



Program Title: Master of Science in Artificial
Intelligence

**Real-Time IoT Botnet Detection Using Machine
and Deep Learning for Scalable Security**

A Final Proposal
Submitted in Partial Fulfillment of Course Research
Methodology in Computer Science MSc. Degree
Requirements for Master of Science in Artificial
Intelligence

Prepared By
Omar AbuAmra 242001458

Supervised By
Dr. Mohammed Anbar

2024

TABLE OF CONTENTS

	Page
TABLE OF CONTENTS	II
Abstract	III
Chapter I	1
INTRODUCTION	1
1.1 Introduction	1-2
1.2 Background	2-4
1.3 Problem Statemnt	4-5
1.4 Research objective	5-7
1.5 Research scope	7-8
Chapter II	9
LITERATURE REVIEW	9
2.1 literature Review Detection System for IOT Botnet Attacks	9-15
Chapter III	16
Research methodology	16
3.1 Research methodology	16-18
3.1.1 Expected outcomes	19
3.1.2 Gantt chart	20
References	21

Abstract

Internet of Things (IoT) Devices have created new security challenges, especially with regard to protecting against botnet assaults. This study filled in the gaps in the IoT environment's current (IDS). We created a scalable, flexible IDS system leveraging microservices architecture and created an extensive dataset that reflected various IoT botnet attack patterns. Principal component analysis and autoencoder-based dimensionality reduction, two sophisticated feature optimization approaches, were used. We conducted an assessment of different machine learning models, which encompassed ensemble methods and deep learning techniques, with the aim of detecting in real-time. An implementation of federated learning was used to train models collaboratively while respecting anonymity. The benchmark dataset that was produced exceeded the diversity and real-world applicability of existing datasets. The framework we developed showed enhanced adaptability in deployment and increased efficiency in resource utilization, resulting in reduced computational demands while preserving detection accuracy. The implementation of real-time detection resulted in minimal delay, while the collaborative approach significantly enhanced the early identification of botnets in their first stages. These results contribute to the improvement of detecting and preventing IoT botnets, benefiting both academic research and industrial practices by providing more reliable, adaptable, and immediate IoT security solutions.

Keywords: Internet of Things (IoT), Botnet attacks, Intrusion Detection System (IDS), Machine learning, Deep learning, Federated learning, Microservices architecture, Feature optimization, Real-time detection, Cybersecurity.

Chapter I

INTRODUCTION

1.1 Introduction

The Internet of Things (IoT) (IoT) has transformed numerous sectors, including smart homes, industrial automation, and healthcare. Nevertheless, the widespread use of IoT devices has brought about substantial security weaknesses, with botnet attacks emerging as a crucial and imminent danger. These attacks leverage the inherent vulnerabilities of IoT devices, presenting significant threats to both individual users and large-scale deployments. (Jithu et al., 2021).

Although modern intrusion detection systems (IDS) that utilize machine learning and deep learning approaches demonstrate potential, there are still some significant deficiencies. Current methods frequently depend on inadequate datasets that do not encompass the complete spectrum of botnet attack behaviors observed in real-world scenarios (Nguyen et al., 2022). Furthermore, the ability to grow across extensive IoT networks and the capacity to identify events in real-time pose considerable difficulties (Al-Haija & Al-Dala'ien, 2022; Kadhim et al., 2022).

This study seeks to overcome these constraints by creating a sophisticated intrusion detection system specifically designed for Internet of Things (IoT) scenarios. The aims that guide our study are as follows: Generate an extensive dataset that encompasses a wide range of Internet of Things (IoT) botnet assault behaviors. Create a scalable system for detecting unauthorized access in IoT networks of significant size. Enhance the feature selection techniques to attain superior accuracy and efficiency in detection. Develop and integrate real-time detection capabilities into the proposed framework. Evaluate the effectiveness of various machine learning and deep learning models. Implement and utilize a collaborative learning strategy to quickly detect and prevent cyber attacks from IoT botnets.

Our methodology include the development of datasets, construction of frameworks, enhancement of features, assessment of models, and implementation of collaborative

learning techniques. We utilize the latest advancements in deep learning models and employ the federated learning technique to overcome their inherent constraints.

Our goal is to make major progress in the field of IoT botnet detection and mitigation by accomplishing these objectives. The anticipated results consist of a standardized dataset for comparison, a flexible and adaptable intrusion detection system framework, refined methods for selecting the most effective features, and improved skills to identify threats in real-time. These contributions will offer strong, adaptable, and capable security solutions for IoT networks, fulfilling a crucial requirement in both academic research and industry application.

The remainder of this paper is structured as follows: Chapter II provides a comprehensive literature review, Chapter III details our methodology, Expected outcomes and Gantt chart.

1.2 Background

Offering creative ideas and improved efficiencies across sectors including healthcare, smart homes, and industrial automation, the Internet of Things (IoT) has transformed many different businesses. The IOT use is fast rising, and the possibilities for transforming uses are expanding dramatically. But this growth of IoT devices has also brought major security flaws, especially in the shape of botnet attacks.

Often with limited processing power, memory, and energy, IoT devices are resource-constrained. They work in different surroundings, which makes consistent security policies challenging to carry out. Moreover, a lot of IoT devices lack strong security systems, which makes them simple targets for hackers. By using these weaknesses, botnets build networks of hacked devices under remote control capable of launching major attacks.

Botnet Attack Mirai One excellent illustration of the terrible power of IoT botnets is the Mirai botnet attack of 2016. Mostly security cameras and routers, this hack compromised over 600,000 IoT devices to execute a large DDoS attack briefly shutting down CNN, Twitter, and Netflix. The hack made clear how urgently strong security policies are needed in Internet of Things networks. (Antonakakis et al., 2017) To offset these risks, several intrusion detection systems (IDS) using machine learning (ML) and deep learning (DL) approaches have been created. To find hostile

activity, these systems examine network traffic, device behavior, and other indicators. Studies on ML and DL models—such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs)—have revealed that they can efficiently spot anomalies and possible hazards in Internet of Things systems. Still, current remedies may include restrictions, including Many IDS depend on particular datasets that might not fully reflect the range of botnet attack techniques found in actual IoT environments.

IoT networks might include hundreds, maybe millions of devices. Current IDS systems could not be able to scale efficiently.

IoT data's high dimensionality calls for effective feature selection methods to guarantee accurate and fast processing.

Many of the present IDS systems lack real-time detection features, so postponing the threat identification and neutralizing power.

IoT networks clearly call for more strong and scalable intrusion detection solutions. Dealing with the above described constraints calls for creating scalable frameworks, complete datasets, best feature selection techniques, and real-time detection systems. This work attempts to close these gaps by improving IoT security using cutting-edge ML and DL methods.

The literature discuss the limitations in Few datasets completely reflect the variety of botnet attack patterns observed in actual IoT systems (Popoola et al., 2021; Srinivasan & P, 2023)

Many times, lacking scalability, current systems call for the creation of modular and distributed systems (Al-Haija & Al-Dala'ien, 2022; Kadhim et al., 2022; Wang et al., 2021)

The reduction in computational load depends on effective feature selection. Although there is need for more improvement, current studies show several feature selection methods (Alshamkhany et al., 2020; Idrissi et al., 2021; Nguyen et al., 2022)

Although some research look at real-time capabilities, low-latency processing and immediate threat reduction still remain very elusive (Al-Haija & Al-Dala'ien, 2022; Alshamkhany et al., 2020; Jithu et al., 2021; Rezaei, 2021)

IoT security policies and procedures might be greatly changed by this study. More efficient intrusion detection systems will help IoT networks to be more resilient generally and build more confidence in IoT technologies. Wider acceptance of IoT solutions across sectors resulting from this could stimulate innovation and economic

growth by means of industry-specific relevance. Furthermore, our results might guide the creation of fresh IoT security guidelines and best practices, so strengthening the digital environment and ensuring its safety.

All things considered, the background of this study emphasizes the great demand for sophisticated security solutions in Internet of Things systems. This work intends to provide a complete intrusion detection system that greatly increases the security of IoT devices against botnet attacks by solving the gaps in dataset variety, scalability, feature optimization, and real-time detection. This project will support the larger field of IoT security by offering strong, scalable, real-time competent solutions that can be generally accepted among several IoT applications.

By means of this research, we hope to not only improve the technical features of IoT security but also help to build a more trustworthy IoT ecosystem. The value of this work in preserving our digital future cannot be emphasized as IoT keeps invading all spheres of our life, from smart homes to crucial infrastructure.

1.3 Problem Statement

The Current proposal using advance machine learning and deep learning methods to detect and reduce botnet attacks in IoT networks.

Because of the varied nature of IoT devices and their limited computational capacity, sophisticated botnet assaults are becoming more and more common on IoT networks. These assaults pose serious risks to individual users, large-scale IoT deployments, and national security since they cause major disruptions, security breaches, and the possible penetration of key infrastructure. Even with advances in intrusion detection systems (IDS), real-time detection and mitigation of botnet attacks is still a major problem, especially in diverse and dynamic IoT contexts. The rapid growth of IoT devices in a variety of fields, including smart cities, healthcare, and industrial systems, amplifies the possibility of harm that botnet attacks might cause. For IoT networks to function in a safe, dependable, and trustworthy manner a prerequisite for the developing Internet of Everything (IoE) effective detection and mitigation of these threats are not just desired, but also necessary. It is predicted that IoT security breaches would have a billiondollar yearly economic impact, therefore reinforcing how critical it is to have strong defenses.

Existing methodologies for identifying botnet assaults in IoT networks encompass an intrusion detection system utilizing deep learning is employed to detect DDoS botnet attacks in IoT networks. This system achieves exceptional accuracy and precision by employing a deep neural network model. (Jithu et al., 2021). An ensemble learning model is employed to detect botnets in IoT environments. This model achieves a remarkable accuracy rate of 100% and significantly reduces the number of required features to just 20. It utilizes a combination of supervised, unsupervised, and regression learning approaches. (Rezaei, 2021). BotIDS is an advanced intrusion detection system that utilizes deep learning algorithms to effectively identify and detect botnet assaults in IoT networks. It has been proven to outperform previous methods, with an impressive validation accuracy of 99.94%. (Idrissi et al., 2021). A cooperative machine learning model was developed to detect IoT botnets at an early stage. The model achieved an accuracy of 99.37% when tested on a dataset consisting of 5023 botnets and 3888 benign samples. (Nguyen et al., 2022). The ELBA-IoT ensemble learning model surpasses current cutting-edge techniques for identifying botnet assaults in IoT networks, attaining a remarkable accuracy of 99.6% while incurring a minimal inference overhead of 40 as. (Abu Al-Haifa and Al-Dalanie, 2022). An

ensemble deep transfer learning approach was used to detect IoT malware, resulting in an accuracy of 93.58% on a dataset consisting of 3000 malicious and 3000 benign [samples](#). (Kadhim et al., 2022).

Despite these developments, the present approaches fall lacking in several critical areas. insufficient ability to change or adjust in the face of novel or unusual conditions. The quickly changing nature of botnet attacks and the wide variety of IoT devices and protocols frequently present difficulties for current designs. Restricted scalability: Several proposed solutions face major difficulties when scaled to cover large, heterogeneous IoT networks. Limited real-time capabilities the fluid nature of IoT environments makes real-time detection and reaction essential, yet current solutions generally fall lacking in this regard. There is not enough representation in the dataset. A large number of models are trained using small datasets that come up short of capturing the richness and diversity of actual IoT botnet attacks.

1.4 Research objective

Research that focuses primarily on developing machine learning models that can scale and adapt to meet the diversity and dynamic nature of IoT networks is desperately needed in order to address these serious shortcomings. creating and utilizing larger, more inclusive datasets that faithfully represent real-world IoT botnet attack

scenarios. developing machine learning models that can scale and adjust to the many and ever-changing IoT network conditions. Improving feature extraction and selection techniques to offer effective real-time response and detection capabilities. Algorithms that are lightweight and efficient can be developed so that they can be employed with high detection accuracy on Internet of Things devices with limited resources. To propose and Assure diversity and applicability to actual situations by compiling an extensive dataset that encompasses a broad spectrum of IoT botnet attack patterns. The diversity required to train models that can accurately identify the entire range of IoT botnet attacks is frequently lacking from current datasets. The accuracy and dependability of detection algorithms will be increased dramatically with a larger dataset. To design and construct a scalable intrusion detection system that may be used to successfully implement resource limitations and heterogeneous device management across large-scale IoT networks. A scalable solution that can safeguard networks of different sizes and compositions is required due to the quick expansion and diversity of IoT deployments, which highlights a significant weakness in present security methods. To develop and analyze new methods for feature extraction and selection that balance detection accuracy and computational efficiency in IoT devices with limited resources. One major problem in existing IoT security systems is providing real-time detection on resource-limited IoT devices, which requires efficient feature selection. To develop and incorporate real-time detection features into the suggested architecture to guarantee prompt detection and handling of botnet threats in ever-changing Internet of Things settings. Real-time detection is essential for successful security because of the dynamic nature of IoT networks and the possibility for botnet infections to spread quickly. This solves a major flaw in many existing systems. To evaluate how well different deep learning and machine learning models including Deep Learning hybrid approaches work to improve intrusion detection systems' efficiency and accuracy in identifying threats on Internet of Things networks. Pushing the limits of existing IoT security capabilities, comparative study of various models and hybrid method exploration can result in more effective detection strategies.

To design and use a collaborative learning approach to identify and isolate threats early in their lifespan, protecting privacy and cutting down on communication overhead, in order to enable the early detection of IoT botnet attacks. Reducing the effects of botnet assaults requires early discovery. A critical gap in IoT security can

be filled by collaborative techniques, which can make use of distributed IoT systems to increase detection speed and accuracy. The main obstacles in IoT botnet detection are intended to be methodically addressed by this study framework, which covers everything from data representation to model deployment and evaluation. By accomplishing these goals, we hope to: Improve IoT security by creating botnet detection systems that are more precise, effective, and flexible. Boost the overall security posture of IoT ecosystems by offering workable solutions that can be integrated into actual IoT deployments. Provide significant methods and insights to the cybersecurity community at large, especially when addressing situations with varying resources. Create a foundation for next studies on intelligent and adaptable IoT security systems.

1.5 Research scope

This work uses advanced machine learning and deep learning techniques to improve the identification and mitigating of botnet assaults in IoT networks. The main goal is to solve important problems in present systems such limited dataset diversity, scalability problems, poor feature selection, and absence of real-time detection capacity. Developing a thorough dataset spanning a broad spectrum of IoT botnet attack behaviors, produced by means of data augmentation techniques, public sources, and data collecting from many IoT contexts, is central focus of this work. Among the several machine learning approaches used in the study will be feature selection and optimization techniques include LASSO regression, Recursive Feature Elimination (RFE), and Principal Component Analysis (PCA). With an eye on using collaborative learning techniques, especially federated learning, several models—including convolutional neural networks (CNNs), recurrent neural networks (RNNs), random forests, and gradient boosting machines—will be reviewed. The approach consists in the construction of a scalable, modular intrusion detection framework, application of real-time data streaming and processing pipelines, feature selection and optimization techniques, evaluation of several machine learning and deep learning models, and implementation of a collaborative learning method for early identification. Both controlled and real-world settings will see extensive validation and testing. Accuracy, precision, recall, F1-score, ROC-AUC, real-time detection latency, and scalability performance in big-scale IoT networks will be among the several

evaluation measures used in the research. Different machine learning models will be compared in a thorough comparison study, evaluated against current IoT botnet detection technologies, and the effect of feature selection on model performance examined.

Improving several important areas—enhancing dataset diversity and representation, improving scalability for large-scale IoT networks, optimizing feature selection for better performance and efficiency, developing real-time detection capabilities, and applying collaborative learning for better early detection—is especially emphasized in this paper. This work intends to greatly increase IoT botnet detection by addressing these key elements, thereby offering more efficient, scalable, real-time capable security solutions for the always growing terrain of IoT networks.

Chapter II

LITERATURE REVIEW

2.1 literature Review Detection System for IOT Botnet Attacks

This study paper presents a deep learning-based intrusion detection method to catch IoT botnet episodes, especially DDoS episodes. The writers developed a Deep Neural Network (DNN) model using the BoT-IoT dataset, which includes both normal and pretend attack traffic from a real IoT network circumstances. The dataset was preprocessed and normalized utilizing min-max scaling. The DNN model was trained and assessed employing various metrics, including accuracy, precision, recall, F-score, and AUC-ROC curve. The outcomes deliver high performance, with a precision of 0.947 and an AUC of 0.98. The key findings show that the offered DNN model outperforms existing techniques in detecting IoT botnet attacks with elevated accuracy and precision. Two major restrictions of this study are that The focus is primarily on DDoS attacks, potentially neglecting different classes of IoT botnet episodes, and The model's performance in real-world, dynamic IoT environments may differ from the simulated dataset used for training and testing(Jithu et al., 2021)

This study aimed to develop an ensemble learning model for optimizing botnet detection accuracy on IoT machines while underrating required features. The experimenters mixed artificial neural networks (ANN) and decision trees (DT) after considering 17 device learning algorithms. They employed a dataset of 990,517 domain names (790,745 standard, 199,772 malicious) with 204 features, which was decreased to 20 essential features. The ensemble approach reached 100% accuracy in catching botnets and bots, with a true positive rate of 20.17% and a true negative rate

of 79.83%. The main constraints were that the study concentrated exclusively on domain name-based detection, potentially ignoring other kinds of botnet activity. The model's interpretation was not pushed on real-time network gridlock, which could introduce further challenges in live circumstances (Rezaei, 2021).

This study presents a deep learning-based intrusion detection method called BotIDS to notice botnet attacks in IoT networks. The experimenters operated convolutional neural networks (CNNs), corresponding to performance in other deep learning techniques like RNN, LSTM, and GRU. They employed the Bot-IoT dataset, which includes about 3.6 million network traffic logs affecting an IoT domain with normal traffic and ten classes of botnet episodes. The CNN model gained 99.94% guarantee accuracy and 0.58% validation loss, with forecast time beneath 0.34 ms. The key results are that the CNN-based approach surpassed other deep learning techniques for this IoT intrusion detection mission. The primary restrictions noted are that The study only used one dataset for evaluation, and Real-world testing in an actual IoT environment was not conducted (Idrissi et al., 2021).

This study paper presents a collaborative machine-learning method for premature detection of IoT botnets. The authors design a model that integrates numerous features from technique calls, network manners, and machine performance to notice malicious movement in the premature phases of the botnet lifecycle. The model is assessed on a 5023 IoT botnet illustrations dataset and 3888 benign samples, achieving 99.37% accuracy. Key results contain the significance of using a collaborative late fusion approach to combine different feature sets and the capability to notice hurtful behavior earlier than full-time series data methods. The primary constraints are The dependency on features collected through a custom sandbox domain, which may restrict generalizability to real-world IoT devices—an

insufficiency of comparison to other state-of-the-art IoT botnet detection methods to contextualize the performance gains (Nguyen et al., 2022).

This research paper suggests ELBA-IoT, an ensemble learning model for noticing botnet attacks in IoT networks. The examination utilizes the N-BaIoT-2021 dataset, incorporating 611,359 regular traffic and 7,737 samples of botnet traffic (4,737 Bashlite and 3,000 Mirai) from five IoT appliances. The writers utilize three decision tree approaches (AdaBoosted, RUSBoosted, and bagged) and assess their ensemble model employing different performance metrics. The results show that ELBA-IoT achieves a high detection accuracy of 99.6% with lower inference overhead (40 μ -seconds), exceeding other state-of-the-art methods. The primary constraints of this study are The dependence on a single dataset, which may restrict the model's generalizability to other IoT environments, and The absence of real-time testing in an existing IoT network environment, which could deliver more insights into the model's functional performance (Al-Haija & Al-Dala'ien, 2022) .

This research document offers a novel method for IoT malware detection employing ensemble deep transfer learning. The approach preprocesses binary malware data into grayscale images and employs two pre-trained CNN models (ResNet101 and VGG16) as feature extractors. The pulled features are combined and provided into a classifier. The study utilized a dataset of 6,000 samples (3,000 malicious and 3,000 benign) gathered from two IT firms in Iraq. The suggested method reached 93.58% accuracy, 94.39% precision, and 92.67% recall on the test set. The primary benefits are the benefit of transfer learning to optimize training with narrow data and the model's competitive performance connected to earlier works, particularly regarding inference time (0.012 seconds). Nevertheless, two fundamental limitations are the moderately small dataset size corresponding to some other studies in the field, which may restrict

generalizability, and the absence of comparison with more recent state-of-the-art malware detection approaches, which could deliver a better complete evaluation of the presented approach's effectiveness (Kadhim et al., 2022).

This work provides an ensemble defensive mechanism termed Def-IDS to guard deep learning-based network intrusion detection systems from adversarial attacks. The two-module strategy the researchers devised combines multi-class generative adversarial networks (MGAN) with multi-source adversarial retraining (MAT). Using the CSE-CIC-IDS2018 dataset—which comprises seven classifications of intrusions across 8.9 million records and regular network traffic—they found On hostile cases, the ensemble model exceeded individual MGAN (71.3%) and MAT (92.9%) modules with 97.9% detection accuracy. Key findings reveal that, while preserving great accuracy on clean samples, the combined strategy efficiently resists known and undiscovered adversarial attacks. The key restrictions observed are that real-world testing in operational networks was not carried out and the evaluation was done on a single dataset. Furthermore, dynamic intrusion detection systems could be concerned about the computing cost of the ensemble process (Wang et al., 2021).

The paper presents a deep learning method for IoT network botnet attack detection: SMote-DRNN. In order to solve the class imbalance in network traffic data, the researchers merged Synthetic Minority Oversampling Technique (SMote) with Deep Recurrent Neural Networks (DRNN.). Comprising around 73 million network flow records with normal traffic and 10 varieties of botnet attacks, their Bot-IoT dataset Outperformance of the baseline DRNN model and other state-of- the-art machine

learning methods, the SMote-DRNN model obtained 100% overall accuracy, 99.50% precision, 99.75% recall, and 99.62% F1-score. Important results show that SMote-DRNN is more robust against underfitting and overfitting as well as on minority attack classes. Two primary constraints observed are higher computational cost and training time than in DRNN by itself. Furthermore, even while performance was assessed on a test set, genuine IoT environment deployment was not carried out (Popoola et al., 2021).

The paper presents an Ensemble Classifier Algorithm with Stacking Process (ECASP) for cybersecurity applications to detect botnet attacks. In order to enhance classification performance, the researchers stacked several machine learning classifiers. They made use of the Cyber Clean Center (CCC) dataset, which features network traffic information for HTTP and IRC ports. Outperformance of the ECASP model above baseline approaches including Extreme Learning Machine (ELM), Convolutional Neural Network (CNN), and Support Vector Machine (SVM) was 94.08% accuracy, 86.5% sensitivity, 85.68% specificity, and 78.24% F-measure. Among important results are ECASP's better performance than individual classifiers on several evaluation criteria. The primary restrictions observed are that the study concentrated mostly on classification performance without any investigation of computational scalability or efficiency. Furthermore, testing in real-world network contexts was not done even when the model was assessed on benchmark datasets (Srinivasan & P, 2023).

The paper presents a machine learning method based on several categorization techniques to identify botnet attacks. Four classifiers—Naïve Bayes, K-Nearest Neighbors, Support Vector Machine, and Decision Trees—were assessed by the researchers. They used the 2.5 million network traffic records in the UNSW-NB15 dataset, with 43 attributes denoting both normal and different attack kinds. The Decision Tree model performed best among 82,000 randomly chosen records with 99.89% testing accuracy, 100% precision, 100% recall, and 100% F-score in identifying botnet attacks. Important results include the efficiency of feature selection applied under Chi-squared criteria, therefore lowering the feature set to 29 for best performance. The primary acknowledged restrictions are that testing was not done in a real-time network context and only a subset of the complete dataset was employed. Furthermore, more advanced methods like deep learning were not investigated even as several classifiers were compared (Alshamkhany et al., 2020).

The combined aim of these research papers is to introduce state-of-the-art machine learning and deep learning model-related studies that can be utilized for IoT botnet attack detection and intrusion detection so that more efficient Intrusion Detection System (IDS) design in an IoT network could occur. The studies used datasets like combined IoT and simulated network traffic Bot-IoT, which introduced benchmarks in the reviewed approaches performance with metrics such as accuracy, precision, recall (or True Positive Rate), F1 score, etc. They also provide ROC curves of the mentioned metrics for methods under study, all to prove the feasibility of the studied techniques proposed. Their papers generally say that their use of deep learning and ensembles furthers the state-of-the-art in detecting IoT botnet attacks compared to Traditional methods, with reasonable accuracy and performance metrics. Although not explicitly mentioned, these works are extensions of the IoT security research and deal with

drawback contexture within customary intrusion detection strategies for IoT networks. The underlying assumption is that machine learning and deep learning (more formally called "end-to-end" training with inference), in particular, can learn narrow patterns in network traffic, implying normal vs. malicious behavior of an IoT device or group thereof! These include multiple models such as Deep Neural Networks (DNN), ensemble learning techniques, and transfer learning methods. This paper's strengths were presenting the recent IoT datasets and exploring the inherent security challenges in a precisely specified domain. However, some of the potential limitations may be overfitting to specific datasets, or it can get computationally complex for resource-limited IOT devices and, in a few cases, has never been tested in real life. Instead, these papers offer positive measures for enhancing IoT security using elaborate machine learning methods. However, more research regarding their practicality in real-world implementation and resource efficiency is required to corroborate the conclusions.

Chapter III

Research methodology

:

3.1 Research methodology

This work aims to improve the identification and mitigating of botnet attacks in IoT networks by means of modern machine learning and deep learning methods. More especially, we want to solve the following difficulties. Limited dataset diversity in current IoT botnet detection systems, Scalability problems in vast-scale IoT networks, ineffective feature selection methods, Lack of real-time detection capabilities.

To suggest a complete dataset including a broad spectrum of IoT botnet attack trends. Emphasizing several kinds of botnet assaults, gather varied datasets from many IoT environments and publically accessible sources. Use synthesis and data augmentation to guarantee a wide variety of attack strategies are represented. Normalize, sanitize, and encode the data such that it fits machine learning models. Apply anomaly detection, correlation analysis, and distribution analysis among other statistical analysis methods to describe the dataset.

To provide a scalable intrusion detection system for big Internet of Things networks. Create a distributed and modular system capable of managing big amounts of data from different IoT devices. Share the computational load using resources from cloud and edge computing. Apply effective data transmission techniques to guarantee flawless integration among several layers of a network. With particular hardware needs including high-performance GPUs for model training and low-power edge devices for dispersed detection, use a hybrid cloud-edge architecture. In order to maximize intrusion detection feature selection techniques. To find the most pertinent features use LASSO regression, Recursive Feature Elimination (RFE), and Principal Component Analysis (PCA). Examine how computational efficiency and model performance change with feature selection. Iteratively test and refine to get the best feature set. Estimate feature relevance using statistical measures of information acquisition and mutual information. To enable inside the suggested framework real-time detection

capability Utilizing Apache Kafka and Spark Streaming, apply real-time data streaming and processing pipelines leveraging technologies Include real-time detecting machine learning models into these pipelines. Guarantee low-latency processing and instantaneous response systems will help to quickly address found hazards. Find temporal trends in network traffic using methods of time series analysis.

to assess several models of machine learning and deep learning for their potency. Train and test several models including CNNs, RNNs, and ensemble learning techniques like Random Forests and Gradient Boosting Machines. Compare models using performance benchmarks like accuracy, precision, recall, F1-score, and ROC-AUC. Cross-validation guarantees generalizability and robustness. Examine model performances using statistical hypothesis testing. using a collaborative learning approach to early IoT botnet attack detection.

Create a collaborative learning system allowing several IoT devices to exchange knowledge and thereby enhance detection accuracy together. Apply federated learning methods to protect data privacy and support group model development. Test the framework in simulated IoT systems to evaluate early threat detection capability.

to guarantee the suggested fixes are sensible and useful. In controlled environments and actual IoT installations, do thorough validation and testing. Run many assault simulations to assess the intrusion detection system's dependability and strength. Get comments from domain experts and iteratively improve the models and framework depending on testing results.

Two main study groups—the research will comprise Controlled Environment Group: Initially tested and validated IoT devices established in a lab environment. b) Real-World Deployment Group: For real-world testing, a range of IoT networks from many sectors—e.g., smart homes, industrial IoT—are selected.

This study will follow rigorous ethical standards; all gathered information will be anonymized and encrypted. Any dataset utilized will have personal identifiable information deleted. Informed permission will be secured from every person or entity engaged for actual testing. Strict security policies will be followed to guard all data and

systems engaged in the research. Before publication, any vulnerabilities found during the investigation will be responsibly shared to relevant stakeholders.

We shall create a specific IoT testbed in our lab for the regulated environment. To have access to various IoT environments for real-world testing, we will form alliances with IoT device producers and network operators. To guarantee appropriate access and data management, all required legal agreements including non-disclosure clauses will be set in place.

Infrastructure for Cloud Computing: Use cloud services (such as AWS or Google Cloud) for high-performance computing requirements and data storage. To replicate various IoT scenarios, set up a variety of edge devices—such as Raspberry Pi, NVIDIA Jetson. Configure both wired and wireless networks to reflect different IoT topologies. Analyze data and model performance statistically using R and Python's SciPy. Framework of Machine Learning: Implement and examine machine learning models with TensorFlow and PyTorch. Processing and analyzing vast amounts of data with Apache Hadoop and Spark is big data tool. Biases could be introduced into the synthetic data producing process. We will thus constantly improve our data generating methods and check our synthetic data against actual samples to help to reduce this. Particularly large-scale IoT networks could provide difficulties for the architecture. We shall handle this using hierarchical detection methods and adaptive sampling strategies. Models might fit particular kinds of attacks too much. We will hence use strict cross-valuation and test on several attack situations to counteract this.

3.1.2 Expected outcomes

A thorough and varied dataset reflecting several IoT botnet attack trends will improve the detection model resilience.

an edge computing and cloud-based scalable intrusion detection system able to manage vast IoT networks. a strategy of feature selection best suited for botnet identification that raises computational efficiency and model performance.

Integrated into the architecture, real-time detection power's rapid reaction to found hazards. Comparative study of several machine learning and deep learning models finds the best methods for IoT botnet detection. a collaborative learning method maintaining data privacy while raising detection accuracy among several IoT devices.

Results of controlled environment and real-world deployment testing validate the usefulness and simplicity of the suggested answers. Learnings about early botnet attack detection via virtual IoT systems. a collection of best practices and suggestions for using botnet detection grounded on machine learning in various IoT systems.

possible finding of fresh trends or features of IoT botnet attacks by means of thorough dataset analysis. better knowledge of the trade-offs among IoT network detection accuracy, processing needs, and model complexity. Identification of any restrictions or difficulties in implementing the suggested approaches to actual IoT systems, hence guiding next studies. These results would be very beneficial for IoT security since they provide both theoretical progress and pragmatic ways to identify and stop botnet attacks in intricate IoT environments.

3.1.3 Gantt Chart



References

- Al-Haija, Q. A., & Al-Dala'ien, M. (2022). ELBA-IoT: An Ensemble Learning Model for Botnet Attack Detection in IoT Networks. *Journal of Sensor and Actuator Networks*, 11(1). <https://doi.org/10.3390/jsan11010018>
- Alshamkhany, M., Alshamkhany, W., Mansour, M., Khan, M., Dhou, S., & Aloul, F. (2020). Botnet Attack Detection using Machine Learning. *Proceedings of the 2020 14th International Conference on Innovations in Information Technology, IIT 2020*, 203–208. <https://doi.org/10.1109/IIT50501.2020.9299061>
- Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., Durumeric, Z., Alex, J., Luca, H., Michalis Kallitsis, I. •, Kumar, D., Lever, C., Ma, Z., Mason, J., Menscher, D., Seaman, C., Sullivan, N., Thomas, K., & Zhou, Y. (2017). Understanding the Mirai Botnet.
- Idrissi, I., Boukabous, M., Azizi, M., Moussaoui, O., & Fadili, H. El. (2021). Toward a deep learning-based intrusion detection system for iot against botnet attacks. *IAES International Journal of Artificial Intelligence*, 10(1), 110–120. <https://doi.org/10.11591/ijai.v10.i1.pp110-120>
- Jithu, P., Shareena, J., Ramdas, A., & Haripriya, A. P. (2021). Intrusion Detection System for IOT Botnet Attacks Using Deep Learning. *SN Computer Science*, 2(3). <https://doi.org/10.1007/s42979-021-00516-9>
- Kadhim, Q. K., Al-Sudani, A. Q. A. S., Almani, I. A., Alghazali, T., Dabis, H. khalid, Mohammed, A. T., Talib, S. G., Mahmood, R. A., Sahi, Z. T., & Mezaal, Y. S. (2022). IOT-MDEDTL: IoT Malware Detection based on Ensemble Deep Transfer Learning. *Majlesi Journal of Electrical Engineering*, 16(3), 47–54. <https://doi.org/10.52547/mjee.16.3.47>
- Nguyen, G. L., Dumba, B., Ngo, Q. D., Le, H. V., & Nguyen, T. N. (2022). A collaborative approach to early detection of IoT Botnet. *Computers and Electrical Engineering*, 97. <https://doi.org/10.1016/j.compeleceng.2021.107525>
- Popoola, S. I., Adebisi, B., Ande, R., Hammoudeh, M., Anoh, K., & Atayero, A. A. (2021). Smote-drnn: A deep learning algorithm for botnet detection in the internet-of-things networks. *Sensors*, 21(9). <https://doi.org/10.3390/s21092985>
- Rezaei, A. (2021). Using Ensemble Learning Technique for Detecting Botnet on IoT. *SN Computer Science*, 2(3). <https://doi.org/10.1007/s42979-021-00585-w>
- Srinivasan, S., & P, D. (2023). Enhancing the security in cyber-world by detecting the botnets using ensemble classification based machine learning. *Measurement: Sensors*, 25. <https://doi.org/10.1016/j.measen.2022.100624>
- Wang, J., Pan, J., Alqerm, I., & Liu, Y. (2021). Def-IDS: An Ensemble Defense Mechanism against Adversarial Attacks for Deep Learning-based Network Intrusion Detection. *Proceedings - International Conference on Computer Communications and Networks, ICCCN, 2021-July*. <https://doi.org/10.1109/ICCCN52240.2021.9522215>