system (IDS) solutions, especially for wireless sensor networks (WSNs) that use less energy. When the Dragonfly Algorithm (DA) is used to make the Optimal Multilayer Perceptron (OMLP) model work better, it can find intrusions in Wireless Sensor Networks (WSNs). The primary objective of the OMLP model is to detect possible intrusions and precisely categorize their characteristics. This technique enables the DA to optimize the weights and biases of the MLP, hence simplifying the efficient selection of connection weight values. This improvement significantly improves detection performance, thereby augmenting the overall efficacy of the IDS.

According to [7] CNN \& RNN are two deep learning methods used to create an intelligent detection system to identify various network intrusions. Additionally, the system evaluates the proposed solution's performance using several evaluation matrices. The system compares the results of the suggested solution to choose the optimum model for the network intrusion detection system.

According to [8] Large volumes of log data often conceal the malicious activities of insider threats, allowing them to remain undetected. This survey focuses on insider threat detection using deep learning and provides a comprehensive exploration of the issue. It starts by explaining what insider threat detection and related ideas are and why deep learning is better than traditional machine learning methods for dealing with large amounts of complex data from many sources. The article delves into the role of deep learning and log-based anomaly detection in this domain. It compiles datasets relevant to insider threat detection with a particular focus on the CERT Insider Threat Dataset and offers a comparative analysis of existing deep learning techniques applied to the problem. Additionally, the study discusses the challenges faced in this field and how they have shaped opportunities for future research. It shows how a deep learning model can be added to the Elasticsearch-Logstash-Kibana (ELK) stack, bridging the gap between academic and practical views. This gives readers interested in business-related applications new ideas.

According to [9] Nonetheless, because to its dispersed and open designs, which are vulnerable to incursions, privacy & security problems are important hurdles in the effective adoption of cloud computing. Cloud computing's open and dispersed features are more tempting to prospective hackers. Because of the openness of cloud computing, traditional intrusion detection technologies are often

useless. This research investigates the use of deep learning to implement innovative intrusion detection systems that use a believe adaptive security model for intrusion detection.

According to [10] used many IDS methodologies throughout time to get optimal detection accuracy. Machine-learning intrusion detection systems are among the most promising techniques for identifying both known and novel threats. This article (NIDS) examines the many machine learning methods used in the implementation of network-based intrusion detection systems.

III. PROPOSED SYSTEM DESIGN

Our objective is to create and implement a deep learning-based methodology for the IDS. We recognize the intrusion detection system for its elevated incidence of false alarms. We are continuously making efforts to reduce the high false positive rate. We hold the belief that intrusion detection functions as a data analysis process, and we can examine it as a problem of accurately classifying data. From this perspective, we observe that the effectiveness of any classification scheme depends on the quality of the data it receives as input. When the data is more pristine, the likelihood of obtaining accurate results increases. From an anomaly-based IDS perspective, effectively extracting features that distinguish normal from abnormal data can significantly decrease the false positive rate. To fully protect an intrusion detection system, it is necessary to use firewalls and other intrusion prevention tools along with the intrusion detection system. In this study, we used a machine learning method on the pre-processed data to develop a learning model, which was then validated using the whole Botnet dataset. Finally, we calculated the accuracy, detection rate, and false alarm rate to determine the machine learning model's detection efficiency.

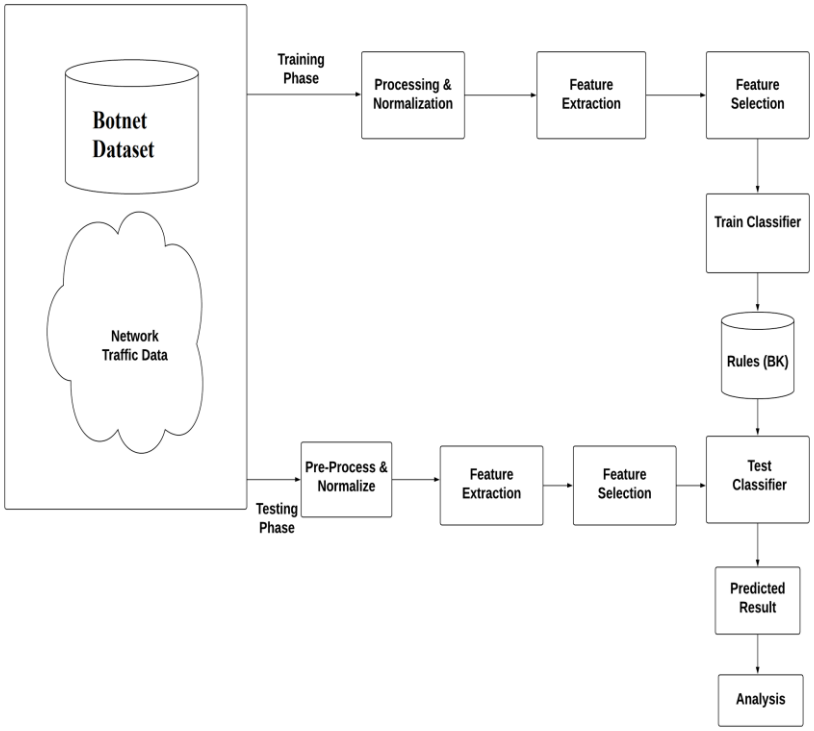

Figure 1: Proposed System Design

Implement Module

Data Collection: We collect data from multiple sources, such as the UCI ML repository, Kaggle, and various real-time data streams. Before performing classification tasks, the data needs to undergo pre-processing to improve results by handling missing values and eliminating redundant features present in the selected dataset. For optimal outcomes during the data mining (DM) process, it is crucial to process the dataset efficiently.

Pre-processing and normalization: The quality and organization of input data are essential for attaining high predictive accuracy in machine learning. This project uses a lot of different pre-processing and normalization methods to make sure that the data fed into the machine learning models is clean, standardized, and good for analysis. This makes it easier to find and stop IoT botnet attacks.

Data Cleaning: The first stage is addressing any absent or compromised data entries in the dataset. This method involves finding and substituting or eliminating records with missing information to avert model errors.

Normalization: We conduct normalization to ensure that all characteristics contribute equally to the model. Network traffic data often includes information with very disparate scales, such as packet size and frequency counts. We use min-max normalization to adjust these characteristics to a range of [0, 1], preventing any one feature from overshadowing others due to its magnitude.

Data Splitting: We partition the data into training, validation, and testing sets after preprocessing and standardizing it. This division facilitates a precise evaluation of model efficacy and guarantees that the model does not over fit to particular data. We use a conventional split ratio of 70% for training and 30% for testing to ensure a fair assessment.

Feature extraction & Selection: Given the potentially large number of variables in IoT network datasets, selecting the most relevant features is essential for efficient and accurate model training. Statistical and correlation-based techniques are used to analyze feature importance, helping to retain only the most significant features that contribute to detecting botnet activities. Feature selection reduces computational complexity and improves model performance by removing irrelevant or redundant data.

Classification: The classification step is essential to the dual machine learning methodology, concentrating on identifying active botnet assaults in the IoT network dataset. During this phase, machine learning algorithms evaluate real-time network traffic data to categorize behaviour as either benign or malicious. This categorization facilitates the swift mitigation of botnet assaults, thereby ensuring improved IoT network dataset security. This research assessed several classification algorithms to establish the most efficient model for detecting botnet assaults.

RESULT AND ANALYSIS

The other machine learning classifiers achieved varying degrees of accuracy across the different feature sets. Ultimately, all three alternative classifiers outperformed ours when applied to the string feature set. Despite this, our combined classifier achieved the highest overall accuracy.

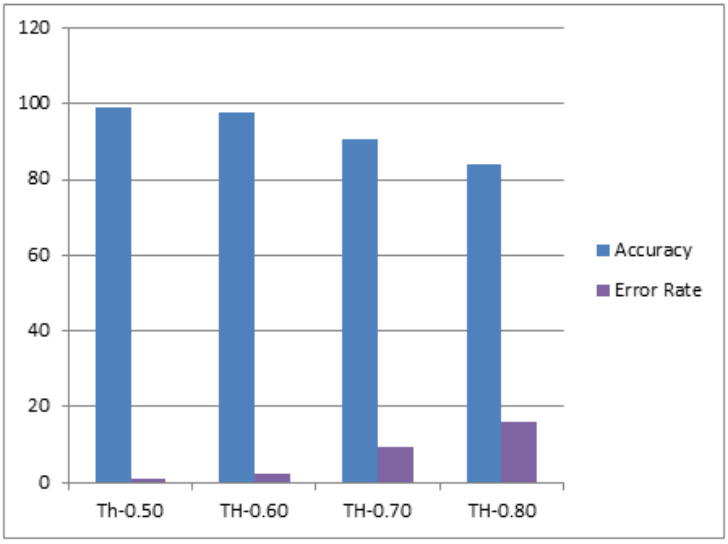

Figure 1.2: Number of unknown attack detections by the system with various thresholds

In Figure 1.2, we had done the same experiment, which is already verified, showing the data set that was used to measure the detection accuracy of the proposed system with the same algorithms. The above figure illustrates how resetting different threshold values can alter classification accuracy.

CONCLUSION

Since intrusion detection research gained significant attention in the security community around a decade ago, a variety of approaches have emerged to address this challenge. Intrusion detection systems differ in the data sources they utilize and the techniques they apply to analyze it. Today, most systems classify data using either abuse detection or anomaly detection, each with its own strengths and weaknesses. Achieving flawless detection, much like perfect security, is unrealistic due to the complexity and rapid evolution of modern systems. Upon completing this survey, we conclude that multiple methods exist for identifying attacks, with both soft computing and classification approaches proving effective. Numerous systems have focused on signal-based anomaly detection, developing diverse rules to facilitate this process. For our research, we employed the Botnet dataset for training and testing purposes.

REFERENCES

[1] Amaizu, Gabriel Chukwunonso, et al. "Investigating Network Intrusion Detection Datasets Using Machine Learning." 2020 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2020

[2] Abdel-Wahab, Mohab Sameh, Ahmed M. Neil, and Ayman Atia. "A Comparative Study of Machine Learning and Deep Learning in Network Anomaly-Based Intrusion Detection Systems." 2020 15th International Conference on Computer Engineering and Systems (ICCES). IEEE, 2020

[3] Abou Khamis, Rana, M. Omair Shafiq, and Ashraf Matrawy. "Investigating Resistance of Deep Learning-based IDS against Adversaries using min-max Optimization." ICC 2020-2020 IEEE International Conference on Communications (ICC). IEEE, 2020.

[4] Atefi, Kayvan, Habibah Hashim, and Touraj Khodadadi. "A Hybrid Anomaly Classification with Deep Learning (DL) and Binary Algorithms (BA) as Optimizer in the Intrusion Detection System (IDS)." 2020 16th IEEE International Colloquium on Signal Processing \& Its Applications (CSPA). IEEE, 2020.

[5] Andresini, Giuseppina, et al. "Multi-channel deep feature learning for intrusion detection." IEEE Access 8 (2020): 53346-53359.

[6] Amaran, Sibi, and R. Madhan Mohan. "An Optimal Multilayer Perceptron with Dragonfly Algorithm for Intrusion Detection in Wireless Sensor Networks." 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2021.

[7] Ashraf, Javed, et al. "Novel deep learning-enabled lstm autoencoder architecture for discovering anomalous events from intelligent transportation systems." IEEE Transactions on Intelligent Transportation Systems (2020).

[8] Al-Emadi, Sara, Aisha Al-Mohannadi, and Felwa Al-Senaid. "Using deep learning techniques for network intrusion detection." 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT). IEEE, 2020.

[9] Al Makdi, Khalid, Frederick T. Sheldon, and Abdullah Abu Hussein. "Trusted Security Model for IDS Using Deep Learning." 2020 3rd International Conference on Signal Processing and Information Security (ICSPIS). IEEE, 2020.

[10] Ahmed, Lubna Ali Hassan, and Yahia Abdalla Mohamed Hamad. "Machine Learning Techniques for Network-based Intrusion Detection System: A Survey Paper." 2021 National Computing Colleges Conference (NCCC). IEEE, 2021.