International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | Apr 2024 www.irjet.net p-ISSN: 2395-0072

Cyber security threats prevention, detection and mitigation using machine learning techniques

Rishu Nitin Verma [1], Kirti Rajeev Tiwari [2], Mohd Idrisi Akram [3], Prof. Simran Patil[4]

[1,2,3]B.E. Student, Dept. of Information Technology, Theem College of Engineering, Maharashtra, India
[4]Professor, Dept. of Information Technology, Theem College of Engineering, Maharashtra, India

---

Abstract - *Communication is essential for individuals with disabilities to fully participate in society. However, those who are blind, deaf, and mute face significant challenges in traditional communication methods. The paper proposes a multimodal approach to facilitate communication for individuals with these multiple disabilities. The approach combines sign language, translator, and assistive technologies to create a comprehensive communication system. Through voice assistant, individuals who are blind can receive information through voice or touch, while translator devices provide real-time feedback. For sign language gestures. Additionally, assistive technologies such as text-to-speech and speech-to-text software enable individuals who are mute to communicate verbally, which can be translated into tactile or visual formats for those who are blind or deaf. By integrating these modalities, individuals with multiple disabilities can overcome communication barriers and engage more effectively with their environment and peers. This multimodal approach offers a promising avenue for enhancing communication. In today's advance world of science and technology, communication field has developed at such extent where we can connect to any part of the world within fraction of minutes and hour. We can send messages, make call or send documents, files to anyone, according to need. Communication has been important to express our thoughts, idea etc. but when it's come about blind, deaf and mute people's it become difficult for them and us to communicate with each other. So, here we have made such a software which shall help them to communicate with each other without depending upon any middle man. With the help of this we would be able to help mute, blind and deaf people to communicate with each other without depending upon each other.*

Key Words: *SDN; ML; SVM; KNN; DDOS; Email spam*.

1. INTRODUCTION

The modern era of cybersecurity is fraught with an expanding spectrum of threats, ranging from the disruptive force of Distributed Denial of Service (DDoS) attacks to the insidious infiltration of email spam campaigns. These threats not only jeopardize critical systems but also undermine the integrity of data and compromise user privacy. Consequently, there is an urgent need to develop resilient and adaptive cybersecurity measures to counter these evolving challenges. In the realm of cyber threat management, encompassing prevention, detection, and mitigation, through a synthesis of Machine Learning (ML) techniques and Software-Defined Networking (SDN) infrastructure. By melding the predictive prowess of ML algorithms with the agile control mechanisms provided by SDN, this approach aims to fortify cybersecurity defenses and mitigate the impact of malicious activities. Central to this investigation are DDoS attacks, notorious for their ability to incapacitate networks by flooding them with an overwhelming influx of traffic. Leveraging SDN capabilities, the study pioneers a methodology for the real-time identification and neutralization of DDoS attacks. ML algorithms including Random Forest, Support Vector Machine (SVM), Decision Tree, k-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression are harnessed to scrutinize network traffic patterns and discern aberrant behavior indicative of potential DDoS assaults. Moreover, the study confronts the persistent menace of email spam, a favored vector for phishing exploits and malware dissemination. Harnessing the adeptness of Naive Bayes, a venerable ML algorithm for text categorization, the research explores strategies to accurately discern and filter out spam emails.

Through the integration of ML algorithms within the SDN framework, this research strives to empower proactive threat mitigation, swift detection, and efficient response strategies against both DDoS attacks and email spam. Rigorous empirical evaluations will ascertain the effectiveness of this integrated ML-SDN solution, contributing to the advancement of cybersecurity practices and reinforcing resilience against the dynamic threat landscape.

2. LITERATURE SURVEY

Conducting a methodical literature review serves as a method to assess and interpret all existing research pertinent to a specific research query, subject, or phenomenon under examination. The study employed comprehensive scientific databases containing full-text papers, along with other relevant scholarly articles within the realm of social sciences. All academic papers and other relevant publications produced between 2009 and March 2020 were considered in the analysis.

2.1. Existing Papers

DDoS detection and mitigation using SDN has been lauded for its potential to offer instantaneous responses to cyber threats. Nonetheless, researchers have noted significant scalability challenges within SDN controllers, particularly when confronted with large-scale DDoS attacks. These scalability limitations have the potential to adversely affect response times and the overall efficacy of DDoS mitigation strategies. As organizations increasingly rely on SDN-based solutions for cyber defence, addressing these scalability issues becomes imperative to ensure the resilience and effectiveness of DDoS detection and mitigation mechanisms. Without robust scalability measures in place, SDN controllers may struggle to handle the sheer volume of traffic associated with sophisticated DDoS attacks, leaving networks vulnerable to prolonged disruptions and downtime. Therefore, future research endeavours should focus on enhancing the scalability of SDN controllers through innovative approaches and optimizations, thereby bolstering their ability to mitigate large-scale DDoS attacks in real-time.[1]

The utilization of Machine Learning (ML) for detecting Distributed Denial of Service (DDoS) attacks within the framework of Software-Defined Networking (SDN) has emerged as a promising avenue. This approach, particularly employing Random Forest and Support Vector Machine (SVM) algorithms, has shown considerable potential in identifying DDoS incidents effectively. However, despite the promising initial results, researchers have highlighted an ongoing challenge in aligning ML models with the constantly evolving tactics employed by cyber attackers. This underscores the importance of a continual process of model retraining to ensure the adaptability of detection mechanisms to emerging threats. Given the persistent evolution of cyber threats, the static nature of ML models poses a significant obstacle to maintaining robust detection capabilities. Consequently, there is an urgent need for innovative methodologies that facilitate dynamic model adaptation and retraining.[2]

email spam detection employing ML techniques, a prominent issue revolves around the presence of incorrect positives and negatives. This quandary arises when ML models mistakenly classify legitimate emails as spam or vice versa, potentially resulting in user inconvenience or vulnerabilities in security.[3]

3. SYSTEM ARCHITECTURE

In the pursuit of accomplishing the objective of recognizing and alleviating DDoS attacks in SDN networks, a model has been devised within the application tier. The framework of the executed methodology encompasses four principal components: the flow aggregation module, feature extension module, anomaly identification module, and anomaly mitigation module. The function of the flow aggregator is to methodically amass data pertaining to traffic flow from the flow tables of individual switches. These flow records are subsequently relayed to the feature extension module, which generates supplementary attributes for each entry. A detection of email spam by categorising it into spam and ham (hold and modify).

3.1. Design

DDoS attack detection system for SDN networks our system is composed of three primary components, outlined as follows: Packet feature extraction: Upon packet ingress into the system, we leverage the OpenFlow protocol to facilitate the extraction of packet attributes such as IP addresses, ports, etc. These attributes are subsequently organized into flow tables and flow entries. Following this, the collected information is transmitted to the detection module residing at the Controller after a predefined time interval. Detection mechanism: We employ machine learning algorithms to compare statistical data with the input dataset. Mitigation strategy: Following the comparison process, if any potentially harmful network data is identified, this module will enact mitigation measures to ensure system integrity during an attack scenario.

3.2. Requirement Analysis

The cyber security system requires the integration of ML techniques and SDN capabilities for effective threat prevention, detection, and mitigation. Functional requirements include utilizing SDN for DDoS detection and ML algorithms such as Random Forest, SVM, Decision Tree, kNN, Naive Bayes, and Logistic Regression for DDoS detection, along with Naive Bayes for email spam detection. Non-functional requirements encompass ensuring real-time monitoring and analysis, low latency, scalability to handle large volumes of traffic, security in communication and data protection, reliability in robustness and fault tolerance, adaptability to evolving threats, and interoperability with existing infrastructure and protocols. By addressing these requirements, the system can effectively safeguard against cyber threats.

The following are the different kinds of requirement for our project:

Table-1: Requirements of our system

Software Requirements	Hardware Requirements
Oracle virtualbox	Windows 10 or latest version
Python	8 GB RAM
RYU	Intel core processor i3
Mininet	100GB free Hard Disk
Hping3	1MBps internet connection

3.3. Proposed System

The proposed system integrates Software-Defined Networking (SDN) for real-time Distributed Denial of Service (DDoS) detection and mitigation, alongside a suite of Machine Learning (ML) algorithms such as Random Forest, Support Vector Machine (SVM), Decision Tree, k-Nearest Neighbors (kNN), Naive Bayes, and Logistic Regression. Leveraging the flexibility of SDN, the system actively monitors network traffic patterns and behaviours. ML algorithms scrutinize this data, identifying anomalous patterns indicative of potential DDoS attacks. By employing sophisticated techniques like Random Forest and SVM, the system enhances its capability to detect and respond to DDoS threats promptly and effectively.

In addition to DDoS detection, the system addresses the pervasive issue of email spam using Naive Bayes, a well-established ML algorithm for text classification tasks. Naive Bayes analyses email content and metadata, accurately discerning spam messages from legitimate ones. This proactive approach to email spam detection ensures that malicious emails are identified and filtered out before they reach users' inboxes, thereby reducing the risk of phishing attempts and malware dissemination.

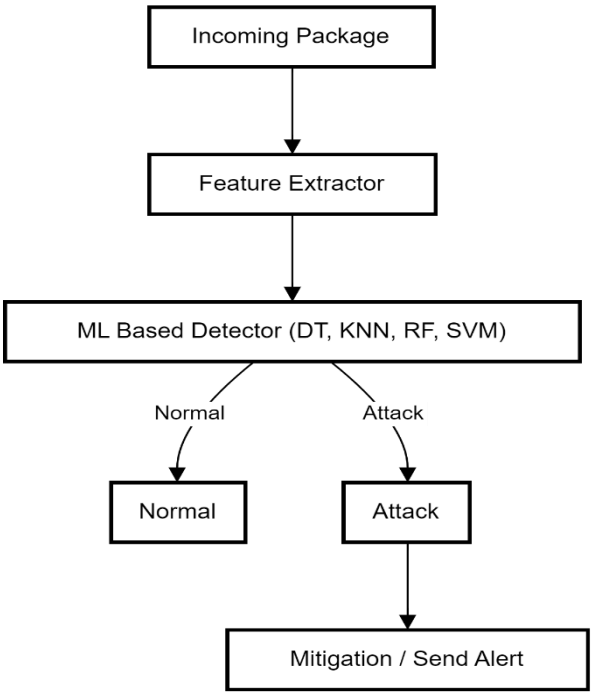

Fig-1: Proposed System of our system

The attack will be performed from the Mininet VM on the Ryu controller VM which will be of 3 types TCP, UDP and SYN flooding on which Ryu controller will apply ML and detect the DDoS attack and classify the normal and flooding packets after which the mitigation will be performed on the effected data packets along with classifying emails into spam and ham.

3.4. System Process

The process begins with the initialization of traffic flow. This could be network traffic in the case of DDoS mitigation or incoming emails in the case of spam classification. The traffic is then divided into normal traffic and attack traffic (or legitimate emails and spam emails). This could be done based on some initial filtering criteria. For normal traffic, feature extraction is performed to identify the characteristics of the normal network behavior or legitimate emails. For attack traffic, data preprocessing is done to clean and transform the raw attack data or spam emails into a usable format.

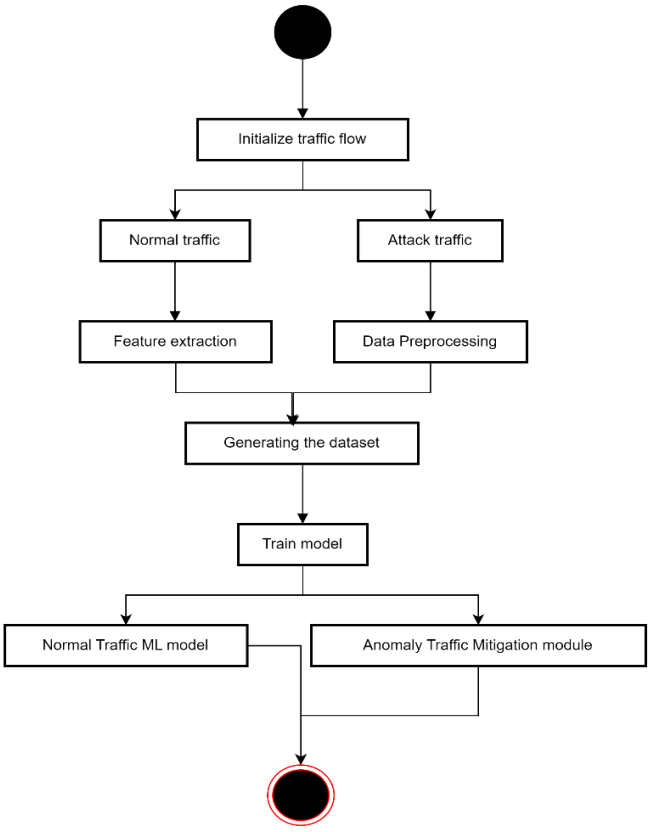

Fig-2: Flow of Cyber Security System

A dataset is generated from these processed data. This dataset would contain features extracted from both types of traffics (or emails) and their corresponding labels (normal/attack or legitimate/spam). A model is trained using this dataset. This could involve algorithms like SVM, KNN, DT, RF for DDoS detection or Naive Bayes for email spam classification. The choice of algorithm would depend on the specific requirements and constraints of the system. The trained model is then used to classify incoming traffic as either normal or an attack (or emails as either legitimate or spam). In the case of network traffic, there would be separate modules for handling normal traffic and mitigating anomaly traffic.

4. RESULTS

Integrating SDN for DDoS detection and ML algorithms (Random Forest, SVM, Decision Tree, kNN, Naive Bayes, Logistic Regression) alongside Naive Bayes for email spam detection show effective threat identification and mitigation. ML algorithm performance varied, aiding in selection for optimal threat detection. Naive Bayes demonstrated high accuracy in spam filtering. These findings underscore the efficacy of combining SDN and ML for proactive threat management, enhancing cybersecurity resilience against evolving threats.

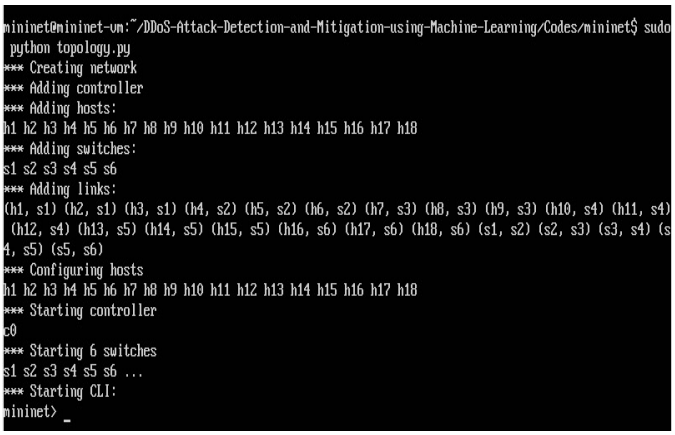

Fig-3: Creating Topology

Here the topology is being created for DDoS attack with 18 hubs and 16 switches interconnected with the Mininet controller for performing DDoS attack.

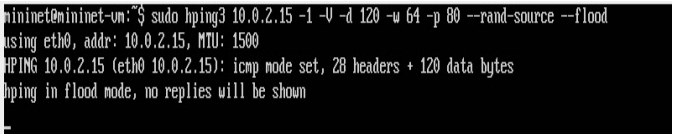

Fig-4: UDP Flooding

Hping3 is a python 3 command used for UDP flooding the network with sudo mentioning the root privileges than hping command -1 denotes UDP command, -V for verbose, -d for data size, -w for window size, -p for packet size and rand-source for hiding your IP address.

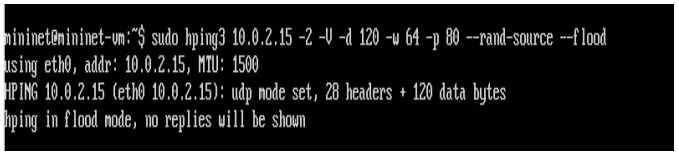

Fig-5: TCP Flooding

Hping3 is a python 3 command used for TCP flooding the network with sudo mentioning the root privileges than hping command -2 denotes UDP command, -V for verbose, -d for data size, -w for window size, -p for packet size and rand-source for hiding your IP address.

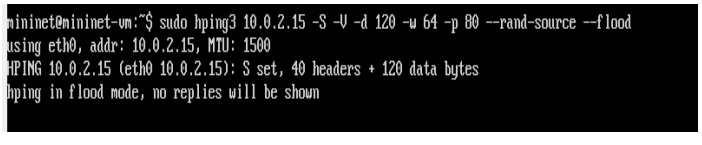

Fig-6: SYN Flooding

Hping3 is a python 3 command used for SYN flooding the network with sudo mentioning the root privileges than hping command -S denotes SYN command, -V for verbose,

-d for data size, -w for window size, -p for packet size and rand-source for hiding your IP address.

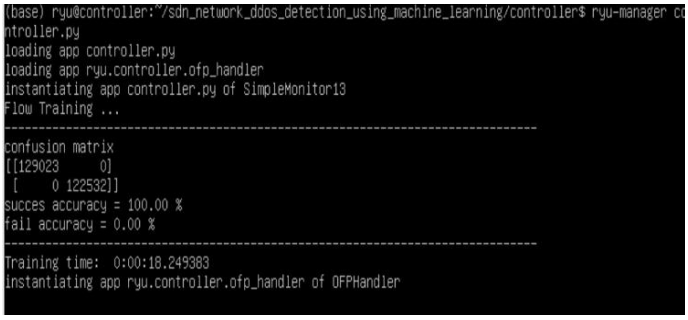

Fig-7: Confusion matrix for DDoS System

The controller executes a specific script, potentially named "controller.py," which houses a component called "SimpleMonitor13." This component trains a machine learning model using information about network flow. The displayed results indicate the model achieved perfect accuracy in attack detection and then we use this data to mitigate the DDoS attack, meaning it correctly identified all instances without any errors. This suggests the model is ready for deployment within the SDN environment, where it could potentially collaborate with Ryu's OpenFlow capabilities to pinpoint and block malicious traffic during a DDoS attack.

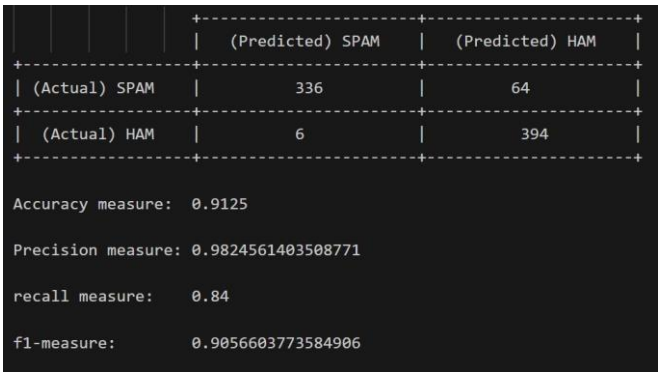

Fig-8: Confusion matrix for E-mail Spam Detection System

The output is from a Naive Bayes classifier used for email spam detection. The confusion matrix shows the model's predictions: 336 true positives, 394 true negatives, 64 false negatives, and 6 false positives. The accuracy is approximately 91.25%, indicating a high rate of correct predictions. Precision is about 98.25%, showing a low false positive rate. Recall is 84%, indicating the model's ability to find all positive samples. The F1 score, a measure of the model's overall effectiveness, is approximately 90.57%. The model tends to predict spam conservatively, leading to more false negatives than false positives. This approach reduces the risk of blocking legitimate emails but allows some spam emails to pass through.

5. CONCLUSIONS

The fusion of Software-Defined Networking (SDN) for detecting and countering Distributed Denial of Service (DDoS) attacks, in conjunction with a suite of Machine Learning (ML) algorithms like Random Forest, Support Vector Machine (SVM), Decision Tree, k-Nearest Neighbors (kNN), Naive Bayes, and Logistic Regression, alongside Naive Bayes for identifying email spam, offers a holistic strategy to fortify cybersecurity resilience. Through visual representations of network traffic patterns and ML algorithm efficacy, potential DDoS threats can be swiftly recognized and mitigated. Furthermore, visual aids such as word clouds and scatter plots facilitate the scrutiny of email content, enabling precise spam detection and filtering. By harnessing these visual cues, organizations can make informed choices about algorithm usage and enhancement, thereby strengthening their defence mechanisms against evolving cyber risks. This proactive and adaptable approach empowers stakeholders to stay one step ahead of cyber adversaries, protecting vital assets and upholding the reliability and accessibility of network infrastructure.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all those who contributed to the successful completion of this project. Foremost, we extend our deepest appreciation to Prof. Sonali Karthik our project Co-ordinater & our Guide Prof. Simran Patil whose guidance, support, and invaluable insights were instrumental throughout. Their expertise and encouragement were indispensable in overcoming challenges and steering us in the right direction. We are also immensely grateful to the entire project team for their dedication, collaboration, and hard work, which were vital in realizing this vision. Each team member's unique skills and contributions played a crucial role in the development, implementation, and testing phases of the project. Additionally, we acknowledge the support and resources provided by our institution, which were essential to the project's success.

REFERENCES

[1] A lightweight DDoS detection scheme under SDN context by Kun Jia, Chaoge Liu, Qixu Liu, Junnan Wang, Jiazhi Liu & Feng Liu

[2] Karki, Diwos & Dawadi, Babu. (2021). Machine Learning based DDoS Detection System in Software-Defined Networking

[3] N. Mageshkumar, A. Vijayaraj, N. Arunpriya, A. Sangeetha. Efficient spam filtering through intelligent text modification detection using machine learning,

Materials Today: Proceedings, Volume 64, Part 1, 2022.

[4] Deep Learning-Based Approach for Detecting DDoS Attack on Software-Defined Networking Controller by Amran Mansoor, Mohammed Anbar, Abdullah Ahmed Bahashwan, Basim Ahmad Alabsi, Shaza Dawood Ahmed Rihan, Shaza Dawood Ahmed Rihan.

[5] Jalal Bhayo, Syed Attique Shah, Sufian Hameed, Awais Ahmed, Jamal Nasir, Dirk Draheim, Towards a machine learning-based framework for DDOS attack detection in software-defined IoT (SD-IoT) networks,Engineering Applications of Artificial Intelligence Volume 123, Part C, 2023, 106432, ISSN 0952-1976.