VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

Project Report on

Real Time Detection and Mitigation of DDoS Attacks using Network Traffic Classification

In partial fulfillment of the Fourth Year (Semester–VII), Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2024-2025

Mrs. Geocey Shejy

Submitted by

Himesh Hotwani, D17C / 25
Neha Lotwani, D17C / 41
Aaryan Mahadik , D17C / 42
Pratham Shetty, D17C / 58

(2024-25)

VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

CERTIFICATE of Approval

This is to certify that **Himesh Hotwani (D17C 25), Neha Lotwani (D17C 41), Aaryan Mahadik (D17C 42), Pratham Shetty (D17C 58)** of Fourth Year Computer Engineering studying under the University of Mumbai has satisfactorily presented the project on "***Real Time Detection and Mitigation of DDoS Attacks using Network Traffic Classification***" as a part of the coursework of PROJECT-I for Semester-VII under the guidance of ***Mrs. Geocey Shejy*** in the year 2024-2025.

_____

Date

_____ _____

Internal Examiner External Examiner

_____ _____ _____

Project Mentor Head of the Department Principal

Dr. Mrs. Nupur Giri Dr. J. M. Nair

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Geocey Shejy** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr. (Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department

COURSE OUTCOMES FOR B.E PROJECT

Learners will be to:-

Course Outcome	Description of the Course Outcome
CO 1	Do literature survey/industrial visit and identify the problem of the selected project topic.
CO2	Apply basic engineering fundamental in the domain of practical applications FOR problem identification, formulation and solution
CO 3	Attempt & Design a problem solution in a right approach to complex problems
CO 4	Cultivate the habit of working in a team
CO 5	Correlate the theoretical and experimental/simulations results and draw the proper inferences
CO 6	Demonstrate the knowledge, skills and attitudes of a professional engineer & Prepare report as per the standard guidelines.

INDEX

ABSTRACT

Distributed Denial of Service (DDoS) attacks have become a significant and persistent threat to network security, capable of overwhelming systems, disrupting online services, and causing substantial financial and reputational losses for businesses. With the increasing scale and sophistication of these attacks, traditional security measures often fall short in providing effective protection. This project focuses on developing a robust, real-time system for the detection and mitigation of DDoS attacks using advanced network traffic classification techniques.

The proposed solution leverages machine learning algorithms to analyze network traffic and identify abnormal patterns indicative of potential DDoS attacks. By utilizing real-time monitoring and adaptive learning models, the system can detect various forms of DDoS attacks, including SYN floods, UDP floods, and HTTP-based attacks, even when they employ sophisticated evasion tactics. The system integrates automated mitigation strategies, such as traffic filtering and rate limiting, to minimize the impact of detected attacks, ensuring minimal disruption to legitimate network traffic.

The project employs a comprehensive approach, combining real-world network traffic datasets and simulated attack scenarios to train and validate the machine learning models. This report details the methodologies used for data collection, feature extraction, model training, and system integration. It also covers the design, implementation, and rigorous testing of the system to evaluate its performance in real-world scenarios.

Results from the project demonstrate the effectiveness of the proposed system in accurately detecting and mitigating DDoS attacks with low false-positive rates and minimal latency. The report discusses the challenges encountered during the development process, including data variability and model scalability, and presents solutions to address these issues. Finally, the report outlines the limitations of the current system and explores future directions for enhancing its adaptability, accuracy, and scalability, with a focus on integrating deep learning techniques and expanding the system's capabilities to handle evolving threats in dynamic network environments.

INTRODUCTION

1.1 Introduction

Network systems are increasingly vulnerable to various forms of cyberattacks, particularly Distributed Denial of Service (DDoS) attacks. These attacks overwhelm networks with excessive traffic, leading to significant disruptions and substantial financial losses. When system resources become overloaded due to DDoS assaults, legitimate users find themselves unable to access essential services. Traditional defense mechanisms often fall short in countering these sophisticated and large-scale threats, rendering them ineffective.

DDoS attacks specifically target the stability of networks by inundating systems with malicious traffic. This highlights the urgent need for a robust, real-time solution capable of reliably classifying network data and swiftly neutralizing such threats. The primary objective of this project is to develop a system that can rapidly detect and mitigate harmful activities, thereby safeguarding the network and ensuring continuous service availability.

The rising frequency and complexity of DDoS attacks pose a serious challenge to network security, as they can easily overwhelm systems with unauthorized traffic. Conventional defense strategies frequently struggle to fend off these advanced threats, underscoring the necessity for innovative approaches.

This project aims to address these challenges through network traffic classification, resulting in a real-time system for detecting and mitigating DDoS attacks. By employing machine learning algorithms, the system can differentiate between malicious and legitimate traffic, continuously monitoring network activity to detect anomalies. This proactive approach enables immediate responses to threats, helping to maintain service availability while minimizing disruptions to network performance.

1.2 Motivation

The motivation behind this project stems from the urgent need to enhance cybersecurity measures in light of the growing threat posed by Distributed Denial of Service (DDoS) attacks. As businesses and organizations increasingly rely on online services for their operations, the potential impact of DDoS attacks has become a pressing concern. The following factors illustrate the driving forces behind the development of a real-time DDoS detection and mitigation system:

1. **Increasing Frequency and Complexity of DDoS Attacks**: The frequency and complexity of DDoS attacks have risen dramatically in recent years. Attackers are employing more sophisticated methods, such as multi-vector attacks that combine different techniques to overwhelm systems. This evolution demands advanced detection and mitigation strategies that can keep pace with these evolving threats.
2. **Financial Implications for Businesses**: DDoS attacks can have severe financial consequences for organizations, ranging from direct costs associated with service downtime to loss of customer trust and reputation. For instance, studies have shown that even short periods of downtime can lead to significant revenue loss for e-commerce businesses. The motivation to develop effective DDoS defenses is underscored by the need to protect financial interests and maintain business continuity.
3. **Regulatory and Compliance Requirements**: Many industries are subject to regulatory requirements that mandate specific security measures to protect against cyber threats. Failure to comply with these regulations can lead to legal consequences and hefty fines. Developing a robust DDoS detection and mitigation system helps organizations meet these compliance standards and demonstrate their commitment to cybersecurity.
4. **Advancements in Machine Learning**: The rapid advancements in machine learning and data analysis technologies present an opportunity to enhance DDoS detection capabilities significantly. By leveraging these technologies, we can develop intelligent systems that can analyze vast amounts of network traffic in real time, learning to identify abnormal patterns indicative of DDoS attacks. This potential motivates the exploration of innovative solutions that combine traditional security measures with modern machine learning techniques.
5. **Desire for Continuous Improvement**: The cybersecurity landscape is constantly evolving, with attackers always seeking new methods to exploit vulnerabilities. The motivation to continuously improve DDoS detection and mitigation capabilities is driven by the need to stay ahead of potential threats. By developing a system that incorporates adaptive learning and real-time analysis, we aim to create a proactive security solution that evolves alongside the threat landscape.
6. **Real-World Impacts and Case Studies**: Numerous high-profile DDoS attacks have made headlines, disrupting major companies and causing widespread damage. Cases such as the attack on Dyn in 2016, which brought down several major websites, and the continuous targeting of financial institutions demonstrate the real-world consequences of inadequate DDoS defenses. The motivation to prevent similar incidents drives the urgency for developing a reliable detection and mitigation system.

1.3 Problem Definition

The real-time detection and mitigation of DDoS attacks present significant challenges, primarily due to the complexity and evolving nature of these attacks. Traditional rule-based systems often struggle to keep pace with the constantly changing tactics employed by attackers. Attackers use techniques such as IP spoofing, botnets, and multi-vector attacks, making it difficult for static rules to effectively identify and filter out malicious traffic.

The primary issue addressed in this project is the need for a solution that can detect DDoS attacks as they occur and respond quickly enough to prevent or minimize damage. Effective DDoS detection systems must be capable of analyzing massive amounts of network traffic in real time, identifying patterns indicative of an attack, and distinguishing between legitimate spikes in traffic and malicious activity. This requires a solution that is not only accurate but also scalable, adaptive, and efficient. The proposed system must be able to handle high volumes of network traffic without significantly impacting the performance of the network, while also being flexible enough to adapt to new attack patterns as they emerge.

1.4 Relevance of project

The relevance of this project extends beyond its technical contributions; it addresses critical challenges faced by organizations in today's interconnected digital environment. The following points highlight the significance of developing a real-time DDoS detection and mitigation system:

1. **Mitigating Operational Risks**: Organizations across various sectors are increasingly vulnerable to DDoS attacks, which can lead to severe operational disruptions. By providing a reliable mechanism for detection and mitigation, this project aims to minimize downtime, ensuring that services remain available to legitimate users even during an attack.
2. **Preserving User Trust**: For businesses that rely heavily on online transactions and services, maintaining user trust is paramount. Frequent or prolonged service interruptions can erode customer confidence. A robust DDoS mitigation system can help maintain service availability, thereby preserving user trust and satisfaction.
3. **Supporting Critical Infrastructure**: Many industries, such as finance, healthcare, and public services, rely on digital infrastructure to provide essential services. The potential impact of DDoS attacks on critical infrastructure can be catastrophic. This project supports the protection of such infrastructures by enhancing their resilience against these attacks.
4. **Cost-Effectiveness**: Implementing effective DDoS detection and mitigation strategies can significantly reduce the financial burden on organizations. The costs associated with service downtime, loss of customers, and damage to reputation can far exceed the investment required to implement a robust defense system. This project seeks to provide an economical solution that can protect against substantial losses.

5. **Facilitating Research and Development**: The project contributes to the broader field of cybersecurity by providing a framework for future research and development. By exploring innovative detection and mitigation techniques, it encourages further exploration of machine learning applications in network security, which could lead to advancements in other areas of cybersecurity.
6. **Aligning with Industry Trends**: As organizations increasingly adopt cloud computing, Internet of Things (IoT), and other digital innovations, the relevance of this project aligns with industry trends that emphasize the need for adaptive security measures. By focusing on real-time detection, the project addresses the requirements of modern network environments that demand agility and resilience.

1.5 Methodology Used

Data Handling & Feature Extraction

Collect data from various sources, including CAIDA, DARPA, and custom datasets related to network traffic. Utilize tools such as Wireshark for data capture and analysis. Preprocess and label the data to ensure accuracy and relevance, then extract key features such as packet size, flow duration, source/destination IP addresses, and statistical properties of traffic patterns to facilitate effective classification.

Traffic Classification & Detection

Implement both supervised machine learning algorithms (e.g., Random Forest, Support Vector Machine) and unsupervised methods (e.g., K-means clustering) to detect DDoS attacks. Develop a real-time classification system that utilizes stream processing techniques to ensure low-latency detection of anomalous traffic patterns, enabling prompt identification of potential threats.

Attack Mitigation & Optimization

Create strategies to filter suspicious traffic, such as implementing rate limiting and integrating IP reputation systems. Optimize the performance of the detection and mitigation system through parallel processing, load balancing, and efficient data handling to minimize the impact on legitimate users while effectively neutralizing threats.

Evaluation & Deployment

Assess the effectiveness of the detection and mitigation system using accuracy, precision, recall, and performance metrics. Establish a continuous retraining process for the machine learning models to adapt to evolving attack patterns. Deploy the solution in a cloud environment to ensure scalability and seamless integration with existing network infrastructure, facilitating ongoing monitoring and threat response.

LITERATURE SURVEY

2.1 Research Papers referred

1. **Paper Title:** "Online Network Traffic Classification Based on External Attention and Convolution by IP Packet Header" (2023)
   - **Abstract:** This paper presents a novel approach for classifying online network traffic using external attention mechanisms combined with convolutional techniques applied to IP packet headers. The method aims to enhance the accuracy of traffic classification in real-time, allowing for better detection of potential DDoS attacks.
   - **Inference Drawn:** The study demonstrates that incorporating attention mechanisms significantly improves classification performance by focusing on relevant features within the packet headers, thus enabling more accurate detection of malicious traffic.

2. **Paper Title:** "Mitigating Distributed Denial of Service Attack in Cloud Computing Environment Using Threshold-Based Technique" (2016)
   - **Abstract:** This research explores a threshold-based technique for mitigating DDoS attacks within cloud computing environments. The method involves setting specific thresholds for traffic volume to identify and mitigate potential attacks effectively.
   - **Inference Drawn:** The findings suggest that a proactive approach using predefined thresholds can help prevent service disruptions during DDoS attacks. However, the system's effectiveness is limited by its reliance on accurate threshold settings and may require adjustments based on traffic patterns.

3. **Paper Title:** "A Hierarchical Ensemble Graph Neural Network for DDoS Attack Detection" (2024)
   - **Abstract:** This paper introduces a hierarchical ensemble approach that utilizes graph neural networks to enhance DDoS attack detection capabilities. By modeling network traffic as a graph, the method captures complex relationships and interactions between different nodes, improving detection accuracy.
   - **Inference Drawn:** The research indicates that employing graph neural networks offers significant advantages over traditional methods, particularly in identifying intricate patterns associated with DDoS attacks. The hierarchical structure allows for scalable detection across large networks.

4. **Paper Title:** "A CNN-Based Approach for Efficient Detection and Mitigation of DDoS Attacks" (2023)
   - **Abstract:** This paper proposes a convolutional neural network (CNN)-based framework for the efficient detection and mitigation of DDoS attacks. The framework is designed to process network traffic in real-time, allowing for swift identification and response to potential threats.

- **Inference Drawn:** The results demonstrate that CNNs can accurately learn from traffic data to detect anomalies, enabling timely mitigation strategies and highlighting the value of deep learning in enhancing DDoS defenses.

2.2 Patent search

Patent Title: "Systems and Methods for Mitigating DDoS Attacks on Internet Protocol Networks"

- **Summary**: This patent describes a system using cryptographic techniques to filter IP packets, ensuring only validated traffic reaches the target, thereby mitigating DDoS attacks.
- **Link**: US Patent Application #20240333758

Patent Title: "Adaptive DDoS Defense Mechanism"

- **Summary:** This patent presents an adaptive defense mechanism that adjusts strategies based on evolving attack patterns, utilizing machine learning algorithms for improved responsiveness.
- **Link:** European Patent

Patent Title: "System and Method for Distributed Denial of Service (DDoS) Attack Prevention"

- **Summary:** This patent provides a method for identifying and mitigating DDoS attacks through distributed systems. It uses network monitoring to analyze traffic patterns and detect unusual activity, enabling a response mechanism that reroutes or blocks malicious traffic to protect network resources
- **Link:** USPTO Patent Search

2.3 Existing Systems

External Attention with Convolution on IP Packet Headers (2023):

- **System**: A traffic classification system using external attention mechanisms combined with convolutional techniques.
- **Description**: This system processes IP packet headers, leveraging attention mechanisms to focus on critical features. The convolutional layers further enhance classification accuracy, enabling efficient real-time traffic monitoring and detection of potential DDoS attacks.

Threshold-Based Technique for DDoS Mitigation in Cloud Environments (2016):

- **System**: A threshold-based monitoring and mitigation system.

- **Description**: This method sets specific thresholds for network traffic volume. When the traffic exceeds these thresholds, the system identifies and mitigates the attack. It requires fine-tuning and proactive adjustment of thresholds to adapt to changing traffic patterns and effectively prevent service disruptions.

Hierarchical Ensemble Graph Neural Network for DDoS Detection (2024):

- **System**: A graph neural network (GNN)-based hierarchical ensemble system.
- **Description**: This approach models network traffic as a graph, capturing complex interactions and relationships between nodes. The hierarchical structure makes it scalable, allowing for detection in large networks. This system excels at identifying intricate patterns associated with DDoS attacks, offering advantages over traditional detection methods.

CNN-Based DDoS Detection and Mitigation Framework (2023):

- **System**: A convolutional neural network (CNN)-based real-time detection and mitigation framework.
- **Description**: This system processes network traffic using CNNs to identify anomalies swiftly. It enables immediate responses to potential threats, emphasizing the application of deep learning for effective DDoS defense mechanisms.

2.4 Lacuna in the existing systems

- **High false positive rates:** Many existing systems struggle to accurately distinguish between legitimate traffic spikes and actual DDoS attacks.This can lead to unnecessary mitigation actions that affect normal users.
- **Delayed response:** Some systems have significant lag between detection and mitigation.This delay can allow attacks to cause damage before countermeasures are implemented.
- **Insufficient granularity in mitigation**: Some systems use overly broad mitigation techniques that can affect legitimate traffic.They may lack the ability to selectively filter or rate-limit specific traffic sources.
- **Limited machine learning integration:** While some advanced systems use ML, many still rely primarily on static rules or simple statistical analysis.This limits their ability to identify complex attack patterns or adapt to new threats
- **Inability to detect sophisticated attacks**: Traditional signature-based systems often fail to identify new or evolving attack patterns. They may miss low-and-slow attacks that don't trigger volume-based thresholds.

2.5 Comparison of existing systems and proposed area of work

Aspect	External Attention with Convolution (2023)	Threshold-Based Technique (2016)	Hierarchical Ensemble GNN (2024)	CNN-Based Detection Framework (2023)	Proposed Area of Work
Classification Approach	External attention + convolution on IP packet headers	Threshold-based	Graph neural networks (GNN)	Convolutional neural network (CNN)	External attention + additional ML techniques
Real-Time Capability	Yes, real-time classification	Limited by threshold settings	Yes, scalable across large networks	Yes, real-time detection and mitigation	Yes, focus on improving real-time performance
Focus	Focuses on critical features in IP packet headers	Focuses on traffic volume thresholds	Focuses on relationships between nodes in a network graph	Focuses on traffic anomalies via CNN	Further attention on improving critical feature extraction and speed
Mitigation Strategy	Detection of DDoS via feature extraction, no explicit mitigation	Mitigates via predefined traffic volume thresholds	Detects, but no explicit mitigation strategy	Detects and mitigates in real-time	Enhanced detection + efficient mitigation based on learning
Scalability	Limited by convolutional structure	Scales based on threshold adjustments	High scalability due to hierarchical GNN approach	Real-time but mitigation scalability may be a challenge	Improved scalability with attention-based focus and granularity
Accuracy	High accuracy via attention mechanisms	Relies on accurate threshold settings	High accuracy due to complex pattern detection	High accuracy from CNN's ability to learn from anomalies	Aims to enhance accuracy using attention and ML techniques
False Positives	Likely low due to feature-specific attention	High, due to fixed thresholds	Lower due to ability to model complex relationships	Moderate, depends on CNN training and configuration	Minimized by refining detection via adaptive learning
Lag Between Detection & Response	Minimal, designed for real-time	High, depends on thresholds	Minimal, scalable approach	Minimal due to real-time nature	Aim to reduce lag through optimized response mechanisms
Handling Sophisticated Attacks	Good at identifying relevant packet header features	Poor, cannot handle complex attacks beyond volume spikes	Excellent, can detect complex and low-and-slow attacks	Good, CNN can learn complex patterns	Targeting better handling of sophisticated attacks via improved models

Table 1. comparison of existing systems and proposed area of work

2.6 Focus Area

The Focus Area of the proposed work is to enhance real-time DDoS detection and mitigation by improving the classification of network traffic. Specifically, it aims to:

1. Leverage external attention mechanisms to focus on critical features within network traffic data, such as IP packet headers, to improve detection accuracy.
2. Reduce false positives by refining detection methods to better distinguish between legitimate traffic spikes and actual DDoS attacks.
3. Enhance the ability to detect sophisticated attacks, including low-and-slow or evolving attack patterns, which are often missed by traditional systems.
4. Optimize response time, minimizing the delay between detection and mitigation to prevent damage during an attack.
5. Increase scalability and granularity in mitigation efforts, ensuring that legitimate traffic is not unnecessarily disrupted while effectively blocking malicious traffic.

REQUIREMENTS

3.1 Proposed model

The proposed model for Real-Time Detection and Mitigation of DDoS Attacks using Network Traffic Classification is designed with a focus on preventing large-scale attacks by identifying and neutralizing malicious traffic in real time. The model leverages machine learning techniques to classify network traffic and distinguish between benign and malicious packets, with an emphasis on identifying Distributed Denial of Service (DDoS) attacks.

Key Components of the Model:

1. Data Collection and Traffic Monitoring:
    - The model captures and analyzes network traffic using packet sniffing tools like Wireshark or Tcpdump. Network traffic is continuously monitored on VM 3 (Monitoring Machine), which collects packet-level data from the Web Server (VM 1).
    - The traffic data is then processed to extract relevant features, including packet size, inter-arrival time, protocol type, number of connections, and bandwidth usage.
2. Feature Engineering:
    - From the captured traffic data, the model derives statistical and temporal features that are essential for identifying patterns typical of DDoS attacks.
    - Some of the extracted features include:
        - Flow Duration: Total time duration for which a specific connection persists.
        - Packet Size: Distribution of packet sizes, which may vary in attack scenarios.
        - Inter-arrival Time: Time between consecutive packets, which may decrease during attacks.
        - Source IP and Port Analysis: Identifying abnormal traffic from specific IP addresses or ports.
    - These features serve as inputs to machine learning classifiers.
3. Machine Learning-based Traffic Classification:
    - The core of the proposed model is built around supervised machine learning algorithms trained to differentiate between normal and attack traffic.
    - Several models have been implemented and tested, including:
        - Random Forest Classifier: Used for its ability to handle large datasets and interpret feature importance, making it effective for anomaly detection.
        - Support Vector Machine (SVM): Applied to classify traffic based on a decision boundary that separates normal from malicious flows.

        - Neural Networks: Utilized to capture complex patterns and relationships in the data that may indicate an ongoing DDoS attack.
        - K-Nearest Neighbors (KNN) and Decision Trees were also tested for comparison, though with lower accuracy in large datasets.
    - These models are trained using public DDoS attack datasets like CSE-CIC-IDS2018, which provide labeled traffic data (benign vs. malicious) for supervised learning.
4. Real-Time Traffic Classification:
    - Once trained, the models are deployed on VM 3 (Monitoring Machine), where they classify incoming network traffic in real-time. The system is designed to operate with minimal delay, ensuring that DDoS attacks are detected as they occur.
    - The classification model outputs a decision on whether the traffic flow is benign or malicious. For malicious flows, the system identifies the specific type of DDoS attack (e.g., SYN Flood, HTTP Flood, ICMP Flood).
5. Mitigation Mechanism:
    - Upon detection of a DDoS attack, the model initiates a mitigation strategy to prevent the malicious traffic from overwhelming the Web Server (VM 1). The mitigation can include:
        - Blocking IPs: Blacklisting IP addresses or IP ranges that are generating malicious traffic.
        - Rate Limiting: Reducing the traffic rate from suspected sources to mitigate the attack's impact.
        - Traffic Redirection: Redirecting malicious traffic to a honeypot or a dummy server to absorb the attack without affecting the web server.
    - These mitigation actions are automatically triggered by the detection system to neutralize the attack in real-time

3.2 Functional Requirements

- Traffic Classification: The system must classify network traffic in real-time into benign and malicious categories.
- DDoS Attack Detection: The system should detect volumetric DDoS attacks with high accuracy.
- Traffic Monitoring: Continuous monitoring of network traffic should be done to identify patterns indicative of a DDoS attack.
- Mitigation: On detecting an attack, mitigation strategies should be triggered automatically to reduce impact.
- Report Generation: Generate real-time reports and logs of network traffic, classification results, and mitigation actions.

3.3 Non-Functional Requirements

- Scalability: The system should be scalable to handle traffic from multiple sources.
- Performance: The detection system must have low latency to ensure real-time detection.
- Reliability: High reliability is required for detecting attacks without false positives or false negatives.
- Usability: The system interface for monitoring and management should be user-friendly.
- Security: The monitoring and mitigation components should be secure against tampering by attackers.
- Interoperability: The system should be capable of integrating with different network architectures and technologies.

3.4 Hardware & Software Requirements

Hardware:

- GPU (for machine learning model training)
- RAM (for handling large volumes of traffic data)

Software:

- Wireshark (for network traffic capture and analysis)
- Python (for developing machine learning models and implementing detection algorithms)
- C (for performance-critical components and system integration)

Tools:

- TensorFlow or Scikit-learn (for machine learning)
- Elasticsearch (for log storage and search)
- Kibana (for visualizing log data)

3.5.Technology and Tools utilized

- Machine Learning Models: Random Forest, Decision Tree, SVM (Support Vector Machine), and Neural Networks were utilized for classifying network traffic.
- Network Traffic Features: Statistical features from network flows such as packet size, inter-arrival time, and protocol distribution.
- Traffic Capture: Tools like Wireshark and Tcpdump were used to capture traffic for training the machine learning models.
- Virtual Machines: VMware or VirtualBox to simulate a multi-machine environment with dedicated VMs for different roles.
- Network Datasets: Public datasets like CSE-CIC-IDS2018 and CICIDS2017 were used for training the models.

3.6.Constraints of working

- Latency: The system may experience slight delays during traffic classification under very high traffic loads.
- Network Bandwidth: High-volume traffic during DDoS attacks can saturate bandwidth, affecting the detection model's performance.
- Resource Limitations: VM resources (like CPU and RAM) might constrain the system's ability to handle large-scale DDoS attacks.
- Model Accuracy: The accuracy of detection depends on the quality and quantity of the training dataset used for the machine learning model.

PROPOSED DESIGN

4.1 Block Diagram of the system

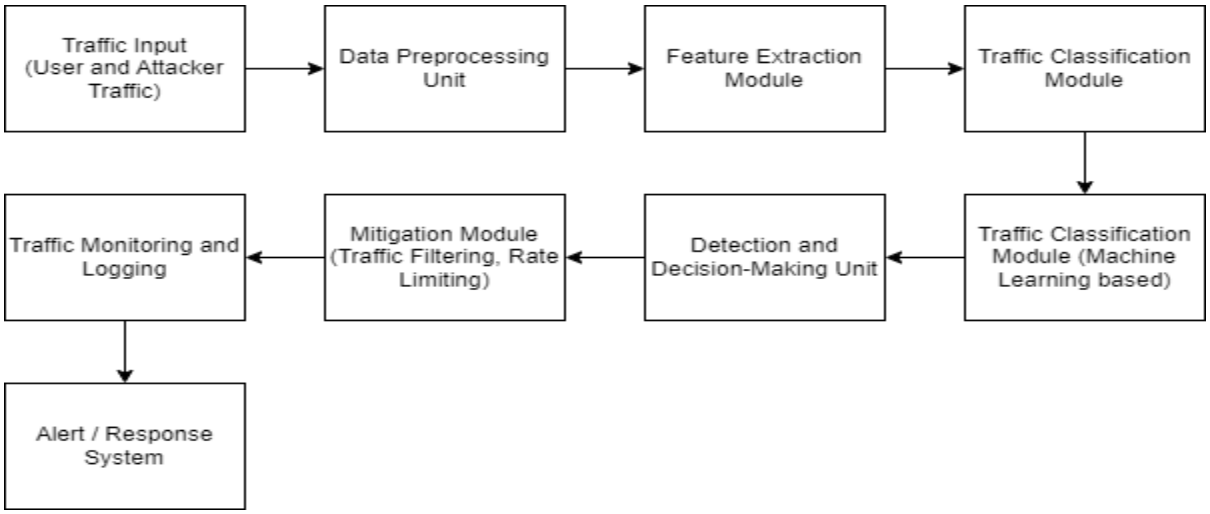

Fig. 1 Block Diagram

- Traffic Input: The system receives both legitimate user traffic and potentially malicious attacker traffic.
- Data Preprocessing Unit: Prepares the incoming traffic by cleaning and organizing it for further analysis.
- Feature Extraction Module: Identifies and extracts key features or patterns from the traffic to facilitate effective analysis.
- Traffic Classification Module: Classifies the traffic based on extracted features, distinguishing between legitimate and malicious traffic.
- Machine Learning-Based Traffic Classification: Uses machine learning models to enhance the accuracy of traffic classification, improving the detection of anomalies.
- Detection and Decision-Making Unit: Analyzes the classified traffic, detecting anomalies or threats, and decides on appropriate response actions.
- Mitigation Module (Traffic Filtering, Rate Limiting): Implements actions like traffic filtering and rate limiting to mitigate threats posed by malicious traffic.
- Traffic Monitoring and Logging: Continuously monitors network traffic and logs relevant data for further analysis, auditing, or incident investigation.
- Alert / Response System: Triggers alerts or initiates automated responses whenever suspicious or malicious traffic is detected, ensuring timely action.

4.2 Modular diagram

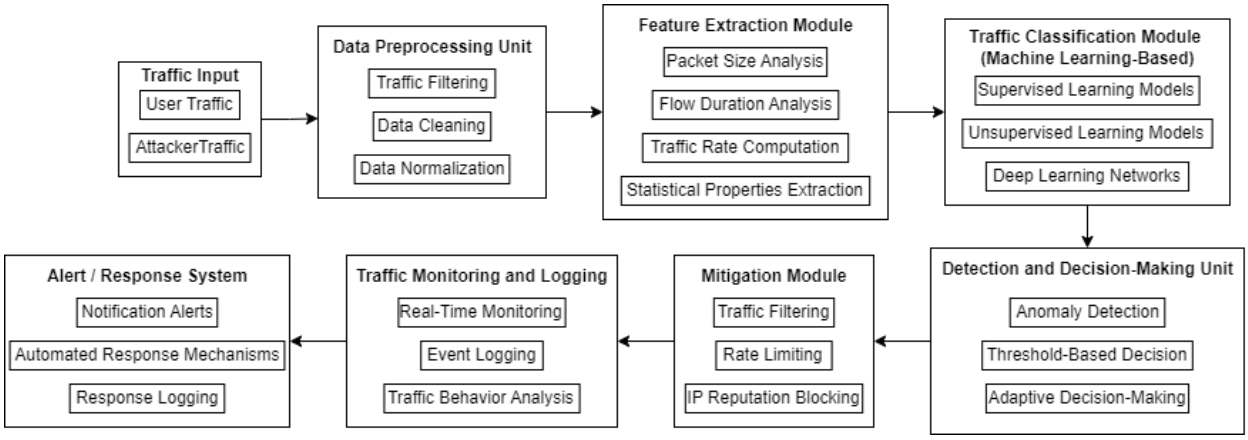

Fig. 2 Modular Diagram

- Traffic Input: Both user and attacker traffic enter the system.
- Data Preprocessing Unit: Cleans, filters, and normalizes traffic data for analysis.
- Feature Extraction Module: Analyzes packet size, flow duration, traffic rates, and other statistical properties.
- Traffic Classification Module (Machine Learning-Based): Classifies traffic using supervised, unsupervised, and deep learning models.
- Detection and Decision-Making Unit: Detects anomalies, makes threshold-based decisions, and applies adaptive decision-making strategies.
- Mitigation Module: Applies measures like traffic filtering, rate limiting, and IP reputation blocking to handle malicious traffic.
- Traffic Monitoring and Logging: Monitors traffic in real-time, logs events, and analyzes traffic behavior.
- Alert / Response System: Issues alerts, triggers automated responses, and logs system actions

4.3 Detailed Design (Flowchart)

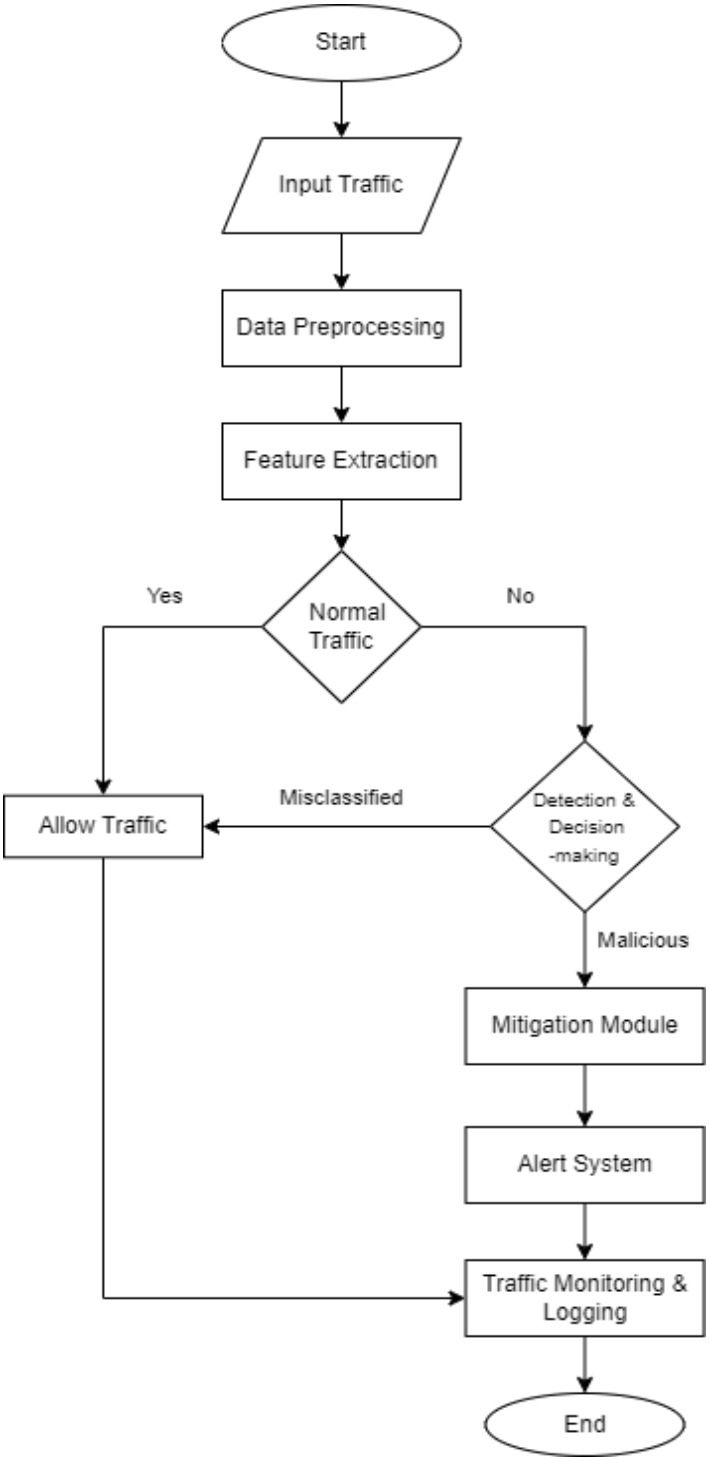

Fig. 3 Flow chart

- Input Traffic: Both user and attacker traffic enter the system.
- Data Preprocessing: Initial cleaning and preparation of traffic data for analysis.
- Feature Extraction: Key characteristics of the traffic are identified to aid in classification.
- Traffic Classification: The system categorizes traffic as either normal or suspicious.
- Decision Point:
    1. Normal Traffic: Allowed to proceed to the web server or targeted services.
    2. Suspicious Traffic: Redirected to a Detection & Decision-Making unit.
- Detection & Decision-Making:
    1. If the traffic is identified as malicious, the Mitigation Module is activated.
    2. If deemed legitimate but misclassified, it is allowed to pass through.
- Mitigation Module: Implements defense mechanisms, such as blocking or rate limiting, to counter detected attacks.
- Alert / Response System: Notifies administrators and can trigger automated responses when necessary.
- Traffic Monitoring & Logging: Continuously observes all traffic to ensure system effectiveness.
- End: The system operates in a continuous cycle to ensure ongoing detection and prevention of threats.

4.4. Proposed algorithms

To effectively detect and classify Distributed Denial of Service (DDoS) attacks, several machine learning algorithms were employed, each offering unique strengths in terms of accuracy, scalability, and feature handling. These models were chosen based on their ability to work with high-dimensional network traffic data and their robustness in classifying benign and malicious traffic. Below are the algorithms used in our project:

1. Gradient Boosting Machines (GBM)

Gradient Boosting algorithms, specifically XGBoost and LightGBM, were implemented due to their superior performance in handling large datasets and high-dimensional features. These algorithms work by building multiple decision trees sequentially, where each new tree corrects errors made by the previous ones. This iterative process improves the model's accuracy and robustness.

- XGBoost (Extreme Gradient Boosting): Known for its speed and performance, XGBoost efficiently handles large datasets. It uses a variety of optimization techniques, such as regularization and parallel processing, to reduce overfitting and improve model performance.
- LightGBM (Light Gradient Boosting Machine): LightGBM offers faster training times and lower memory usage than XGBoost. It is particularly well-suited for

high-dimensional datasets, making it ideal for network traffic data. LightGBM uses a leaf-wise tree growth approach, allowing it to achieve lower loss than depth-wise methods.

Both GBM models were crucial in detecting complex patterns and non-linear relationships between traffic features, enhancing the model's ability to identify DDoS attacks accurately.

2. Random Forest

Random Forest is a powerful ensemble learning method that constructs a multitude of decision trees at training time and outputs the class that is the mode of the individual trees' predictions. It is particularly robust against overfitting due to the averaging of multiple trees' results.

- Random Forest works well with high-dimensional datasets, making it ideal for analyzing network traffic data that contains a variety of features, such as packet size, source IPs, and flow duration.
- It also provides feature importance scores, which help in understanding the key indicators of DDoS attacks, such as abnormal traffic spikes, irregular connection times, and unusually large packet volumes.

Random Forest's ability to handle large datasets and provide insights into feature significance made it a valuable tool in our detection pipeline.

3. Deep Neural Networks (DNN)

Deep Neural Networks (DNN) were employed to capture the complex relationships and patterns inherent in network traffic data. DNNs consist of multiple layers of interconnected neurons that can model non-linear functions and relationships, making them highly suitable for detecting sophisticated attack patterns.

- The network architecture was designed to handle high-dimensional input data, which included various network traffic features.
- DNNs were particularly effective in identifying subtle and intricate differences between benign and malicious traffic flows, which may be difficult to detect using simpler models.
- However, DNNs required more computational resources and careful tuning of hyperparameters, such as learning rate, batch size, and the number of hidden layers.

Despite these challenges, DNNs provided strong predictive performance, especially for complex attack scenarios where traditional models might struggle.

4. Support Vector Machines (SVM)

Support Vector Machines (SVM) are widely used for classification tasks, particularly in high-dimensional spaces. In this project, SVM was applied to classify network traffic as benign or malicious based on the extracted features.

- SVM works by finding a hyperplane that maximizes the margin between the different classes (benign and malicious traffic).
- While SVM is effective in handling high-dimensional data, it struggles with large datasets due to its computational complexity. To mitigate this, we explored approximation techniques like the Linear SVM and kernel approximations, which allowed SVM to handle our large dataset more efficiently.

SVM showed good classification performance but required more resources and time for training when dealing with a large volume of network traffic data.

5. Logistic Regression with L1 Regularization

Logistic Regression was used with L1 regularization (Lasso) to provide a simple yet effective model for classifying DDoS attacks. L1 regularization introduces a penalty that forces certain coefficients to zero, effectively performing feature selection.

- Logistic Regression was chosen for its simplicity and interpretability, as well as its ability to handle high-dimensional data.
- The L1 regularization helped reduce the number of irrelevant features, focusing on the most important ones for DDoS attack detection.
- However, as a linear model, Logistic Regression does not capture complex, non-linear relationships as well as the other models in our pipeline.

Despite its limitations, Logistic Regression offered a lightweight, interpretable model that performed well when computational efficiency was a priority.

RESULTS AND DISCUSSION

The dataset provided includes the results of several tested machine learning models, including decision trees, logistic regression, support vector machines (SVMs), Naive Bayes, neural networks, and ensemble methods. These models were evaluated based on their accuracy (both on validation and test datasets) and total cost during testing and validation phases.

Key observations from the data:

1. **Decision Trees**:
    - Several variations of decision tree models (e.g., Models 1, 2.1, 2.2, 3.1, etc.) have been tested. The accuracy for some decision trees, such as Model 1, Model 2.1, and Model 3.1, reaches 100% on both validation and test sets, indicating near-perfect performance on this classification task.
    - For other decision trees, like Models 2.2 and 3.2, the accuracy decreases to around 98%, with a corresponding increase in total cost. Further decreases are seen in Models 2.3 and 3.3, where the accuracy drops to around 92%, indicating a potential trade-off between complexity and model performance.
2. **Logistic Regression**:
    - Efficient logistic regression models (e.g., Models 2.4, 3.4, and 3.5) show significantly lower accuracy, around 64-65% on both validation and test datasets. This suggests that logistic regression, particularly in its efficient form, may not be as well-suited for this specific task compared to other models like decision trees or SVMs.
    - Binary GLM Logistic Regression (Model 3.4) performs better than its efficient counterparts, achieving 88-89% accuracy but still not matching the performance of decision trees or SVMs.
3. **Support Vector Machines (SVMs)**:
    - SVMs (e.g., Models 3.9 to 3.14) exhibit varied performance. The highest accuracy achieved by SVM models reaches about 98%, with total costs varying between 932 and 6724. This indicates strong performance, though SVM models with kernel functions (Models 3.23 and 3.24) show slightly lower accuracy, around 97%.
    - SVM models are known for their ability to handle high-dimensional data, and their performance here seems competitive but slightly below the top-performing models like decision trees.
4. **Naive Bayes**:
    - Naive Bayes models (e.g., Models 3.7 and 3.8) show accuracy rates ranging from 85% to 97%. While not as accurate as decision trees or neural networks, Naive Bayes tends to be computationally efficient and may still be viable depending on the specific use case.

5. **Neural Networks**:
    - Neural networks (e.g., Models 3.18 to 3.22) achieve high accuracy, with some models reaching nearly 100%. These results suggest that neural networks are effective at handling this classification problem, especially given their capacity for modeling complex relationships in data.
    - The accuracy of neural networks on the test set (99-100%) also indicates that these models generalize well, maintaining high performance across different datasets.
6. **Ensemble Methods**:
    - Ensemble models (e.g., Models 3.15, 3.16, and 3.17) demonstrate near-perfect accuracy, reaching up to 100% on the test set. This suggests that combining multiple models may offer the best performance, as ensemble methods typically reduce variance and improve generalization.
    - The total cost associated with these models is also relatively low, indicating that they are efficient in terms of resource utilization.

General Trends:

- **Performance Variation**: There is significant variation in performance across different models. While decision trees, neural networks, and ensemble methods dominate in terms of accuracy, logistic regression and SVM models show lower performance on this task. However, SVMs still offer competitive accuracy in many cases.
- **Accuracy and Cost Relationship**: Higher accuracy generally corresponds to lower total cost, particularly for top-performing models like decision trees, neural networks, and ensemble methods. This suggests that the most accurate models are also the most resource-efficient, which is ideal for real-time or large-scale implementations.

In conclusion, the models tested demonstrate a wide range of accuracy and resource consumption, with decision trees, neural networks, and ensemble methods showing the highest potential for this classification task. These findings are consistent with general expectations in machine learning, where complex models tend to outperform simpler models (e.g., logistic regression) on more intricate tasks.

Favorite	Model Number	Model Type	Status	Accuracy % (Validation)	Total Cost (Validation)	Accuracy % (Test)	Total Cost (Test)
0	1	Tree	Tested	99.9816106	12	100	0
0	2.1	Tree	Tested	99.9816106	12	100	0
0	2.2	Tree	Tested	98.12581411	1223	98.19162631	295
0	2.3	Tree	Tested	91.96076929	5246	91.98798504	1307
0	2.4	Efficient Logistic Regressic	Tested	64.97433147	22856	64.16968062	5845
0	2.5	Efficient Linear SVM	Tested	64.96207187	22864	64.1451603	5849
0	3.1	Tree	Tested	99.9816106	12	100	0
0	3.2	Tree	Tested	98.12581411	1223	98.19162631	295
0	3.3	Tree	Tested	91.96076929	5246	91.98798504	1307
0	3.4	Binary GLM Logistic Regres	Tested	88.95870048	7205	88.69000184	1845
0	3.5	Efficient Logistic Regressic	Tested	64.97433147	22856	64.16968062	5845
0	3.6	Efficient Linear SVM	Tested	64.96207187	22864	64.1451603	5849
0	3.7	Naive Bayes	Tested	85.55819477	9424	85.25102679	2406
0	3.8	Naive Bayes	Tested	97.02551529	1941	97.52344756	404
0	3.9	SVM	Tested	89.69580875	6724	89.55434316	1704
0	3.1	SVM	Tested	97.83924603	1410	97.86060197	349
0	3.11	SVM	Tested	98.57175695	932	98.50426041	244
0	3.12	SVM	Tested	98.14113861	1213	98.41843928	258
0	3.13	SVM	Tested	97.85610298	1399	97.93416294	337
0	3.14	SVM	Tested	93.84414987	4017	94.22546435	942
0	3.15	Ensemble	Tested	99.44831814	360	99.47281309	86
0	3.16	Ensemble	Tested	99.9938702	4	100	0
0	3.17	Ensemble	Tested	98.12581411	1223	98.19162631	295
0	3.18	Neural Network	Tested	99.8145736	121	99.84674799	25
0	3.19	Neural Network	Tested	99.87740403	80	99.95095936	8
0	3.2	Neural Network	Tested	99.80691135	126	99.92030896	13
0	3.21	Neural Network	Tested	99.86667688	87	99.92030896	13
0	3.22	Neural Network	Tested	99.83296299	109	99.77931711	36
0	3.23	Kernel	Tested	97.81012949	1429	97.92190278	339
0	3.24	Kernel	Tested	97.15730595	1855	97.20468338	456

Fig. 4 CSV file containing accuracy values of the algorithms trained and tested on the dataset

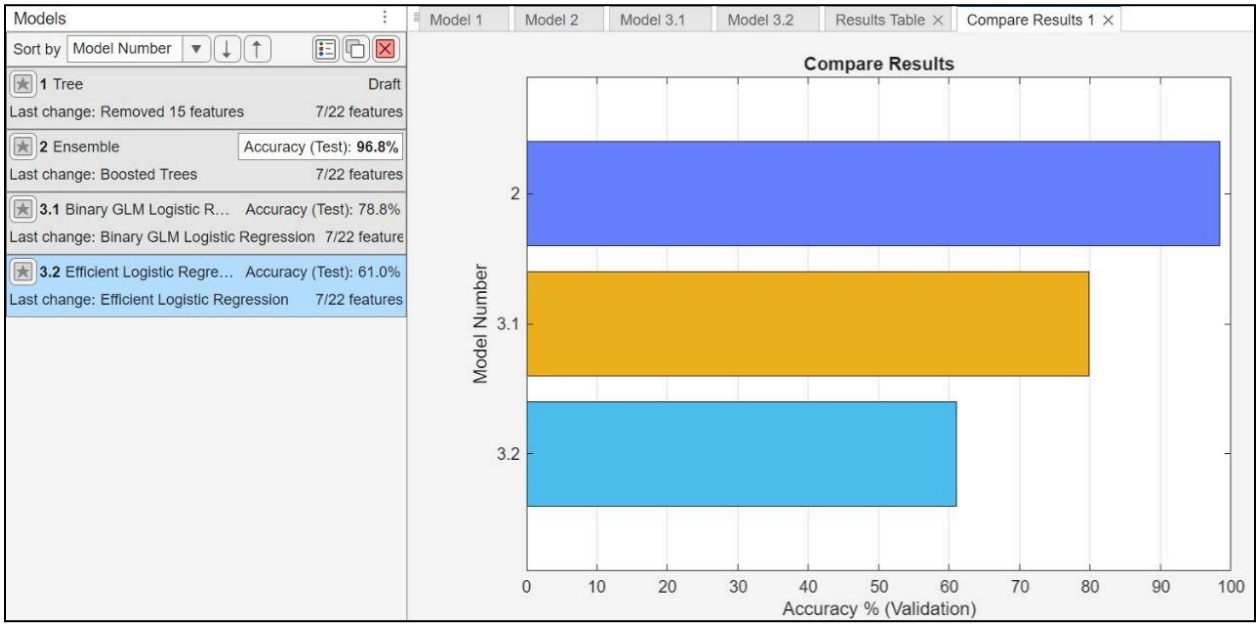

Fig. 5 Result comparison in MATLAB's Classification Learner

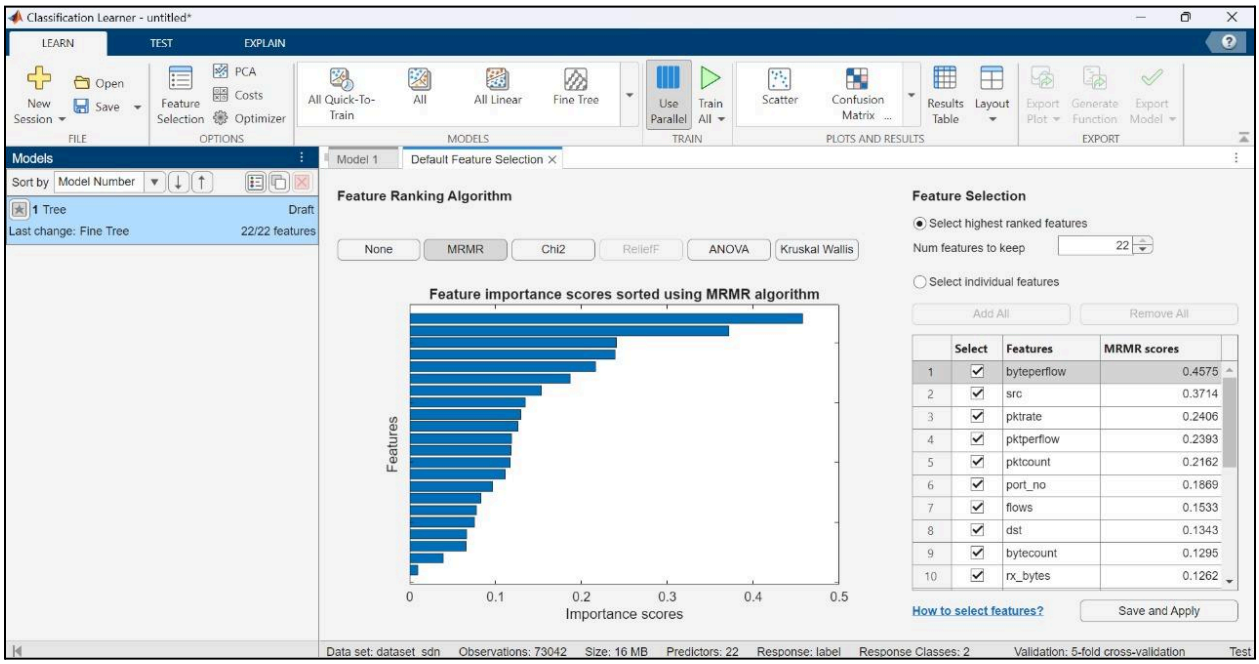

Fig. 6 Top features of the dataset gathered using the Minimum Redundancy Maximum Relevance (mRMR) algorithm

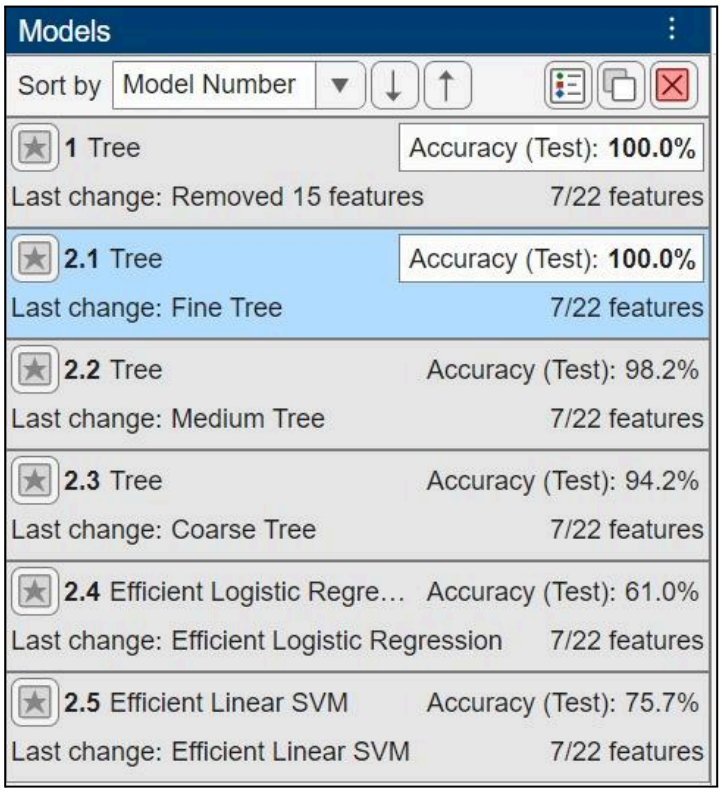

Fig. 7 Accuracy values of some of the algorithms

PLANS FOR THE NEXT SEMESTER

- Detection can be done by using various types of deep learning techniques and neural networks.
- The future plan is to be able to detect and mitigate DDoS attacks.
- Enhance Real-Time Detection: Aim for faster response times during attacks.
- Develop Mitigation Strategies: Create automated response mechanisms.
- Use Advanced Feature Extraction: Improve data preprocessing for accuracy.
- Integrate Continuous Learning: Adapt models based on emerging attack patterns.
- Conduct Performance Evaluations: Regularly assess effectiveness against new threats.
- Focus on Scalability: Ensure models can handle increased traffic loads.

CONCLUSION

The increasing frequency of Distributed Denial of Service (DDoS) attacks highlights the need for advanced, scalable solutions to protect critical online services. Traditional security measures often fail to detect and mitigate such large-scale attacks in real time, especially as attackers evolve. The proposed approach addresses these challenges with enhanced detection, sophisticated mitigation strategies, and scalability, making it ideal for countering modern DDoS threats.

Our model uses cutting-edge machine learning to classify network traffic with high precision, ensuring timely threat detection. By capturing key traffic features and employing a feedback loop for continuous learning, the system adapts to new attack patterns, improving accuracy. Multiple virtual machines for traffic capture, analysis, and mitigation enable seamless integration into diverse network architectures, providing flexibility and reliability.

The model's real-time mitigation strategies—such as IP blocking, rate limiting, and traffic redirection—prevent malicious traffic from overwhelming servers, ensuring immediate responses to attacks while minimizing downtime. Its scalable design handles large traffic volumes, crucial for high-traffic environments. The system’s adaptability to various network settings and compatibility with existing security frameworks make it a versatile solution for organizations of all sizes.

This development enhances network stability by overcoming the limitations of existing systems, which often struggle with the scale and complexity of modern DDoS attacks. It ensures robust protection, contributing to the resilience of essential online services. Future improvements, including deeper learning-based detection and cloud-based deployment, could further strengthen its capabilities, making it a critical tool in combating evolving cyber threats.

REFERENCES

[1] Najar, Ashfaq Ahmad, and S. Manohar Naik. "Cyber-secure SDN: A CNN-based approach for efficient detection and mitigation of DDoS attacks." Computers & Security 139 (2024): 103716.

[2] Azab, Ahmad, et al. "Network traffic classification: Techniques, datasets, and challenges." Digital Communications and Networks 10.3 (2024): 676-692.

[3] Hu, Yahui, et al. "Online network traffic classification based on external attention and convolution by IP packet header." Computer Networks (2024): 110656.

[4] Bakar, Rana Abu, et al. "FTG-Net-E: A hierarchical ensemble graph neural network for DDoS attack detection." Computer Networks 250 (2024): 110508.

[5] Bharot, Nitesh & Verma, Priyanka & Suraparaju, Veenadhari & Gupta, Sanjeev. (2016). Mitigating Distributed Denial of Service Attack in Cloud Computing Environment using Threshold based Technique. Indian Journal of Science and Technology. 9. 10.17485/ijst/2016/v9i38/98811.

APPENDIX

a. List of Figures

b. List of tables