UTILIZING ARTIFICIAL INTELLIGENCE FOR PROACTIVE ANALYSIS IN DDOS ATTACKS USING MACHINE LEARNING

BY

OYEBOLA BLESSING OYELOWO

ABSTRACT

Artificial intelligence is increasingly becoming part of our lifestyle. Businesses and critical infrastructures are blending towards this evolution, the rise in this sophistication is emerging more sophisticated attacks such as DDoS attacks which is posing a huge threat to the confidentiality, integrity and availability of the AI system, leading to financial loss and loss of privacy. This abstract discusses the proposal on how to proactively detect DDoS attacks in real-time using the CICDDOS2019 dataset for training and testing datasets using the hybrid machine learning, SVM and DBSCAN techniques to develop the DDoS attack detection algorithm.

TABLE OF CONTENT

ABSTRACT

TABLE OF CONTENTS

LISTS OF TABLES AND FIGURES

INTRODUCTION
- **Overview**
- **Problem Statement**
- **Research Questions**
- **Research Objectives**
- **Scope**
- **Ethical Issues**

LITERATURE REVIEW1
- **Review of Different Methods2**
- **Review of Used Datasets3**
- **Types of DDoS Attacks4**

RESEARCH METHODOLOGY1
- **Potential Dataset to Use2**
- **Proposed Methods3**
- **Proposed Algorithm4**

Procedure5

CONCLUSION1

REFERENCES1

APPENDIX1

LISTS OF TABLES AND FIGURES

Lists of Tables:

Table 1: Literature review of related works

Table 2: Types of DDoS Attacks

Lists of Figures:

Figure 1: SVM Algorithm

Figure 2: DBSCAN Algorithm

Figure 3: Proposed Algorithm

Figure 4:

1.0INTRODUCTION

1.1 Overview

Distributed Denial of Service (DDoS) attack is one of the most widespread and dangerous attacks in recent times and its amount increases annually (Emina et al., 2019). Its aims to disrupt or weaken the service of the victim, it can occur either in high-rate traffic, characterized by flooding the targeted system with a massive amount of traffic for a while, sometimes minutes or a few hours, causing resource exhaustion, or Low rate traffic, construed by slowly and subtlety sending malicious packet or traffic to the targeted system over an extended period, he malicious traffic usually contains minimal data and its goal is to overwhelm the targeted service, thereby achieving persistence gradually (Nooribakhsh et al., 2020).

According to Kaspersky Labs, the longest attack seen in recent years was witnessed in the final quarter of 2018 and lasted for almost 14 days and 329 hours, (Furfaro et al., 2020), In the last decade, there has been an emergence of more DDoS attacks due to the changes in the scalability of networks and topology (Wang et al., 2015; Website_Prolexic, 2017), NETSCOUT Arbor's Worldwide Infrastructure Security Report 2018 recounted that 91% of the enterprises suffered from overloading of Internet bandwidth due to 1.7 Tbps Distributed DoS (DDoS) attacks recorded in 2018 (Tsochev et al., 2020). Based on their performance and behaviours, DDOS attacks are divided into three forms of attacks; **Bandwidth-Based DDoS Attacks**; which flood the target network with a massive amount of data traffic, thereby overwhelming the capacity and causing congestion and disruption, UDP Flood, DNS Amplification and ICMP flood attack, **Protocol-based DDoS Attacks**; target vulnerabilities in the network protocols and overwhelms it, examples include SYN flood attacks and Ping of Death and **Application-based DDoS attacks;** which targets the application layer to

overwhelm the resources such as the memory and cu, it includes the HTTP flood and Zero Day attacks

With the evolution in digital interconnectivity and complexities of DDoS attack techniques and motivation, traditional reactive approaches to detecting and preventing DDOS attacks have become insufficient and tools like Firewalls and antivirus cannot handle the complexities, therefore there is a need for a proactive approach to utilizing artificial intelligence to detect and prevent DDoS threats before they are impactful. Within recent years, artificial intelligence has become a compelling force in combating complex attacks such as DDoS attacks, offering capabilities in detecting anomalies in DDoS attacks, analysis and response (Khalaf et al., 2019). AI-based approaches offers a powerful tool for proactively and dynamically detecting anomalies in DDoS attacks. AI systems, are characterized by their capabilities to leverage big data to analyze network traffics in real time (Garcia et al., 2021).

One major advantage of AI is their ability to adapt and learn from evolving threats, autonomously detecting anomalies and bad traffic from normal behavior, typical Artificial Intelligence techniques include Machine Learning and Deep learning (Zhang et al., 2017). By employing the power of AI, it aims to revolutionize the paradigm from reactive approach to proactive approach to detecting and preventing DDoS attacks, thereby enhancing their ability to mitigate DDoS threats.

1.2 Problem Statement (100)

DDoS attacks have been shown to be one of the attacks that pose a significant threat in recent times and increases yearly, causing significant financial losses and reputational damage to businesses. Traditional detection techniques are ineffective to combat against the sophistication and dynamics of the evolving DDoS attacks techniques. Therefore, accurately attributing DDoS attacks to their adversaries remains a challenge, hence, hindering effective response strategies.

1.3 Research Questions (100)

1. What are the current challenges associated with proactively identify pattern that demonstrate DDoS attacks?
2. How can AI techniques be effectively utilized to analyze and identify patterns that indicate DDoS attacks?
3. What type of AI techniques can be employed to accurately detect and DDoS attacks in real-time?
4. How to train the AI algorithm to facilitate the implementation of real-time adaptive response to combat the impact of DDoS attacks?

1.4 Research Objectives

1. To investigate current methodologies and technologies used in cyber threat intelligence for DDOS attack detection and attribution.

2. To review the traits that can exhibit a high rate of unusual traffic or specific DDOS attacks.
3. To identify challenges and gaps in proactive analysis techniques for DDOS attacks.
4. To develop innovative approaches to leverage cyber threat intelligence for early detection and proactive mitigation of DDOS attacks.

1.5 Scope

This paper primarily focuses on employing AI techniques, specifically, ML techniques to analyse DDoS attacks proactively. It encompasses the development of an algorithm that is trained to detect at both network and application levels of DDoS attacks.

1.6 Ethical Issues

I, Oyebola Oyelowo, hereby declare the proposal is my original work and all the materials belonging to others have been cited correctly.

2.0 LITERATURE REVIEW

2.1 Review of Different Methods

Researchers have proposed different methods for proactive and reactive protection mechanisms for detecting and preventing DDoS attacks. Thapngam et al (2014) proposed a behaviour-based Support Vector Machine (SVM) for detecting DDoS attacks using Pearson's correlation coefficient employed to extract repeatable features from packet arrivals in DDoS traffic but not in flash crowd traffic. Alternatively, Banitalebi et al. (2021) proposed combined entropy-based methods and classification algorithms for detecting high and low volumes of DDoS attacks. The experimental algorithm had high accuracy in detection and a low alarm rate except for naïve Bayes. Thapngam's approach used both generated and real datasets to test and optimize the detection system heuristic approach is not effective for sophisticated DDoS attacks and the approach lacks responsiveness to emerging threats, while Banitalebi relies on outdated and invalid datasets for training and a change in the network topology can effectiveness the accuracy of the detection system.

Tuan et al, 2019 conducted comprehensive research on botnet and DDoS attacks and supervised and unsupervised machine learning techniques used to detect DDoS attacks including Support Vector Machine (SVM), Artificial Neural Networks (ANN), Naïve Bayes (NB), Decision Tree (DT) and Unsupervised Learning (USML) (K-means, X-means, etc.). in contrast, Sahay et al. (2017) proposed ArOMA, a framework for detecting anomalies, and network monitoring using SDN to mitigate DDoS attacks. The limitations of the framework Luong et al. (2020) accounted for the high rate and severity

of DDoS attacks in Software Defined Networks (SDN) systems specifically, Dithub and Dyn attacks and the limitations of statistical-based methods in detecting abnormal traffic, In addition to that, the paper developed IDS-DDoS software that collects packets to generate and classify flows and notifies the SDN controller of malicious flows. The limitation of this algorithm is the limited data in training the algorithm.

Wanga et al., (2020) addressed DDoS attacks with their severity on network security, specifically Software Defined Network and developed the Hidden Markov Model (HMM), Autoencoder, ResNet and Gated Recurrent Unit which employed the Deep Belief Network (DBN) to detect DDoS attacks. In contrast, Khedr et al. (2023) proposed a multilayer DDoS attack detection and mitigation approach for Software Defined Network (SDN) based Internet of Things (IoTs) that used different algorithms including Support Vector Machine (SVM), Gaussian Naive Bayes (GNB), k-nearest Neighbor (KNN), Binomial Logistic Regression (BLR), Decision Tree (DT), and Random Forest (RF) to train the data, demonstrating high accuracy, precision and average detection time, however, its limitations are centered towards real-time IoT traffic analysis, comprehensive attack type detection, dependency on specific ML models and evaluation with Real-World IoT deployments. Wanga's model is a powerful approach to detecting DDoS attacks with the ability to perform classification tasks, however, the traditional DDoS detection techniques used are limited to adapt to changing attack patterns and are less effective for sophisticated attacks and the model might struggle with complex network environments. Khedr also used an innovative approach to developing the framework in that it used feature engineering and small window size for feature extraction, however, the limitation of relying on specific computed featured sets can limit its adaptability and evolving DDoS attacks.

Najar et al. (2022) employed different machine learning techniques including Random Forest (RF) and Multi-Layer Perceptron (MLP), Support Vector Machine (SVM) and K-Nearest Neighbour (KNN). Aktar et al. (2023) primarily focus on the deep learning approach to detecting DDoS attacks using contractive autoencoder achieving accuracy ranging between 93.41% and 97.58% and limitations in Hyperparameter tuning and data imbalance and its ability to detect zero-day attacks are not included in the trained data. Sadhwani et al. (2023) proposed a Lightweight Intrusion Detection System using machine learning and deep learning including logistic regression, naive Bayes, K-Nearest Neighbour, decision trees, random forest, support vector machine long short-term memory (LSTM), and convolutional neural network (CNN). Benmohamed et al. (2024) introduced an Encoder-Stacked Dee Neural network (ES-DNN) which leverages multi-layer perceptrons (MLP) for accurate DDoS attack detection, achieving an accuracy of 99.94%.

No	Authors	Point of Literature (summary)	Output
1	Tuan et al., 2019	The lit. gives comprehensive research on botnet and DDoS attacks and machine learning	Confusion matrix which is used for comparing the performance of the algorithms

		detection techniques including both Supervised and Unsupervised learning	
2	Thapngam et al., 2014	The lit proposes a behavior-based method for detecting the DDoS attacks based on Pearson's correlation	
3	Luong et al., 2020	This paper presented the high rate of DDoS attacks in Software Defined Networks (SDN) systems emphasising Github and Dyn attacks,	accuracy, precision recall and F1-Score to detect and mitigate DDoS attacks. Decision tree model has the best performance among machine learning
4	Najar et al. 2022	The document focuses on the detection of DDoS attacks using Random Forest (RF) and Multi-Layer Perceptron (MLP)	The RF model achieved an accuracy of 99.13% on both train and validation data and 97% on full test data, while the MLP model showed an accuracy of 97.96% on train data, 98.53% on validation data, and 74% on the full test dataset.
5	Wanga et al., 2020	The lit addressed DDoS attacks and their adverse effect on network security, especially Software Defined Networks (SDN),	Accuracy, Precision, Recall, and F1-Score with an accuracy rate of 96.28% and minimizing construction error
6	Khedr et al., 2023	The paper proposed a framework called FMDADM, a multilayer DDoS attack	high accuracy, precision, recall, and F1-score in detecting various attack scenarios achieving high accuracy, precision, recall, and average detection time.
7		The paper discussed the method of detecting DD0S attacks in cloud computing,	high accuracy and reduced misclassification errors compared to existing methods.
8	**Bawany et al., 2017**	Proposed a framework called ProDefense for uses SDN to detect and mitigate DDoS attacks against DDoS attacks in smart cities	
9	Aktar et al., 2023	The lit. primarily focuses on the deep learning model approach to detecting anomalies in network traffic, especially in DDoS attacks using contractive auto encoder. It elaborates on the limitations of traditional machine learning techniques for detecting unknown attacks and the potential of deep learning models for complex, large-scale network environments.	accuracy ranging between 93.41% and 97.58% on the CIC-DDoS2019 dataset, 96.08% and 92.45% on NSL-KDD, and CIC-IDS2017 datasets, respectively
10	Sadhwani et al., 2023	The paper emphasises the vulnerability in lightweight IoT networks to DDoS attacks and the need for effective detection models to safeguard critical infrastructure., it further proposes	Using accuracy, recall, precision, F1 score, training time, prediction time, and total time as metrics, achieving 100% accuracy

		a light weight intrusion detection system that integrates machine learning	
11	Benmohamed et al., 2024	With the insurgence in cyber attacks including DDoS attacks on network infrastructure, the paper introduces Encoder-Stacked Deep Neural Networks for DDoS Attack Detection which leverages stacked/bagged multilayer perceptrons (MLP) for precise DDoS attack detection	Accuracy rate of 99.94% for the CICDDS2017 dataset and 98.86% for the CICDDoS2019 dataset.

Table 1: Literature Review of Related Works

2.2 Review of Used Datasets

Some of the datasets used in the reviews used in the implementation of the models and algorithms according to the trained and tested dataset are mainly NSL-KDD and CIC-IDS

UNBS-NB 15 dataset is a comprehensive dataset created by researchers in the University of New South Wales (UNSW) in Canberra, Australia for the design of Network Intrusion Detection System.. the dataset captures key features including source files, UNBS-NB 15

DARPA: Defense Advanced Research Project Agency Datasets consist of network traffic data gathered from various military network environments used for evaluating intrusion detection systems. It is developed by MIT Lincoln Laboratory, consisting of 31 features and includes attack types such as Dos, R2L,U2R and Probe.

NSL-KDD: NSL-KDD is a refined version of the KDDCup99 dataset used for intrusion detection systems (IDS), it contains essential methods to develop intrusion detection systems (Meena et al., 2017). The NSL-KDD dataset includes various classes of attacks used for analysis such as Denial of Service (DoS), unauthorized access from a remote machine (R2L), unauthorized access to Local super user (U2R) and Probing. It is a detailed resource for network security, identifying TCP's vulnerability to intrusion refines the IDS models to bolster network protection (Dhanabal et al., 2015).

CSE-CIC-IDS: the dataset is developed by the Canadian Institute for Intrusion Detection System for evaluating Intrusion Detection Systems; the dataset provides a range of network traffic features and evolving attack types based on recent scenarios such as denial-of-service attacks, botnets and SQL Injection, XSS, brute force and web attacks (Thakkar et al. 2020). Furthermore, the dataset identifies the protocols that are vulnerable to intruders utilizing network packets from different sources including payload data and packer header, hosts, firewalls and destinations therefore contributing to the detection of network-based intrusions. Their limitations include imbalanced class distribution, missing and redundant data and the complexities of managing large datasets which could be mitigated and improved through preprocessing techniques, feature engineering and data sampling (Gopalan et al., 2021).

2.3 Types of DDoS attacks

There are several types of attacks and their characteristics based on the three attack categories which are found in the table

Categories of DDoS Attacks	Types of DDoS Attacks	Characteristics of Types
Bandwidth-Based DDoS Attacks	UDP Flood Attacks	UDP flood exploits the connectionless nature of the UDP protocol and floods the victim's computer system or networks with a large number of UDP packets (Halladay et al., 2022)
	DNS Amplification	DNS amplification exploits the DNS to flood thevictim server with an amplified amount of traffic, it targets oen DNS Resolvers, bots query the DNS resolvers with arguments like “ANY”, and the DNS resolver then responds to the spoofed IP address with a large data response (Halladay et al., 2022)
	ICMP Flood Attacks	ICMP flood overwhelms the bandwidth by sending large amounts of ICM echo requests to the victim's server (Halladay et al., 2022)
Protocol-Based DDoS Attacks	TCP SYN Flood Attacks	TCP SYN floods exploit the TCP protocol and send repeated SYN packets to random ports on the victim's system or network using fake IP addresses, thereby disrupting the TCP three-way handshake (Douligeris and Mitrokotsa, 2004).
	Ping to Death Attacks	Ping of Death involves sending a large volume of ICMP packets to the victim's computer, thereby causing the buffer overflow and, thence, system crash (Halladay et al., 2022)
	Smurf Attacks	Smurf attacks are flooding the victim's system with ICMP packets using a spoofed source IP Address to broadcast IP addresses, thereby overwhelming the system traffic (Douligeris and Mitrokotsa, 2004).
Application-Based DDoS Attacks	Zero-Day Attacks	Zero-day flood exploits vulnerabilities in software or systems that are not known to the public, it is also used to launch coordinated attacks and take advantage of the lack of available patches (Halladay et al., 2022).
	HTTP Flood Attacks	HTTP flood targets web servers, overwhelming them with large amounts of HTTP GET or POST requests thereby, causing web disruption (Halladay et al., 2022)

Table 1: Types of DDoS Attacks, (Oyebola Oyelowo)

Table 1 outlines the different types of DDoS attacks based on their categories, the key finding in their characteristics is their ability to overwhelm the target system or network by sending large volumes of traffic and data, the use of spoofed IP addresses and botnet control. These characteristics are crucial to the detection and training of the algorithm to enhance the overall cybersecurity posture.

3.0 RESEARCH METHODOLOGY

3.1 Potential Datasets to Use

In this paper, the CICDDoS2019 dataset is used for the implementation of the algorithm. it is created by the Canadian Institute for Cybersecurity (CIC) wit the University of Brunswick it contains both benign and updated common DDoS attacks which could pass for the real-world data (PCAPs) and labels flows based on timestamps, attack, protocols, source and destination ports and source and destination IP addresses in a CSV file, it also contains 86 traffic features that consists of modern types of DDoS attacks including UDP, NTP, DNS,

3.2 Proposed Method

The evolution and sophistication of DDoS attacks, the proposed method is a hybrid method that combines support vector Machine (SVM) with Density-Based Spatial Clustering of Application with Noise (DBSCAN), the combination of SVM and DBSCAN is a robust approach to creating the model which leverages on the strength of the supervised and unsupervised learning to proactively and effectively detect and mitigate DDoS attack in real-time.

3.2.1 Support Vector Machine (SVM): it is a supervised learning algorithm that is commonly used for classification and regression, it performs effectively in binary classification tasks, using both linear and non-linear approaches to classify data (Dwivedi et al., 2020). It maximizes the margin between the margin and the closet data points called support vectors by finding the optimal hyperplane that best separates the data into different classes**.** SVM performs well when handling both linear and non-linear classification problems, it has different kernel functions such as Linear, radial basis, polynomial and sigmoid (Nguyen et al., 2024).

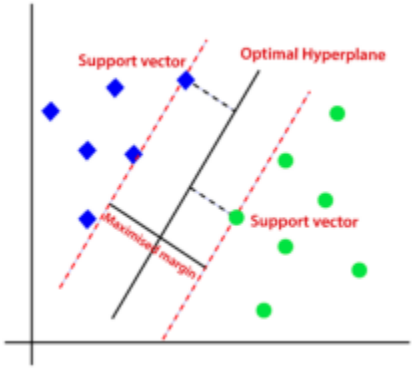

Figure 1: SVM Algorithm Dwivedi et al. (2020)

SVM has enough strength to choose the detection of DDoS attacks in real-time, it is highly effective in Higher Dimensional Spaces, is commonly found in network traffic and can handle large feature sets without significant performance degradation., furthermore,

SVMs have different kernel functions giving them room for flexibility to model both linear and non-linear relationships, thereby being able to handle complex patterns (Dwivedi et al., 2020). Lastly, it is excellent for avoiding overfitting and data imbalances by employing the right parameters which is important for maintaining generalizations in real-life scenarios (Chang and Yang, 2017).

3.2.2 Density-Based Spatial Clustering of Application with Noise (DBSCAN): the DBSCAN is an unsupervised learning algorithm that is used for identifying clusters of data points based on density, it identifies novel DDoS attack patterns that are not seen during the training.

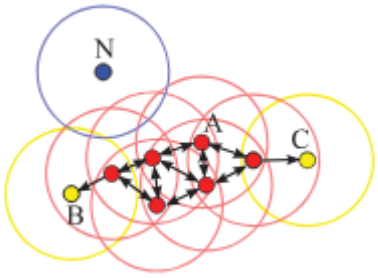

Figure 2: DBSCAN Algorithm (Schubert et al., 2017)

The arrows indicate the direct density reachability and points B and C represent the density connected since both B and C are density reachable from A and N is not density reachable and can be a noise point (Schubert et al., 2017).

DBSCAN thrives in detecting clusters of varying shapes to solve the problem of irregular and complex patterns and does not need to know the number of clusters specified beforehand, furthermore, it is capable of identifying noise and outliers in the data, useful in network traffic analysis where anomalous signals are sparse (Schubert et al., 2017). Lastly, DBSCAN clusters data points based on density which makes it effective for identifying high volumes of traffic which could indicate DDoS attacks.

Integrating both SVM and DBSCAN is a combined strength to produce a robust DDoS attack detection system that extracts features in real-time and continuously monitors network traffic.

3.3 Proposed Algorithm

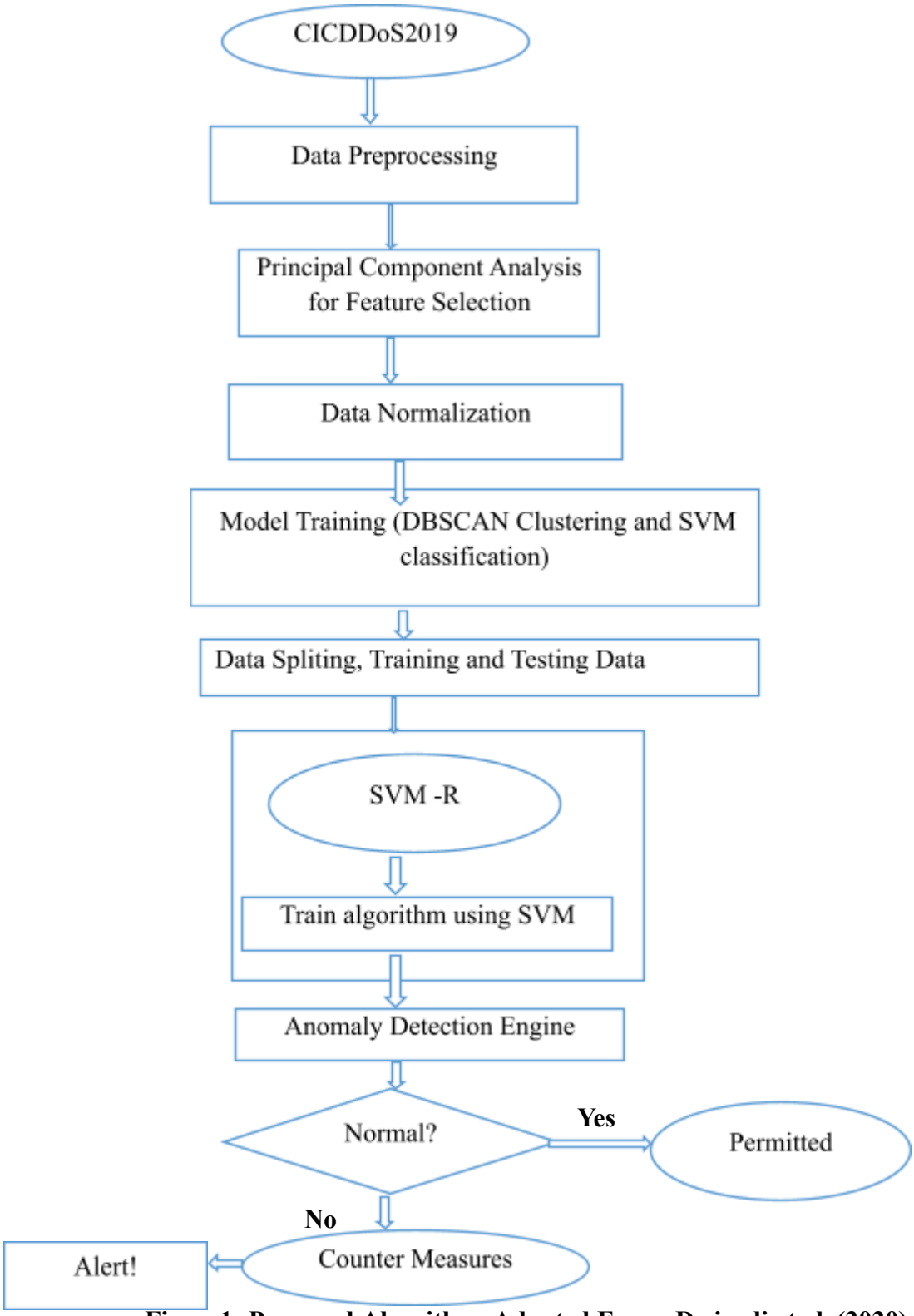

Figure1: Proposed Algorithm, Adapted From: Dwivedi et al. (2020)

The proposed diagram is a flowchart representation of how the algorithm will be implemented. It contains the step-by-step procedure of the development process, starting from the collected data, after the data has been trained to detect anomalies, it confirms that every data that comes in is normal, if not, it generates an alert to the system or the organization.

3.4 Procedure

The procedure to implement the DDoS detection algorithm for proactive analysis utilizing the SVM and DBSCAN is in five stages which are, data collection, data preprocessing, model training and model testing (Huang et al., 2018);

3.4.1 Data Collection

In this stage, the dataset is collected, the CICDDoS2019 dataset, which contains various types of network traffic data including normal attacks and the types of DDoS attacks, the dataset collected is in CSV format.

3.4.2 Data Preprocessing

After the dataset is collected, it will undergo pre-processing to minimize noise and overfittings and make it well-suitable for training the algorithms which primarily operate on numerical data. To enable compatibility in numerical features, **One hot encoding** is required where some features such as flowID, destination IP and source IP are transformed into numerical representations. One hot encoding converts the categorical features to binary columns which are either 0, indicating no corresponding category in the data or 1, which indicates a category data.

- **Data Cleaning:** this refers to the process of identifying and correcting errors or missing values in the dataset, it is done using **elimination;** for data samples with features values NAN**, 10E10;** for features valued INF **and 0;** for features with negative values
- **Feature Selection**: this involves identifying and removing highly correlated features and reducing feature space, Principal Component Analysis (PCA) will be used.
- **Data Normalization**: it involves normalizing the data by rescaling the values of the features to the range of (0, 1), using the min-max scaling method (Wang and Li, 2024);

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Where x_n, the value is after normalization, $x_{max,}$ is the maximum value feature and x is the minimum value of the current attributed feature (Wang and Li, 2024)

3.4.3 Model Training and Testing:

DBSCAN: After preprocessing it, the DBSCAN is clustered, the optimal value for the epsilon parameter is calculated, and the DBSCAN is applied to the dataset to detect

clusters of normal traffic and detect outliers and the clustered separates into normal and anomalous detection traffic patterns, it takes the raw data and outputs the cluster flow (Najafimehr et al., 2022).

SVM: After the raw data is clustered, there is a need for the classifier to label the clusters, the SVM performs the work of classification. Additionally, the proposed algorithm will be trained using the training data. The training stage will begin with an initialization by selecting the kernel and hyperparameter, in the proposed model, the RBF kernel function and the C hyperparameter will be selected, and then, the dataset is split into two, the training dataset, 80% and the testing dataset, 20%. Lastly, the SVM Model is trained, optimizing the hyperparameter through cross-validation (Najafimehr et al., 2022).

\Evaluation: The implementation will be performed using Python Programming Language using the Windows operating system. The 'scikit-learn' library is used for implementing the DBSCAN and SVM algorithms (Pedregosa et al., 2011).

3.4.4 Performance Metrics

The performance metrics of the proposed algorithm is evaluated the performance of the

will be evaluated on the SVM and DBSCAN using the test dataset and tested for some metrics;

- **Performance Metrics of SVM**

Precision: precision is the ratio of the corrected predicted positives observed to the total predicted positives (Najafimeh et al., 2022);

$$Precision = \frac{True\ Positives}{True\ positives+False\ Positives}$$

Accuracy: Accuracy is the ratio of the correctly predicted instances (true positives and true negatives) to the total instances (Najafimeh et al., 2022);

$$Accuracy = \frac{True\ Positives+True\ Negatives}{Total\ Instances}$$

*Where True Instances = True Positives+ True Negative + False Positives + False Negative (*Najafimeh et al., 2022*)*

F1-score: The F1 score is the weighted average of precision and recall, thereby providing the balance between the two (Najafimeh et al., 2022);

$$F1 - Score = 2\,x\,\frac{Precision\ x\ Recall}{Precision + Recall}$$

False Positive Rate (FPR): FPR indicates how many false positives are recognized by the algorithm (Najafimeh et al., 2022);

$$FPR = \frac{False\ Positives}{False\ Positives + True\ Negatives}$$

Recall: it indicates how many true positives are correctly recognized by the algorithm by comparing true positives with actual positives (Najafimeh et al., 2022);

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Then, the performance of the algorithm's effectiveness in detecting a DDoS attack will also be tested.

- **Performance Metrics of DBSCAN**

  The accuracy of the DBSCAN clustering needs to be tested using **Purity;** to determine the extent to which the resulted clusters' points belong to the same class, either normal or DDoS and the **homogeneity score**; shows how much these points belong to a single class (Najafimeh et al., 2022);

$$Homogeneity = 1 - \frac{H(C|K)}{H(C)}$$

Where H(C|K) is the conditional entropy of the classes given the cluster assignments (Najafimeh et al., 2022),

H(C) is the entropy of the classes.

$$purity = \frac{1}{N} \sum_{k} \max_{j} |\omega_k \cap c_j|$$

Figure 2: Purity folmular, (Najafimeh et al., 2022)

Where, N is the number of all points used in the clustering procedure, Ω = {w1, …wk}is the set of resulted clusters, and C = {c1, ..., cJ} is the set of the class labels (Najafimeh et al., 2022).

4.0 CONCLUSION

In conclusion, the primary objective of this proposal is to investigate current methods and techniques used in DDoS attack detetction. Through comprehensive research, the characteristics that are associated with DDoS attacks were identified including attack signatures, unusual traffic patterns and overwhelming resources. In addition, the gaps in proactive research of DDoS attack detection were also highlighted. This paper outlines a comprehensive approach to proactive innovative approaches to detecting DDoS attacks in artificial intelligence and develop an anomalous-based DDoS attack detection system using hybrid machine learning of SVM and DBSCAN. The paper aims to address these objectives and contribute to the advancement of cybersecurity research

5.0 REFERENCE

1. Aktar, S. and Nur, A.Y., 2023. Towards DDoS attack detection using deep learning approach. *Computers & Security*, *129*, p.103251.
2. Alduailij, M., Khan, Q.W., Tahir, M., Sardaraz, M., Alduailij, M. and Malik, F., 2022. Machine-learning-based DDoS attack detection using mutual information and random forest feature importance method. *Symmetry*, *14*(6), p.1095.
3. Banitalebi Dehkordi, A., Soltanaghaei, M. and Boroujeni, F.Z., 2021. The DDoS attacks detection through machine learning and statistical methods in SDN. *The Journal of Supercomputing*, *77*(3), pp.2383-2415.
4. Bawany, N.Z., Shamsi, J.A. and Salah, K., 2017. DDoS attack detection and mitigation using SDN: methods, practices, and solutions. *Arabian Journal for Science and Engineering*, *42*, pp.425-441.
5. Benmohamed, E., Thaljaoui, A., Elkhediri, S., Aladhadh, S. and Alohali, M., 2024. E-SDNN: encoder-stacked deep neural networks for DDOS attack detection. *Neural Computing and Applications*, pp.1-13.
6. Chang, Y., Li, W. and Yang, Z., 2017, July. Network intrusion detection based on random forest and support vector machine. In *2017 IEEE international conference on computational science and engineering (CSE) and IEEE international conference on embedded and ubiquitous computing (EUC)* (Vol. 1, pp. 635-638). IEEE.
7. Douligeris, C. and Mitrokotsa, A., 2004. DDoS attacks and defense mechanisms: classification and state-of-the-art. *Computer networks*, *44*(5), pp.643-666.
8. Dwivedi, S., Vardhan, M., Tripathi, S. and Shukla, A.K., 2020. Implementation of adaptive scheme in evolutionary technique for anomaly-based intrusion detection. *Evolutionary intelligence*, *13*(1), pp.103-117.
9. Furfaro, A., Pace, P. and Parise, A., 2020. Facing DDoS bandwidth flooding attacks. *Simulation Modelling Practice and Theory*, *98*, p.101984.
10. Garcia, N., Alcaniz, T., González-Vidal, A., Bernabe, J.B., Rivera, D. and Skarmeta, A., 2021. Distributed real-time SlowDoS attacks detection over encrypted traffic using Artificial Intelligence. *Journal of Network and Computer Applications*, *173*, p.102871.
11. Halladay, J., Cullen, D., Briner, N., Warren, J., Fye, K., Basnet, R., Bergen, J. and Doleck, T., 2022. Detection and characterization of DDoS attacks using time-based features. *IEEE Access*, *10*, pp.49794-49807.
12. Huang, J., Tang, Y. and Chen, S., 2018. Energy demand forecasting: combining cointegration analysis and artificial intelligence algorithm. *Mathematical Problems in Engineering*, *2018*, pp.1-13.
13. Khalaf, B.A., Mostafa, S.A., Mustapha, A., Mohammed, M.A. and Abduallah, W.M., 2019. Comprehensive review of artificial intelligence and statistical approaches in distributed denial of service attack and defense methods. *IEEE Access*, *7*, pp.51691-51713.

14. Khedr, W.I., Gouda, A.E. and Mohamed, E.R., 2023. FMDADM: A multi-layer DDoS attack detection and mitigation framework using machine learning for stateful SDN-based IoT networks. *IEEE Access*, *11*, pp.28934-28954.
15. Lawal, M.A., Shaikh, R.A. and Hassan, S.R., 2021. A DDoS attack mitigation framework for IoT networks using fog computing. *Procedia Computer Science*, *182*, pp.13-20
16. Luong, T.K., Tran, T.D. and Le, G.T., 2020, November. DDoS attack detection and defense in SDN based on machine learning. In *2020 7th NAFOSTED conference on information and computer science (NICS)* (pp. 31-35). IEEE.
17. Najafimehr, M., Zarifzadeh, S. and Mostafavi, S., 2022. A hybrid machine learning approach for detecting unprecedented DDoS attacks. *The Journal of Supercomputing*, *78*(6), pp.8106-8136.
18. Najar, A.A. and Manohar Naik, S., 2022. DDoS attack detection using MLP and Random Forest Algorithms. *International Journal of Information Technology*, *14*(5), pp.2317-2327.
19. Nguyen, T.L., Kao, H., Nguyen, T.T., Horng, M.F. and Shieh, C.S., 2024. Unknown DDoS Attack Detection with Fuzzy C-Means Clustering and Spatial Location Constraint Prototype Loss. *Computers, Materials & Continua*, *78*(2).
20. Nooribakhsh, M. and Mollamotalebi, M., 2020. A review on statistical approaches for anomaly detection in DDoS attacks. *Information Security Journal: A Global Perspective*, *29*(3), pp.118-133.
21. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, pp.2825-2830.
22. Sadhwani, S., Manibalan, B., Muthalagu, R. and Pawar, P., 2023. A Lightweight Model for DDoS Attack Detection Using Machine Learning Techniques. *Applied Sciences*, *13*(17), p.9937
23. Sahay, R., Blanc, G., Zhang, Z. and Debar, H., 2017. ArOMA: An SDN based autonomic DDoS mitigation framework. *computers & security*, *70*, pp.482-499.
24. Schubert, E., Sander, J., Ester, M., Kriegel, H.P. and Xu, X., 2017. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, *42*(3), pp.1-21.
25. Thapngam, T., Yu, S., Zhou, W. and Makki, S.K., 2014. Distributed Denial of Service (DDoS) detection by traffic pattern analysis. *Peer-to-peer networking and applications*, *7*, pp.346-358.
26. Tsochev, G., Trifonov, R., Nakov, O., Manolov, S. and Pavlova, G., 2020, October. Cyber security: Threats and challenges. In *2020 International Conference Automatics and Informatics (ICAI)* (pp. 1-6). IEEE.
27. Tuan, T. A. et al. (2019) 'Performance Evaluation of Botnet Ddos Attack Detection Using Machine Learning', Evolutionary Intelligence, 13(2), pp. 283–294. doi: 10.1007/s12065-019-00310-w.

28. Wanda, P. and Hiswati, M.E., 2024. Belief-DDoS: stepping up DDoS attack detection model using DBN algorithm. *International Journal of Information Technology*, *16*(1), pp.271-278.
29. Wang, H. and Li, Y., 2024. Overview of DDoS Attack Detection in Software-Defined Networks. *IEEE Access*.
30. Zhang, B., Zhang, T. and Yu, Z., 2017, December. DDoS detection and prevention based on artificial intelligence techniques. In *2017 3rd IEEE International Conference on Computer and Communications (ICCC)* (pp. 1276-1280). IEEE.

6.0 APPENDIX

No	Abbreviation	Meaning
	UDP	User Datagram Protocol
	DNS	Domain Name System
	ICMP	Internet Control Message Protocol
	HTTP	Hypertext Transfer Protocol
	CICDDoS2019	Cybersecurity Industry Consortium for Denial of Service Open Dataset 2019
	NAN	Not a Number
	10E10	10 raised to the power 10
	INF	Infinite
	0	None