DISTRIBUTED DENIAL OF SERVICE ATTACK DETECTION AND PREVENTION MODEL FOR IOT-BASED COMPUTING ENVIRONMENT USING ENSEMBLE MACHINE LEARNING APPROACH

Nicholas Oluwole Ogini[1], Wilfred Adigwe[2] and Noah Oghenefego Ogwara[3]

[1]Department of Computer Science, Delta State University, Abraka, Nigeria
[2]Department of Computer Science,
Delta State University of Science and Technology, Ozoro, Nigeria
[3]Department of Computer Science and Software Engineering, Auckland University of Technology, Auckland, New Zealand

ABSTRACT

Defending against Distributed Denial of Service (DDoS) in the Internet of Things (IoT) computing environment is a challenging task. DDoS attacks are type of collective attack in which attackers work together to compromise internet security and services. The resource-constrained devices used in IoT deployments have made it even easier for an attacker to break, because of the vast number of vulnerable IoT devices with significant compute power. This paper proposed an ensemble machine learning (ML) model using the bagging technique to detect and prevent DDoS attacks in the IoT computing environment. We carried out an Machine Learning experiment and evaluated our proposed model with the most recent DDoS attacks (CICDoS2019) dataset. We use seven validation metrics (classification accuracy, precision rate, recall rate, f1-score, Matthews Correlation Coefficient, false negative rate and false positive rate) to evaluate the performance of the proposed model. The results obtained in our experiment shows an improved performance with an overall maximum classification accuracy of 99.75%, precision rate of 99.99%, recall rate of 99.76%, f1-score of 99.87%, Matthews Correlation Coefficient of 0.000000214, false negative rate of 0.24% and 4.42% false positive rate.

KEYWORDS

Internet of Things, DDoS Attacks, Ensemble Machine Learning, Security, & Detection Models.

1. INTRODUCTION

One of the most visible breakthroughs in computing technology over the last decade has been the Internet of Things (IoT). This technology enabled the integration of computer power and network communication capabilities into a wide range of devices, allowing end-users to access new types of services. The utilisation of IoT devices has changed multiple fields such as healthcare, industry, physical security, farming, and even home automation [1].

Significant cyber security implications have resulted from this transition in computing. The issue of protecting the internet from cyber-attacks is growing increasingly difficult. One of the main reasons for this is that the number of connected devices has increased, making it impossible to secure them all due to their variety of types. IoT devices are characterised by a wide range of hardware and software, as well as insecure design, a lack of upgrades, and user engagement [2].

However, when the network grows, the obstacles grow as well. The security difficulties surrounding IoT have a significant impact on the domain's future, raising concerns about the security of devices in use. Meanwhile, A distributed denial of service (DDoS) attack is one of the most prevalent security issues affecting Internet of Things (IoT)-based applications and devices [3].

The infrastructure that interconnects devices operating in an IoT-based environment requires maximum cooperation to tackle security issues because of the rapid growth of connected devices [3]. DDoS attacks are type of collective attack in which attackers work together to compromise internet security and services. In this type of attack, the attacker takes advantage of compromised systems to deny legitimate users access to server resources and then uses those resources to launch a series of attacks against the victim [4]. DDoS attacks have become increasingly common in the cyber security world. DDoS attacks have expanded dramatically since everything became connected through the Internet, because there are now more devices to compromise and launch attacks [4].

The resource-constrained devices used in IoT deployments have made it even easier for an attacker to break, because of the vast number of vulnerable IoT devices with significant compute power, attackers seeking to compromise these devices and exploiting them to establish large-scale botnets against their victims. A botnet is a collection of infected devices or bots, sometimes known as zombies, that share a command-and-control infrastructure and are used for nefarious purposes such as DDoS attacks. These DDoS attacks were expected for various reasons such as the heterogeneous nature of IoT devices in terms of communication medium and protocols, platforms, and devices, IoT systems are extremely diverse., and therefore pose a bigger security risk than traditional computing systems. [5].

All these factors have heightened research interest in the field of IoT security in recent years. The Internet of Things (IoT) environment is made up of several devices with varying levels of traits and capabilities. It includes anything from high-end computers to low-cost microprocessors with limited memory and processing power. Security solutions must be proposed at many levels in this diverse environment. Because the capacities of devices at different levels of the IoT vary, security measures used at each level will have varied dimensions and features [6]. To solve some of these problems associated with DDoS attacks on IoT devices, this paper proposed DDoS detection framework using the ensemble machine learning (ML) techniques to detect attacks in application layer of the IoT-based computing environment. The contribution of the paper is as follows:

- We proposed an ensemble machine learning model using the bagging technique to effectively detect different types of DDoS attacks in IoT computing environment in a timely manner.
- Our proposed model can identify the type of DDoS attacks and initiate a mitigation approach in preventing the attacks.
- The result obtained from our experiment shows an improved performance when compared with similar methods in extant literature

The remaining section of the paper is organized as follows: section 2 discussed some of the related works and the proposed model is presented in section 3. Section 4 discuss the methodology use in the study which includes the data collection, data pre-processing and machine learning experimental work carried out in this study. The results of the experiments and comparison with related works in extant literature are presented in section 5 and finally section 6 concludes the paper and highlight direction for further research.

2. RELATED WORKS

This section reviews the state of the arts research work that proposed a model or framework to detect DDoS attacks in various network environment. The selected papers used in this review on recent works that applies either machine learning (ML) or deep learning (DL) approach in their experimental work with the most up to date state of the art dataset for the effective detection of attacks in a network environment. This enable the study to identify some of the current issues that needs to be addressed. In addition several studies has been conducted to solve security issues concerns with malicious network traffics that causes DDoS attacks in a network environment for example Smys et al. [7] proposed a hybrid IDS that detects DDoS attacks in IoT environments. The model proposed by the authors applies a convolutional neural network algorithm that uses a bidirectional long-term memory architecture for the training its model. However, the sample size of normal (240) and attack (3,890) types in the dataset used for their experiment is very small. The authors reported 98.60% classification accuracy from the experiment conducted in their work. One of the advantage of work reported in [7] is the performance of the detection model in detecting intrusion in IoT environments but the proposed model suffers from its inability to identity each type of attacks and applies appropriate counter measures to a detected attacks in the IoT environment.

Singh-Samom and Taggu [8] proposed a model to detect four types of DDoS attacks using a machine learning (ML) approach. The authors conducted experiments in their study with different ML algorithms using the CICDDoS2019 and UNSW-NB15 DDoS attack datasets. The results of their experiment shows that the RandomForest ML classifier outperforms other ML algorithms. The RandomForest classifier was adopted as the classifier for the model proposed in their work. In addition, the authors evaluated their model using four performance metrics and reported a classification accuracy of 99.92% using CICDDoS2019 dataset and 96.20% using UNSW-NB15 dataset. One of the advantages reported by the authors in [8] is the ability of their detection model to distinguish between multiple types of DDoS attacks. Although the work in [8] has no counter security measures for detected attacks and no clear description the proposed model.

Elsayed et al [9] proposed an IDS to detect DDoS attacks in software defined networking (SDN) environment using deep learning approach. The authors use Recurrent Neural Network (RNN) with autoencoder in the training of their model using the CICDDoS2019 DDoS attacks dataset. The authors also reported 99% classification accuracy from the results obtained from the evaluation of their model. The advantage of the work reported in [9] eliminates the complexity of model training but suffers from high computational requirements during the encoding process and only apply binary classification approaches in its detection engine.

Wei et al [10] proposed a model to detect DDoS attacks using a deep learning approach. The authors combine two deep learning techniques for effective detection and classification of DDoS attacks. First, the authors proposed a feature extraction model that uses autoencoder to extract relevant features that is require for the detection process, and secondly a model that uses Multi-layer Perceptron Network (MLP) to classify DDOS attack types. In addition, the authors reported a classification accuracy of 98.34% based on the results obtained from the evaluation carried out in their work. One of the advantages of the work proposed in [10] its ability to detect different types of attacks, the authors apply a multi-classification approach rather than a binary approach proposed by majority of the related works. However, their model requires high computational resources during training.

Kousar et al [11] proposed a decision tree-based model to detect DDoS attacks in SDN environment. The proposed model focus on attacks that is used to flood the SDN controller. The

authors perform an experiment to detect different types of DDoS attacks using the CIC-DDoS 2019 dataset and also compares the results of the experiment conducted by collecting simulated data from SDN environment using Mininet emulator and RYU controller using different DDoS tools. The authors reported a classification accuracy of 99.99% using the decision tree approach compared to the other ML approach used in their experiment. One of the advantages of the model reported in [11] its ability to detects attacks in a software defined environment which is much complex compares to the IoT environment. Although the work in [11] only apply the binary classification in its detection engine. In addition, the resource computational requirements in software defined environments are higher compared to IoT environments.

Malliga et al[12] proposed a model to detect DDoS attacks in the application, network, and transport layers of the OSI model. The model proposed by the authors uses the Deep Neural Network (DNN) and Long Short-Term Memory (LSTM) approaches. In addition, the authors evaluated their proposed model with CICIDS2017 and CICDDoS2019 DDoS datasets with a maximum classification accuracy of 99.40%. One of the draws back of the works in [12] is the amount of time requires for its training and no preventive measures to tackle detected attacks however, the proposed model is effective in detecting intrusions across the several layers of the open system interconnection layers.

Vuong et al [13] propose a novel multi-tier architecture IDS for detection of DDoS attacks and evaluate their proposed model with the CICDDoS2019 dataset. The authors reported a precision rate and a recall rate of 99.50% and 99.10%, respectively. One of the major advantages of the proposed model reported in [13] is the feature selection approach. Identifying key features relevant to the detection process helps improve the performance and reliability of the model. Although the complexity of the N-tier architecture is resource demanding. Rajagopal et al.

Similarly, the authors in [14] proposed a model that uses decision jungle algorithms with a meta-classification approach to detect DDoS attacks in a cloud computing environment. The experiments reported by the authors was conducted in a production ready Azure virtual machine. The authors also evaluated their proposed model with UNSW NB-15, CICID 2017, and CICDDOS2019 dataset and obtains a maximum classification accuracy of 98% and 97% using the CICDDOS2019 dataset. The feature selection approach reported in [14] was very effective to select relevant features require to build a reliable model however, the training set were less than 70% of the entire set. The performance of such model is more reliable when trained with over 70% of the dataset to reduce the issue of imbalanced data in the learning process. Also, the issue of preventive measures to counter detected attacks was not reported.

Gohil and Kumar, [15] carried out experiments with major supervised ML classification algorithms to identify the best ML models to effectively detect DDoS attacks in a network environment. The results of their experiment show that tree-based classifiers are much better than distance-based classifiers. The authors reported an overall maximum classification accuracy of 96.25% using the tree-based ML classifiers. The model reported in [15] uses a smaller number of dataset even though the proposed model has the ability to detect multiple attacks. Similarly, Shieh et al [16] proposed a DDoS attack detection model that uses a bi-directional LSTM model along with a Gaussian Mixture. The authors conducted an experiment, and they reported a maximum classification accuracy of 98.18%. One of the major advantages of the proposed model in [16] centred on the ability of the model to cope with incremental learning but the model suffers in performance from the detection of zero day attacks in this environment.

In summary, different approaches has been reported in extant literature offering solutions to the detection of DDoS attacks. However, some of the recent works proposed in literature suffers from performance issues with accurate detection and mitigation of DDoS attacks in a timely

manner. In addition, majority of the works reported in literature did not provide mitigation component in their model to manage detected attacks. Similarly, some of the works has failed to evaluated the false positive rate of attacks in various network environment to effectively measure the performance of their model. Using the classification accuracy to measure the effectiveness of a detection model gives rooms for the possibility of false alarms in the detection of attacks. There is need to build a better detection model by considering all relevant evaluation metrics to measure the performance of the future models. In addition, the use of ensemble ML approach has not been explored widely in the detection of DDoS attacks in IoT environment in extant literature and also explored other evaluation metrics to measure the performance of DDoS attacks models is necessary hence the focus of this study to explore other evaluation metrics and ensemble ML technique to build a more reliable detection model that can effectively handle large volume of network traffics in a timely manner.

3. THE PROPOSED MODEL

In this paper, we proposed a conceptualized model to detect and prevent DDoS attacks in the IoT computing environment. The complex nature of the IoT environment and its resource constrained nature make it difficult to combat security issues in this environment.

The proposed model is designed for an IoT network that connects the Internet to a personal or private network. The router at the border will be more computationally powerful and capable than end-of-line devices. Even in such a scenario, we must think about the power and execution time constraints. The internal network will communicate with the cloud-based server via the router that is connected to the internet. The proposed model applies exhaustion of resources mitigation techniques to tackle DDoS attacks in an IoT environment.

This study proposed a DDoS attack identification and prevention machine learning (ML) model using ensemble techniques as shown in Figure 1. The ML model proposed in this paper uses CICDDoS2019 dataset [17] for its training. The proposed ML model in this paper uses the ensemble bagging method. This ML technique creates several instances of a black-box estimator on random subsets of the original training set, then aggregates their unique forecasts to make a final classification. These methods are used to lower the variance of a base estimator (for example, a decision tree) by introducing randomization into the construction mechanism and then constructing an ensemble from it.

In many circumstances, bagging approaches provide a simple way to enhance with respect to a single model without having to change the underlying base algorithm. The ensemble ML model proposed in this paper uses the decision tree algorithms as its base estimator. Bagging approaches perform best with strong and complex models because they reduce overfitting.

The proposed model has a two-level DDoS attack identification module. The first level DDoS attack identification module receives each network packet along with the device ID and forwards the network packets to the ensemble ML model for inspection. In the event of the detection of abnormalities in the suspected network traffic in the first level, the ensemble ML model forward the network packets to the DDoS attack classification module.

The second level is the DDoS attack classification module is responsible for identifying the type of DDoS attack detected by the ensemble ML Model. Once the attack type is identified, the prevention module is notified and initiates the mitigation procedure of the devices that are associated with the attacks by blocking traffic from the devices. The process of blocking the suspected traffic from the device protects the device from further attacks at the same time notify the administrator of the devices for further investigation.

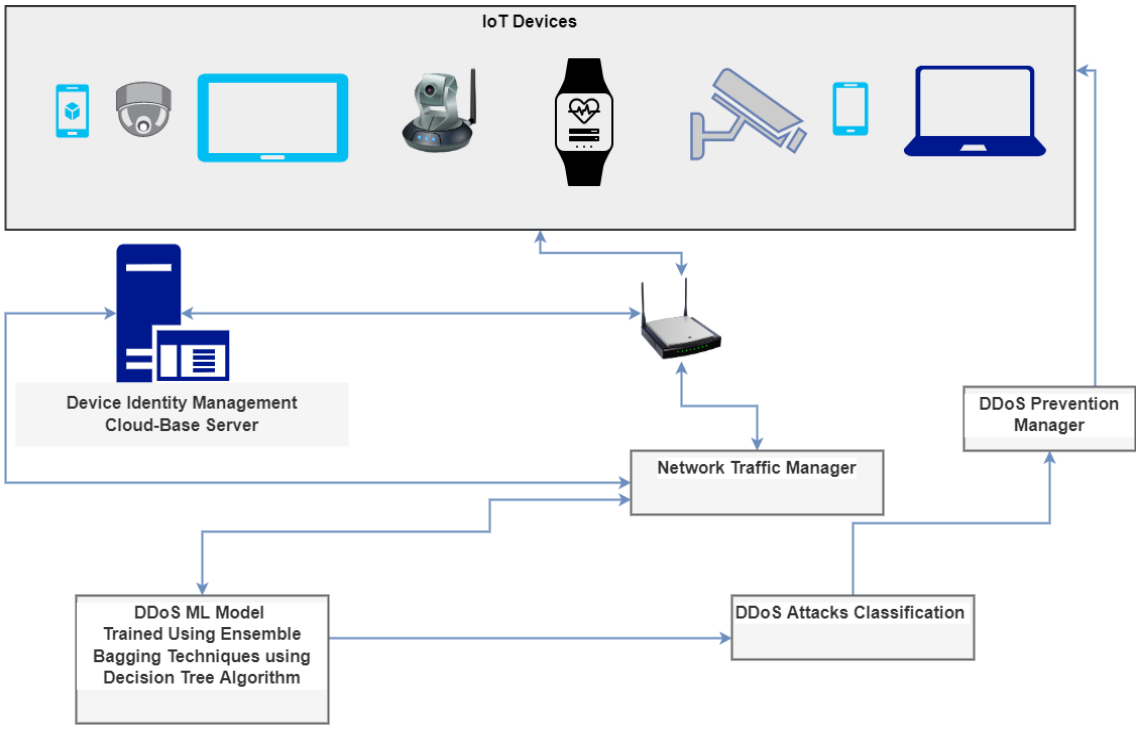

Figure 1. Proposed DDoS Attacks Detection and Prevention Ensemble ML Models

4. METHODOLOGY

This section discusses the data collection and the experimental work carried out in this study. This includes a description of the dataset used, pre-processing of collected data, the ensemble ML algorithms used, the workflow of the proposed model and a discussion of the results obtained in the experimental work.

4.1. Data Collection and Description

This study uses a benchmark DDoS attack dataset named CICDoS2019 [17] collected from a public repository at the University of New Brunswick, Canada. This dataset has been widely used for conducting research on the detection of DDoS attacks in various environments. This study used this dataset to conduct an ML experiment using the ensemble techniques and proposed a model that uses the bagging techniques for attack detection and prevention.

Overall, the dataset contains 50,063,112 records (56,863 benign traffic records and 50,006,249 malicious traffic records) in its training set and 20,364,525 records (56,965 benign traffic records and 20,307,560 malicious traffic records) in its testing set. The entire dataset contains 88 unique features such as source IP, destination IP, source port, destination port, and so on. The experimental work carried out in this study uses all DDoS attack types in its training set, which includes NTP, DNS, LDAP, MSSQL, NetBIOS, SNMP, SSDP, UDP, UDP-Lag, WebDDoS, SYN, and TFTP, However, only seven DDoS attacks are included in the testing dataset (i.e., PortScan, NetBIOS, LDAP, MSSQL, UDP, UDP-Lag, and SYN).

There are two categories of DDoS attacks in the dataset used in this study, namely the reflection-based and exploitation-based attacks, as shown in Figure 2.

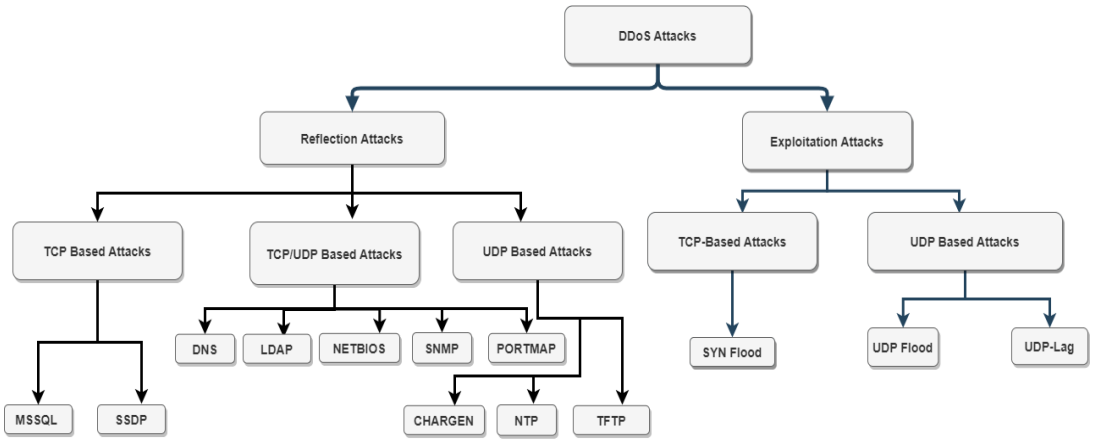

Figure 2. DDoS Attacks Categorization and Hierarchy

A. Reflection-based DDoS Attacks: In this attack category, the identity of the attacker is kept secret by leveraging a legitimate third-party component. The application layer protocol are often used in this attack type by transmitting malicious network packets to reflector servers with the source IP address set to the victim's IP address, causing the victim to be overwhelmed and send a large number of response packets. As shown in Figure 2, TCP-based attacks include MSSQL and SSDP, whereas UDP-based attacks include CharGen, NTP, and TFTP. DNS, LDAP, NETBIOS, and SNMP are examples of attacks that can be carried out using either TCP or UDP.

B. Exploitation-Based DDoS Attacks: This category of attacks also hides the identity of the attacker using third-party tools like the reflection attack. However, this category of attacks focuses on the exploitation of a specific protocol such as the network, transport, and application layers of the Open Systems Interconnection (OSI) model. The Attackers send packets to reflector servers with the source IP address configured to the victim's IP address to overwhelm the victim with response packets.

4.2. Data Pre-processing

The pre-processing of the collected data as used in this study involves the cleaning of the dataset. First, we removed some of the non-numeric features that have no contribution to the detection of attacks. In this study, we removed five features from the dataset, including Flow ID, Source IP, Destination IP, Timestamp, and Similar HTTP. These features contain values that are not numeric. On the other hand, we convert all the values to numeric in the label column, which contains the attack type. We represented each DDoS attack with the value of 1 and the value of 0 was assigned to benign labels. We also replaced all values in the dataset that contain infinity with the value of 1.79769313e+308 and all nan values were also replaced with zero.

Second, after the transformation process on our dataset, we apply the min-max normalization technique to transform the values of each attribute into a uniform distribution with the smallest value is 0 and the highest value as 1 using equation 1.

$$k \in K = \frac{k - K_{min}}{K_{max} - K_{min}}$$ where k is the value of each feature in the dataset (1).

4.3. Ensemble Machine Learning Algorithms

This section briefly discussed the five ensemble ML algorithms used in the experimental work conducted in this study.

***A. The Bagging Classifier*:** The Bagging ML classifier is an ensemble ML technique that is used for solving both classification and regression problems. This classifier is designed in such a way that it combines the output of randomly generated training set to form its final prediction. The Bagging ML classifier can use different supervised ML algorithms as its base estimators. The basic principle of this ML technique uses a group of weaker learners to form a strong learner. This technique grows as many decision trees as possible, which are combined to get the final prediction of the model.

***B. The Random Forest (RF)*:** The RF classifier is a supervised ensemble ML algorithm that can be used to solve classification and regression problems. Multiple decision trees make up the RF, which is an ensemble classifier. The trees in the RF classifier learn independently on a randomly selected part of the training data. By merging the outcomes of several learning models, the RF classifier uses bagging techniques to improve overall results. The most often recurring categories predicted by each learning model of each tree in the classifier define the RF ML classifier's output. Due to the nature of the tree, it is ideal for multi-classification problems, and it is also very effective for binary classification problems.

C. C. The AdaBoost Classifier: The AdaBoost Classifier is a supervised ML classification technique that uses an ensemble of machine learning algorithms. This algorithm is used to turn a weak classifier into a strong one. The algorithms work on the idea that each learner is developed consecutively, with the exception of the first learner, who is grown from the prior learner.

D. D. The Gradient Tree Boosting Classifier: Gradient Tree Boosting, also known as Gradient Boosted Decision Trees (GBDT), is a generalisation of boosting to arbitrary differentiable loss functions. GBDT is an off-the-shelf approach that may be used to solve regression and classification problems in a wide range of fields, including Web search ranking and ecology.

E. E. The Ensemble Voting Classifier: The Voting Classifier's concept is to combine conceptually diverse machine learning classifiers and predict class labels using a majority vote or the average predicted probability (soft vote). A classifier like this can be useful for balancing out the flaws of a group of models that are all doing well. The projected class label for a given sample is the class label that represents the majority (mode) of the class labels predicted by each classifier in majority voting. Soft voting, as opposed to majority voting (hard voting), returns the class label as the maximum of the sum of predicted probabilities. The weights option can be used to assign specific weights to each classifier. The predicted class probabilities for each classifier are gathered, multiplied by the classifier weight, and averaged where weights are available. The class label with the highest average probability is then used to create the final class label.

4.4. Experimental Setup

We carried out our ML experiment on a computer with the following hardware configuration: Intel (R) Core (TM) i7-8700 CPU @3.20GHz, 32GB RAM, and 1TB HD. The experiment carried out in this study was implemented using Python 3.7.1 and the scikit-learn ML library. We used the 6 DDoS attack types (PortScan, NetBIOS, LDAP, MSSQL, UDP, and SYN) in the testing dataset to evaluate our proposed model and repeat the same experiment with other

ensemble ML models. Below is a brief description of the DDoS attack types used in our experiment:

A. ***LDAP DDoS Attack***: In this DDoS attack, the attacker uses an application layer protocol called Lightweight Directory Access Protocol (LDAP) to send requests to a publicly available but insecure LDAP server to produce huge responses in such a way that the attacker can obtain human readable URLs.
B. ***NetBIOS DDoS Attacks***: In this DDoS attack, an attacker exploits the Network Basic Input/Output System (NetBIOS) by sending spoofed "Name Release" or "Name Conflict" signals to a victim system to block all NetBIOS network communication.
C. ***UDP DDoS Attack***: A UDP flood attack sends a large number of UDP packets to random ports on the target machine. As a result, the network's available bandwidth is depleted, the system crashes, and performance suffers. As a result, the target server's firewall may become overburdened.
D. ***SYN DDoS Attack***: The SYN DDoS attack takes advantage of the TCP three-way handshake by sending a large number of SYN packets to the target system until it crashes or malfunctions.
E. ***MSSQL DDoS Attack***: In this attack, an attacker exploits vulnerabilities in Microsoft Structured Query Language (MSSQL) by impersonating a valid MSSQL client and sending scripted requests to the MSSQL server using a faked IP address that appears to be originating from the target server.
F. ***PortMap DDoS Attack***: In this attack type , victims are bombarded with responses from Portmapper servers, overwhelming bandwidth and rendering websites and web-based services inaccessible.

4.5. Evaluation Metrics

This study uses the confusion matrix to validate the performance of the proposed ensemble ML model for DDoS attack classification in IoT computing environment and other different ensemble ML classifiers in the experiment carried out in this study. The confusion matrix is defined as follows:

True Positive (TP): The total number of DDoS attacks that were classified correctly.

True Negative (TN): The total number of benign traffics that were classified correctly.

False Negative (FN): The total number of DDoS attacks that were incorrectly classified as benign traffics.

False Positive (FP): The total numbers of benign traffics that were incorrectly classified as DDoS attacks.

To evaluate the performance of the proposed model and other ensemble ML classifiers used in the experiment, the following metrics were adopted for this study.

A. ***Classification Accuracy (CA):*** This is the total percentage of the correctly classified DDoS attacks and benign traffics in the dataset.

$$CA = \frac{TP+TN}{TP+FP+TN+FN} \times 100 \quad (2)$$

B. ***Precision Rate (PR):*** This is the total percentage of correctly classified results of all DDoS attacks that belongs to the benign labelled in the dataset.

$$PR = \frac{TP}{TP+FP} \times 100 \quad (3)$$

C. ***Recall Rate (RC):*** This is the total percentage of DDoS attacks that are correctly predicted as DDoS attacks in the dataset.

$$RC = \frac{TP}{TP+FN} \times 100 \quad (4)$$

D. ***False Positive Rate (FPR):*** This is the total percentage ratio of DDoS attacks classified wrongly to the actual numbers of the DDoS attacks samples in the dataset.

$$FPR = \frac{FP}{FP+TN} \times 100 \quad (5)$$

E. ***False Negative Rate (FNR):*** This is the total percentage ratio of benign traffics classified wrongly to the actual numbers of the benign traffic samples in the dataset.

$$FNR = \frac{FN}{FN+TP} \; X \; 100 \quad (6)$$

F. ***F1-Score (F1):*** This is the harmonic mean of the proposed classifier which is obtainable from the value of both PR and RC.

$$F1 = 2 \times \frac{PR \times RC}{PR + RC} \quad (7)$$

G. ***Matthews Correlation Coefficient (MCC).*** This metrics is unaffected by the unbalanced datasets issue and can be used to evaluate the performance of the proposed model.

$$MCC = \frac{(TP.TN) - (FP - FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (8)$$

5. PERFORMANCE EVALUATION

The performance evaluation of this study is discussed in the following subsections as follows:

5.1. Results Obtained from Our Experiment

The experiment conducted in this study uses six datasets of different DDoS attack types. The distribution records of the number of samples of each benign and DDoS traffic used in the experiment is shown in Table 1. The experimental results obtained in this study are shown in Table 2.

Table 1. Distribution of the Dataset Used in Our Experiment

Dataset Used	**No of Benign Traffic Samples**	**No of DDoS Traffic Sample**
LDAP	4,070	1,044,505
NETBIOS	84	1,048,491
PORTSMAP	4,734	186,960
SYN	30,661	1,017,915
UDP	2013	1,022,170
MSSQL	3,979	243,392

Table 2. Results Obtained from Our ML Experiment

DDoS Attack Dataset Used	**Classifier**	**TP**	**FP**	**TN**	**FN**	**CA**	**PR**	**RC**	**FPR**	**FNR**	**F1**	**MCC**
LDAP	E1	1040940	99	3971	3565	99.65	99.99	99.66	2.43	0.34	99.82	0.000000126
	E2	1037230	252	3818	7275	99.28	99.98	99.30	6.19	0.70	99.64	0.000000082
	E3	1032778	185	3885	11727	98.86	99.98	98.88	4.55	1.12	99.43	0.000000059
	E4	1032496	210	3860	12009	98.83	99.98	98.85	5.16	1.15	99.41	0.000000058
	E5	1032980	197	3873	11525	98.88	99.98	98.90	4.84	1.10	99.44	0.000000060
NETBIOS	E1	1045513	5	79	2978	99.72	100.00	99.72	5.95	0.28	99.86	0.000000300
	E2	1042966	10	74	5525	99.47	100.00	99.47	11.90	0.53	99.74	0.000000153
	E3	1040502	12	72	7989	99.24	100.00	99.24	14.29	0.76	99.62	0.000000103
	E4	1041602	6	78	6889	99.34	100.00	99.34	7.14	0.66	99.67	0.000000130
	E5	1043508	11	73	4983	99.52	100.00	99.52	13.10	0.48	99.76	0.000000167
PORTSMAP	E1	185435	75	4659	1525	99.17	99.96	99.18	1.58	0.82	99.57	0.000000366
	E2	184957	125	4609	2003	98.89	99.93	98.93	2.64	1.07	99.43	0.000000338
	E3	184971	119	4615	1989	98.90	99.94	98.94	2.51	1.06	99.43	0.000000339
	E4	183535	151	4583	3425	98.13	99.92	98.17	3.19	1.83	99.04	0.000000277
	E5	182955	149	4585	4005	97.83	99.92	97.86	3.15	2.14	98.88	0.000000258
SYN	E1	1012962	325	30336	4953	99.50	99.97	99.51	1.06	0.49	99.74	0.000000028
	E2	1010620	589	30072	7295	99.25	99.94	99.28	1.92	0.72	99.61	0.000000026
	E3	1010926	788	29873	6989	99.26	99.92	99.31	2.57	0.69	99.62	0.000000026
	E4	1009262	989	29672	8653	99.08	99.90	99.15	3.23	0.85	99.52	0.000000025
	E5	1011202	608	30053	6713	99.30	99.94	99.34	1.98	0.66	99.64	0.000000026
UDP	E1	1019683	89	1924	2487	99.75	99.99	99.76	4.42	0.24	99.87	0.000000214
	E2	1017459	110	1903	4711	99.53	99.99	99.54	5.46	0.46	99.76	0.000000141
	E3	1017026	84	1929	5144	99.49	99.99	99.50	4.17	0.50	99.74	0.000000134
	E4	1014879	135	1878	7291	99.27	99.99	99.29	6.71	0.71	99.64	0.000000100
	E5	1016294	95	1918	5876	99.42	99.99	99.43	4.72	0.57	99.71	0.000000121
MSSQL	E1	241403	175	3804	1989	99.13	99.93	99.18	4.40	0.82	99.55	0.000000333
	E2	240495	243	3736	2897	98.73	99.90	98.81	6.11	1.19	99.35	0.000000285
	E3	239103	289	3690	4289	98.15	99.88	98.24	7.26	1.76	99.05	0.000000233
	E4	239505	323	3656	3887	98.30	99.87	98.40	8.12	1.60	99.13	0.000000244
	E5	241493	189	3790	1899	99.16	99.92	99.22	4.75	0.78	99.57	0.000000338
Key	E1-Proposed Model (Ensemble Bagging ML Classifier)											
	E2-Ensemble Random Forest ML Classifier											
	E3- Ensemble AdaBoosting ML Classifier											
	E4- Ensemble Gradient Boosting Tree ML Classifier											
	E5- Ensemble Voting ML Classifier											

Overall, the results of the proposed ML model using the ensemble bagging technique were better than other ensemble ML techniques. For example, in Figure 3, the classification accuracy of the proposed ML model was over 99% across all the six datasets used in our experiment. The highest classification accuracy obtained in our experiment was 99.75% in the UDP dataset using the ensemble bagging technique with decision tree ML algorithm as the base estimator.

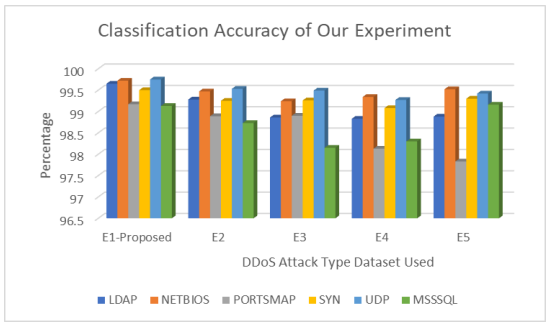

Figure 3. Classification accuracy results

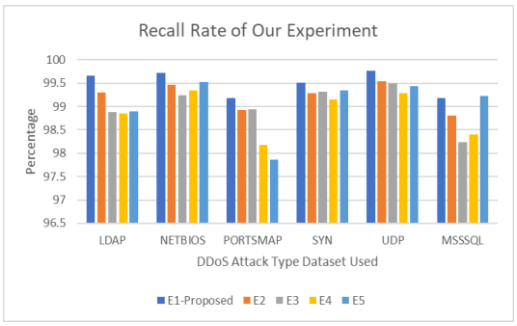

Figure 5. Recall rate results

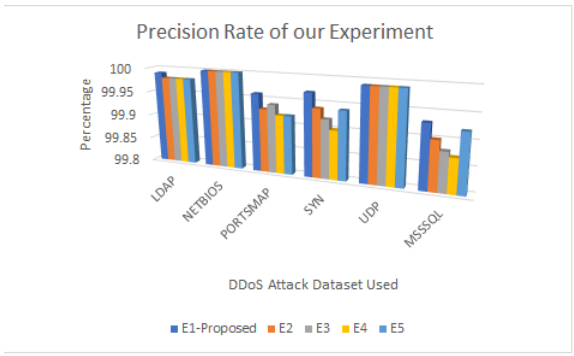

Figure 4. Precision rate results

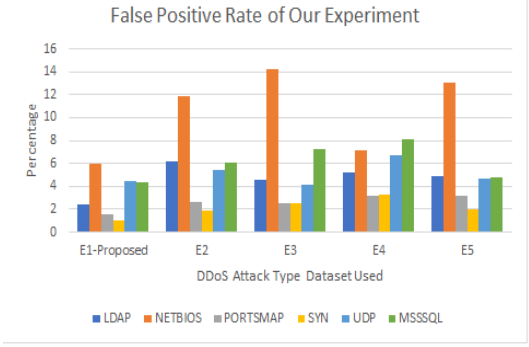

Figure 6. False positive rate results

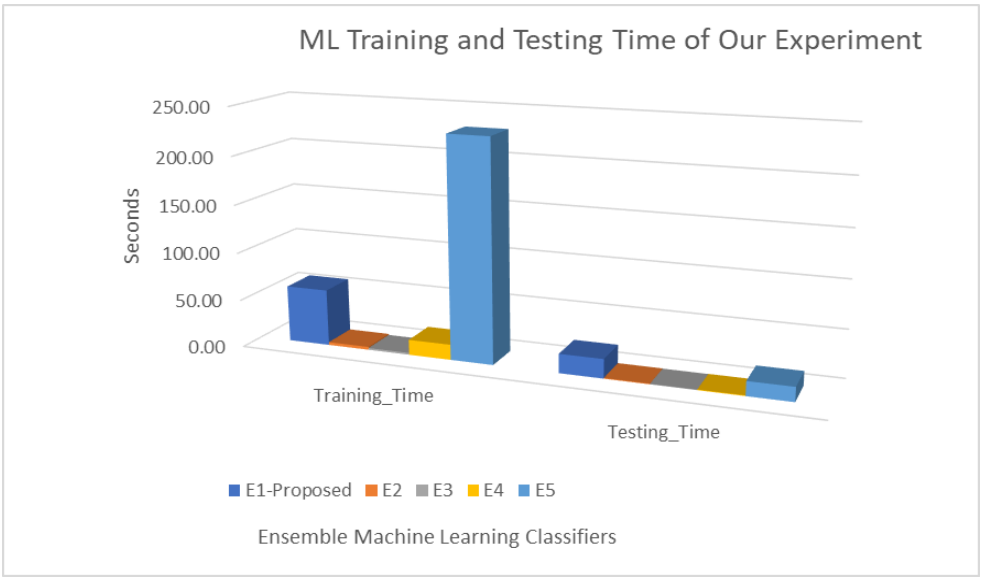

Figure 7. ML training and testing time results

In Figure 4, the precision rate was very high, for example in the NETBIOS dataset sample used in our experiment all the ensemble ML technique recorded 100% precision rate while in LDAP dataset, all ML techniques recorded over 99.98% precision rate. These results are attributed to the smaller number of benign samples and high number of DDoS attack types in the testing dataset used in the experiment., Similarly all ensemble ML techniques used in our experiment achieves over 99% precision rate. On the other hand, the recall rate was quite close to the precision rate as shown in Figure 5, as some of the ensemble ML techniques recorded 98%. However, the proposed ensemble ML model for the detection of DDoS attacks presented in this study recorded over 99% precision rate in all the six-dataset used in the experiment.

The false-positive rate for the proposed ML model from the experiment conducted was below 5% in all datasets as shown in Figure 6. Similarly, we obtained an overall of less than 3% false negative rates in all the dataset used in our experiment. The value of our F1-score was relatively good over 98% in all the dataset used in our experiment. The value of the MCC shows the reliability of the model as shown in Table 2. This helps to improve the performance of the detection of attacks in a real time IoT computing environment. The best FPR was recorded in SYN dataset with 1.06% and the worst FPR was 4.42% in the UDP dataset using the proposed ML model. However, overall, we recorded the worst FPR of 14.29% in the NETBIOS dataset using the AdaBoosting techniques in the experiment conducted in this study.

In addition, the evaluation of the proposed model for each dataset shows that the training time and testing time were outperformed by the RF classifier, AdaBoosting, and gradient boosting classifiers as shown in Figure 7, However, the proposed ensemble ML model using the bagging technique outperforms the ensemble voting classifier in both the training and testing time recorded during the experiment conducted in this study.

Finally, the results obtained in each dataset shown a similar trend with each evaluation metrics used in the experiment as shown in figure 3, 4,5 and 6. It was observed that the classification accuracy, precision rate and recall rate were over 97% in all datasets using the various ensemble ML techniques.

5.2. Results Comparison with Related Works

The results obtained in our experiment was also compares with related work in extant literature as shown in Table 3. The results shows that our proposed model compete favourably with similar research works.

Table 3. Comparison with Related Works

Source	**CA**	**PR**	**RC**	**FPR**	**FNR**	**F1**	**MCC**
Smys et al (2020)	98.60%	100%	100%				
Singh et al(2021)	99.92%	97.79%	93.07%				
Elsayed et al(2020)	99%	99%	99%				
Wei et al (2021)	98.34%	97.91%	98.48%				
Kousar et al (2021)	99.99%	99.40%	99.20%				
Malliga et al (2022)	99.40%						
Vuong et al(2021)		99.50%	99.10%				
Rajagopal et al(2021)	97%	99%	96%				
Gohil and Kumar(2020)	96.25%	96%	96%				
Shieh et al (2021)	98.18%	97.93%	99.84%				
Our Proposed Model	**99.75%**	**99.99%**	**99.76%**	**4.42%**	**0.24%**	**99.87%**	**0.000000214**

6. CONCLUSION AND FUTURE WORKS

This paper proposed a DDoS detection and prevention model to tackle malicious network traffic in IoT computing environment. The proposed model uses the ensemble Bagging ML techniques for the training of its detection engine. We also carried out an ML experiment using five ML ensemble classifiers with the most recent DDoS attack (CICDoS2019) dataset. The result of our experiment shows that the ensemble ML bagging classifier using a decision tree ML algorithm as the base estimator outperforms other ensemble ML classifiers as we recorded an overall classification accuracy of 99.75% and 1.06% to 5.95% false-positive rate. The results obtain in our study follows similar trends with related study and also outperforms some of the proposed work in extant literature. In the future, we intend to apply an efficient feature selection technique to select relevant features to improve the performance of our detection model and implements the models in a real-time IoT environment.

REFERENCES

[1] Spathoulas, G., Giachoudis, N., Damiris, G. P., & Theodoridis, G. (2019). “Collaborative blockchain-based detection of distributed denial of service attacks based on internet of things botnets”, Future Internet, 11(11), 226.

[2] Adat, V., & Gupta, B. B. (2017). “A DDoS attack mitigation framework for internet of things”, In 2017 international conference on communication and signal processing (ICCSP) (pp. 2036-2041). IEEE.

[3] Radhika, R., & Kulothungan, K. (2019). “Mitigation of Distributed Denial of Service Attacks on the Internet of Things”, Appl. Math, 13(5), 831-837.

[4] Manavi, M. T. (2018). “Defense mechanisms against distributed denial of service attacks: A survey”, Computers & Electrical Engineering, 72, 26-38.

[5] Sicari, S., Rizzardi, A., Miorandi, D., & Coen-Porisini, A. (2018). REATO: REActing TO Denial of Service attacks in the Internet of Things. Computer Networks, 137, 37-48.

[6] Bertino, E., & Islam, N. (2017). “Botnets and internet of things security”, Computer, 50(2), 76-79.

[7] Smys, S., Basar, A., & Wang, H. (2020). "Hybrid intrusion detection system for internet of things (IoT)", Journal of ISMAC, 2(04), 190-199.

[8] Singh-Samom, P., & Taggu, A. (2021). "Distributed denial of service (DDoS) attacks detection: A machine learning approach", In Applied Soft Computing and Communication Networks (pp. 75-87). Springer, Singapore.

[9] Elsayed, M. S., Le-Khac, N. A., Dev, S., & Jurcut, A. D. (2020). "Ddosnet: A deep-learning model for detecting network attacks", In 2020 IEEE 21st International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM) (pp. 391-396). IEEE.

[10] Wei, Y., Jang-Jaccard, J., Sabrina, F., Singh, A., Xu, W., & Camtepe, S. (2021). "Ae-mlp: A hybrid deep learning approach for ddos detection and classification", IEEE Access, 9, 146810-146821.

[11] Kousar, H., Mulla, M. M., Shettar, P., & Narayan, D. G. (2021). "Detection of DDoS Attacks in Software Defined Network using Decision Tree", In 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT) (pp. 783-788). IEEE.

[12] Malliga, S., Kogilavani, S. V., & Sowmya, R. (2022). "Deep discover: Deep learning models for detecting distributed denial of service (DDOS) attacks", In AIP Conference Proceedings (Vol. 2393, No. 1, p. 020191). AIP Publishing LLC.

[13] Vuong, T. H., Thi, C. V. N., & Ha, Q. T. (2021). "N-tier machine learning-based architecture for DDoS attack detection", In Asian Conference on Intelligent Information and Database Systems (pp. 375-385). Springer, Cham.

[14] Rajagopal, S., Kundapur, P. P., & Hareesha, K. S. (2021). "Towards effective network intrusion detection: from concept to creation on Azure cloud", IEEE Access, 9, 19723-19742.

[15] Gohil, M., & Kumar, S. (2020). "Evaluation of classification algorithms for distributed denial of service attack detection", In 2020 IEEE Third International Conference on Artificial Intelligence and Knowledge Engineering (AIKE) (pp. 138-141). IEEE.

[16] Shieh, C. S., Lin, W. W., Nguyen, T. T., Chen, C. H., Horng, M. F., & Miu, D. (2021). "Detection of unknown ddos attacks with deep learning and gaussian mixture model", Applied Sciences, 11(11), 5213.

[17] Sharafaldin, I., Lashkari, A. H., Hakak, S., & Ghorbani, A. A. (2019). "Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy", In 2019 International Carnahan Conference on Security Technology (ICCST) (pp. 1-8). IEEE.