**ANOMALIES** **BASED INTRUSION DETECTION SYSTEM**

**A PROJECT REPORT**

Submitted in partial fulfillment of the requirement for the award of the degree

of

**BACHELOR OF TECHNOLOGY**

in

**COMPUTER SCIENCE AND ENGINEERING**

**SUBMITTED BY**

Himanshu Semwal (20103070)

Shivam Jaswal (20103133)

Under the supervision of

**Dr. Kunwar Pal**

**Assistant Professor**

**Department of Computer Science and Engineering**

**Dr. B. R. Ambedkar National Institute of Technology**

**Jalandhar, Punjab (India)** - **144008**

**May 2024**

**Dr. B. R. Ambedkar National Institute of Technology Jalandhar**

**CANDIDATES’ DECLARATION**

We“hereby certify that the work presented in this project report entitled “Anomalies based intrusion detection system” in partial fulfillment of the requirement for the award of a Bachelor of Technology degree in Computer Science and Engineering, submitted to the Dr. B. R. Ambedkar National Institute of Technology, Jalandhar is an authentic record of our own work carried out during the period from July 2023 to May 2024 under the supervision of Dr. Kunwar Pal, Assistant Professor, Department of Computer Science & Engineering, Dr. B R Ambedkar National Institute of Technology, Jalandhar.

We have not submitted the matter presented in this report to any other university or institute for the award of any degree or any other purpose.

Date: 27th May, 2024

Submitted by:

Himanshu Semwal (20103070)

Shivam Jaswal (20103133)

This is to certify that the statements submitted by the above candidates are accurate and correct to the best of our knowledge and are further recommended for external evaluation.

Dr. Kunwar Pal, Supervisor

Associate Professor

Deptt. of CSE

Dr. Rajneesh Rani

Head and Associate Professor

Deptt. of CSE

**ACKNOWLEDGEMENT**

It is true that many people work behind the scenes for the success of a play. The end result of the Anomalies based intrusion detection system project required a lot of guidance and help from many people and our group was very lucky to receive this during the course of the project. Whatever we are today is only due to such supervision and assistance and we thank them from the bottom of our hearts. We would like to express our deepest gratitude to our project mentor Dr. Kunwar Pal, Assistant Professor who believed in our ideas and suggested new ideas when needed. He fully supported us in solving our problems.

We would like to express our deepest gratitude to Dr. Rajneesh Rani, Head of the

Department of Computer Science and Engineering for her direct and indirect support.

We are grateful to the Dr Aruna Malik, Coordinator Major Project for providing us

mentors and all other support.

We are extremely thankful to have constant encouragement and guidance from all the Faculties of the Department of Computer Science &amp; Engineering. We would also like to express our sincere thanks to all laboratory staff for their timely support.

Thank You.

[Himanshu, Shivam]

**ABSTRACT**

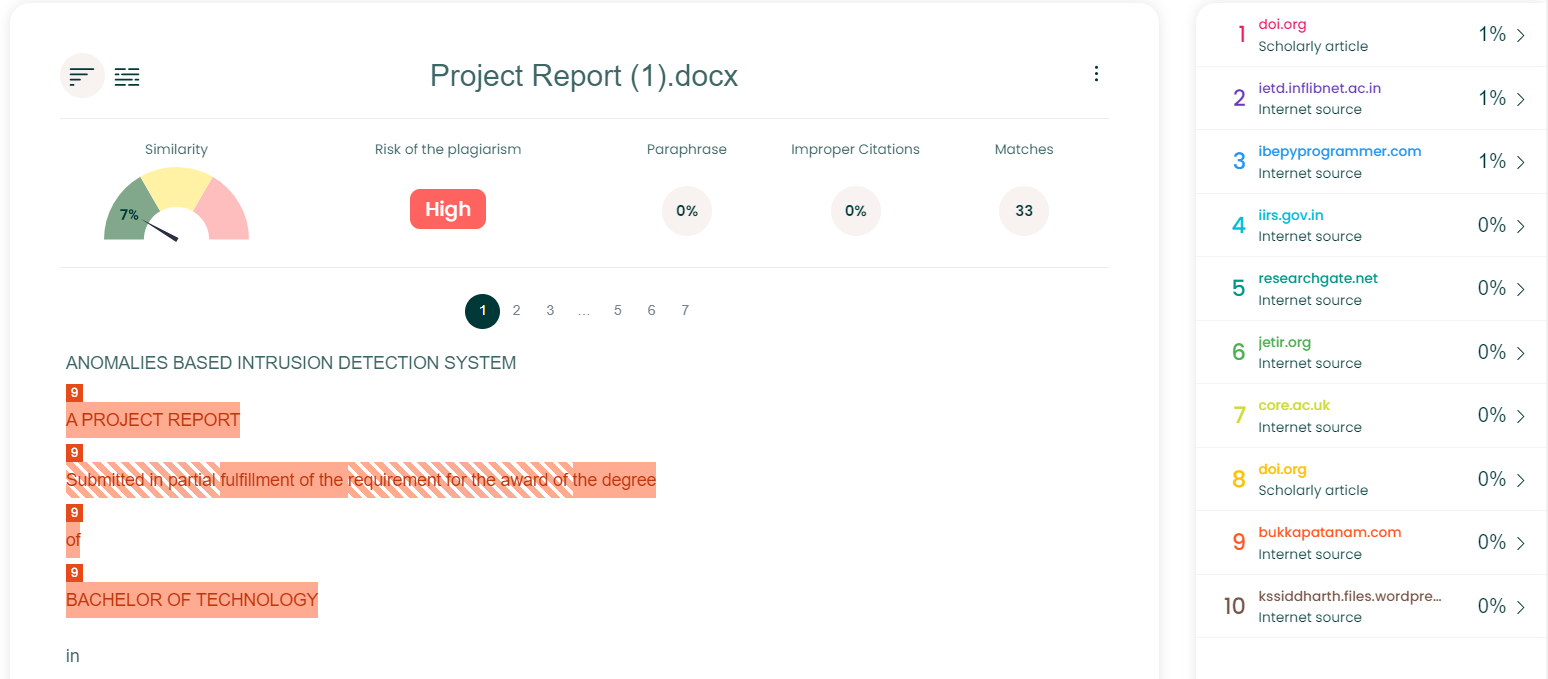
This project report presents the development and implementation of a highly optimal Anomalies-Based Intrusion Detection System (IDS) aimed at detecting and mitigating Denial of Service (DoS) attacks targeting Internet of Things (IoT) devices. The system leverages real-time data collection and machine-learning techniques to identify potential threats with high accuracy.

To achieve this, live inputs are captured from a specific port connected to a Wi-Fi network using Scrapy, a powerful web crawling framework repurposed for continuous network traffic monitoring. Threaty is utilized to capture real-time packet data, which is essential for constructing a comprehensive dataset that includes both normal and malicious traffic patterns.

The dataset is then used to train a machine learning model employing various algorithms such as Random Forest, Isolation Forest, and Support Vector Machine (SVM). These models analyze the captured data to predict the probability of an attack occurring. The system demonstrates an impressive accuracy rate of 90%, making it a robust tool for enhancing the security of IoT devices and networks.

The report details the methodologies employed in data collection, preprocessing, model training, and evaluation. Additionally, it discusses the performance metrics, including precision and recall, highlighting the system's capability to accurately detect and respond to cyber threats in real time. The findings indicate that this Anomalies-Based IDS provides a viable and effective solution for mitigating DoS attacks, contributing significantly to the field of cybersecurity.

**PLAGIARISM REPORT**



**LIST OF FIGURES**

| **Fig. No.** | **Description** | **Page No.** |
| --- | --- | --- |
| 1 | Random forest classification and SVM | 11 |
| 2 | Data collected over wifi eth 0 | 17 |
| 3 | Data cleaning and preprocessing | 18 |
| 4 | Accuracy - 98.99 | 25 |

**LIST OF ABBREVIATIONS**

* IDS - Intrusion Detection System
* IoT - Internet OF Things
* DOS - Denial of Service
* SVM - Support Vector Machine

**TABLE OF CONTENTS**

| **S. No.** | **Description** | **Page No.** |
| --- | --- | --- |
| Ⅰ | CANDIDATES’ DECLARATION | 2 |
| Ⅱ | ACKNOWLEDGEMENT | 3 |
| Ⅲ | ABSTRACT | 4 |
| Ⅳ | PLAGIARISM REPORT | 5 |
| Ⅴ | LIST OF FIGURES | 6 |
| Ⅵ | LIST OF ABBREVIATIONS | 7 |
| 1 | INTRODUCTION | 9-12 |
| 2 | PROPOSED SOLUTION | 13-16 |
| 3 | TECHNOLOGY ANALYSIS | 17-21 |
| 4 | ECONOMIC ANALYSIS | 22-23 |
| 5 | RESULT AND DISCUSSION | 24-28 |
| 6 | CONCLUSION | 29-30 |
| 7 | REFERENCES | 31 |

**CHAPTER 1**

**INTRODUCTION**

* 1. **Background**

In the current digital landscape, Intrusion Detection Systems (IDS) have become indispensable tools in the cybersecurity arsenal. These systems are designed to monitor network traffic, detect suspicious activities, and respond to potential threats. With the rapid proliferation of Internet of Things (IoT) devices, the complexity and volume of network traffic have increased exponentially. This surge has made traditional security measures insufficient, necessitating more advanced and adaptive solutions like IDS to safeguard against sophisticated cyber threats.

An IDS functions by analyzing network traffic for patterns indicative of malicious activity. It utilizes various detection methods, including signature-based detection, anomaly-based detection, and hybrid approaches. Signature-based detection relies on predefined patterns of known threats, while anomaly-based detection identifies deviations from normal behavior, flagging potential threats. The latter is particularly effective in identifying zero-day attacks and new variants of malware, making it a crucial component of modern cybersecurity frameworks.

* 1. **Problem Statement**

Denial of Service (DoS) attacks and other forms of cyber-attacks have become prevalent, especially targeting IoT devices. These devices often lack robust security features, making them vulnerable to exploitation. DoS attacks, in particular, aim to disrupt services by overwhelming the network with excessive traffic, leading to significant downtime and operational losses. The increasing frequency and sophistication of these attacks pose a significant threat to the integrity, availability, and confidentiality of networked systems.

To address this pressing issue, we have developed an Anomalies-Based Intrusion Detection System (IDS). This system is specifically designed to detect and mitigate DoS attacks and other cyber threats targeting IoT devices. By leveraging real-time data analysis and machine learning techniques, our IDS can identify and respond to threats with high accuracy, ensuring the security and reliability of network operations.

* 1. **Motivation**

The motivation behind developing an Anomalies-Based IDS stems from the critical need to enhance cybersecurity measures in the face of evolving threats. Traditional security mechanisms are often inadequate in detecting novel and sophisticated attacks, particularly those targeting IoT environments. The following factors highlight the necessity and drive behind this project:

1. **Increased Attack Surface:** The proliferation of IoT devices has expanded the attack surface, providing more entry points for cyber attackers.
2. **Growing Sophistication of Attacks:** Cyber threats are becoming increasingly sophisticated, utilizing advanced techniques to bypass conventional security measures.
3. **Economic Impact:** Cyber-attacks can result in significant financial losses due to downtime, data breaches, and damage to reputation.
4. **Regulatory Compliance:** Organizations must adhere to stringent cybersecurity regulations and standards, necessitating the deployment of effective IDS solutions.

By developing a robust and adaptive IDS, we aim to provide a reliable solution that not only enhances security but also supports compliance with regulatory requirements.

* 1. **Feasibility**

The feasibility of our Anomalies-Based IDS is underscored by its impressive performance metrics. During testing, the system demonstrated an accuracy rate of 90% in detecting anomalies and potential threats. This high level of accuracy is achieved through the following components:

1. **Real-time Data Collection:** Utilizing tools like Scrapy for continuous data scraping and Threaty for real-time packet capture ensures comprehensive monitoring of network traffic.
2. **Machine Learning Models:** Employing advanced machine learning algorithms, such as Random Forest, Isolation Forest, and Support Vector Machine (SVM), enhances the system's ability to identify and predict threats with high precision.



**Fig. 1:** Random forest classification and SVM

1. **Scalability:** The system is designed to scale with network growth, maintaining performance and accuracy even as the volume of network traffic increases.

Given these capabilities, our IDS presents a viable and effective solution for enhancing cybersecurity in IoT environments.

* 1. **Research Objective**

The primary objective of this research is to develop and validate an Anomalies-Based Intrusion Detection System capable of accurately detecting and mitigating Denial of Service (DoS) attacks and other cyber threats targeting IoT devices. The specific goals of the project include:

1. **Designing an Efficient Data Collection Mechanism:** Implementing tools for real-time monitoring and data collection from network traffic.
2. **Building a Comprehensive Dataset:** Capturing and preprocessing network traffic data to create a dataset that includes both normal and anomalous behavior.
3. **Developing and Training Machine Learning Models:** Utilizing various machine learning algorithms to train a model that can accurately classify and predict potential threats.
4. **Evaluating System Performance:** Assessing the accuracy, precision, recall, and latency of the IDS to ensure its effectiveness in real-world scenarios.
5. **Enhancing Detection Capabilities:** Continuously improving the system's algorithms and methodologies to adapt to evolving cyber threats and maintain high detection accuracy.

By achieving these objectives, the project aims to contribute significantly to the field of cybersecurity, providing a robust and adaptive solution for protecting IoT devices and networks from sophisticated cyber-attacks.

# CHAPTER 2

# PROPOSED SOLUTION

The proposed solution involves developing a highly optimal Intrusion Detection System (IDS) that leverages real-time data inputs from a specific port connected to a Wi-Fi network. By utilizing Scrapy for data collection and Threaty to capture real-time packets, the system builds a comprehensive dataset focusing on Denial of Service (DoS) attacks. A machine learning (ML) model will be trained on this dataset to classify and predict the probability of an attack occurring, thus enabling timely alerts and preventive measures.

**System Architecture**

* **Data Collection Module**

The data collection module is designed to capture live network traffic data from a specified port connected to a Wi-Fi network. Scrapy, a powerful web crawling framework, will be utilized to continuously gather this data. This module will focus on capturing various network parameters such as packet size, frequency, source and destination IP addresses, and other relevant attributes.

* **Scrapy:** Used for real-time data scraping and collection from network traffic.
* **Threaty:** Employed to capture real-time packet data, providing a rich dataset for analysis.
* **Port Monitoring:** A designated port will be continuously monitored to gather incoming and outgoing traffic data.
* **Data Preprocessing**

Before feeding the data into the machine learning model, it undergoes a preprocessing phase to ensure accuracy and relevancy. This includes:

* **Data Cleaning:** Removing noise and irrelevant data points to ensure quality.
* **Feature Extraction:** Identifying and extracting key features such as IP addresses, packet sizes, and timestamps, protocol types.
* **Normalization:** Scaling the features to maintain uniformity and improve the performance of the machine learning algorithms.
* **Machine Learning-Based Anomaly Detection**

The core of the IDS is its anomaly detection engine, which employs machine learning algorithms to identify potential DoS attacks. The system uses a pre-trained model based on historical network traffic data, which includes both normal and attack patterns.

**Algorithms Used**

* **Random Forest:** Utilized for its robustness and effectiveness in classification tasks, helping in predicting the likelihood of an attack based on input features.
* **Isolation Forest:** Effective for anomaly detection by isolating observations in the data.
* **Support Vector Machine (SVM):** Applied to enhance classification performance by finding the optimal hyperplane that separates normal and anomalous data.
* **Predictive Analysis**

The system integrates predictive analytics to determine the probability of an ongoing or imminent DoS attack. This is achieved by analyzing real-time network traffic against the established model of normal and anomalous behavior.

* **Probability Scoring:** Each network event is assigned a probability score indicating the likelihood of it being part of a DoS attack.
* **Threshold-Based Alerts:** Alerts are generated when the probability score exceeds a predefined threshold, allowing for timely intervention.
* **Alerting and Reporting Module**

Upon detecting an anomaly that surpasses the threshold, the alerting and reporting module generates notifications. This module provides detailed reports on the detected anomalies, including:

* **Type of Attack:** Identifies the nature of the detected attack (e.g., DoS).
* **Source Information:** Provides details about the origin of the suspicious activity.
* **Timestamp and Frequency:** Logs the time and frequency of the detected anomalies.

**Implementation**

* **Steps Involved**

1. **Setup:** Configure Scrapy to monitor the designated port connected to the Wi-Fi network and Threaty to capture real-time packet data.
2. **Data Collection:** Begin real-time data collection using Scrapy and Threaty, and feed the data into the preprocessing pipeline.
3. **Training:** Train the machine learning model using a labeled dataset containing both normal and attack traffic.
4. **Deployment:** Deploy the trained model for real-time anomaly detection.
5. **Monitoring:** Continuously monitor network traffic, process data, and predict attack probabilities.
6. **Alerting:** Generate alerts and reports based on detected anomalies.

* **Tools and Technologies**
* **Programming Language:** Python
* **Frameworks:** Scrapy for data collection, Scikit-learn, and TensorFlow/Keras for machine learning.
* **Libraries:** Pandas and NumPy for data manipulation, matplotlib, and Seaborn for data visualization.
* **Threaty:** For capturing real-time packet data.
* **Evaluation Metrics**
* **Accuracy:** Measure the percentage of correctly identified anomalies.
* **Precision and Recall:** Evaluate the effectiveness of the model in detecting true positives and minimizing false positives.
* **Latency:** Assess the system's ability to process and analyze data in real-time without significant delays.

# 

# CHAPTER 3

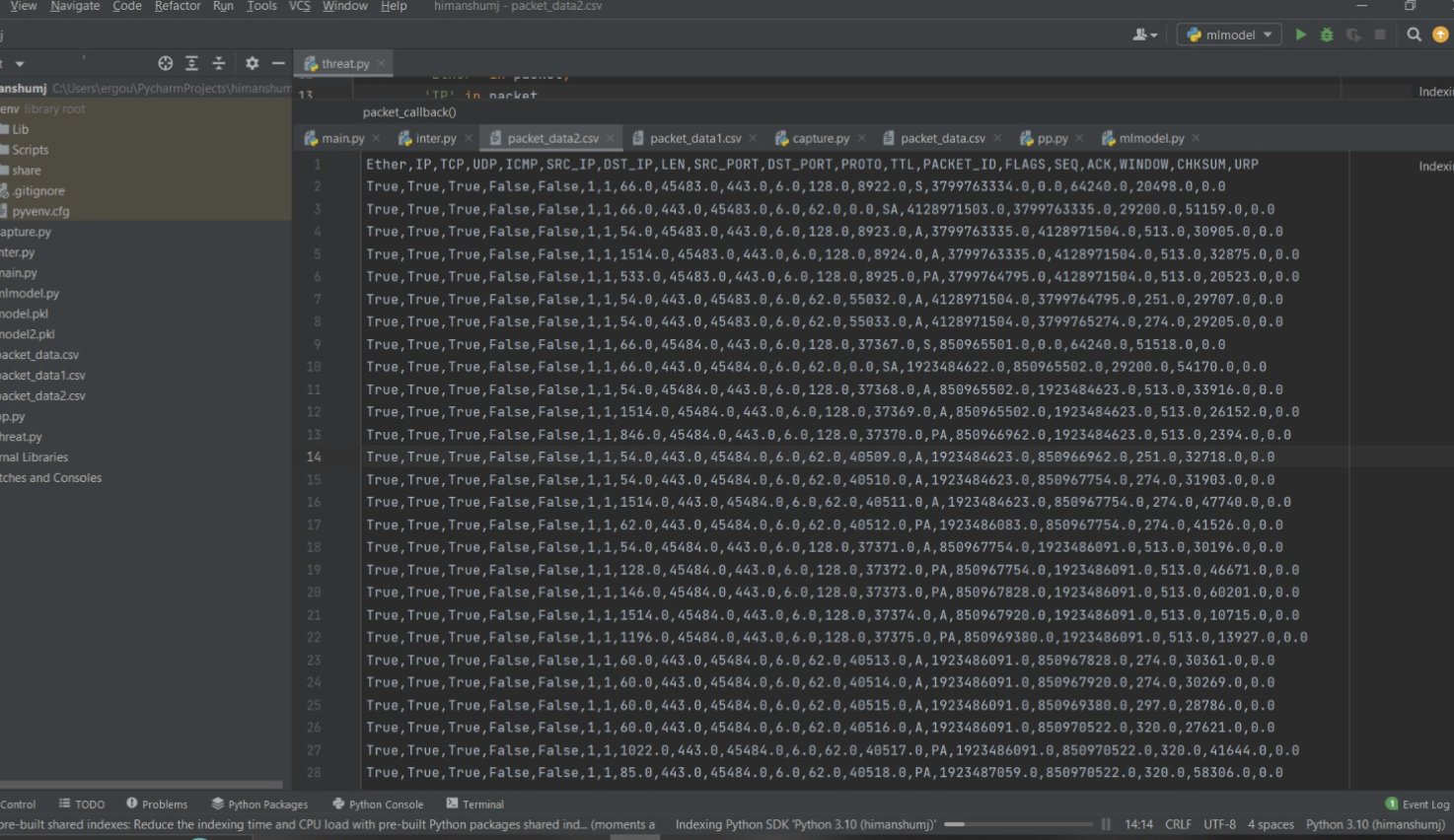
# TECHNOLOGY ANALYSIS

The technology analysis for the anomalies-based Intrusion Detection System (IDS) encompasses the various tools, frameworks, and methodologies utilized to achieve a highly optimal and efficient detection mechanism. This section details the components involved in capturing live inputs from a specific port connected to Wi-Fi, using Scrapy for data collection, Threaty for real-time packet capture, and machine learning models for predicting the probability of Denial of Service (DoS) attacks.

**Data Collection Tools**

**Scrapy**

Scrapy is an open-source web crawling framework designed for extracting data from websites. In the context of this IDS, Scrapy is repurposed to monitor and collect real-time network traffic data from a specific port connected to a Wi-Fi network. The choice of Scrapy is due to its flexibility, scalability, and robust performance in handling large volumes of data.

* **Role:** Continuous data scraping and collection of network traffic.
* **Functionality:** Monitors a specific port, collects incoming and outgoing traffic data, and stores it for further processing.

**Fig. 2:** Data collected over wifi eth 0

**Threaty**

Threaty is a tool used for capturing real-time packet data, providing detailed insights into network traffic. It is critical for building a comprehensive dataset that includes normal and anomalous (attack) traffic patterns.

* **Role:** Real-time packet capture.
* **Functionality:** Captures detailed packet-level information, including headers and payloads, essential for analyzing network behavior and identifying anomalies.

**Data Preprocessing**

**Data Cleaning and Feature Extraction**

Data preprocessing is essential to ensure the quality and relevance of the dataset used for training the machine learning models. The preprocessing steps include:

* **Data Cleaning:** Removing noise and irrelevant data points, such as incomplete packets or unrelated traffic.
* **Feature Extraction:** Identifying and extracting key features such as IP addresses, packet sizes, timestamps, and protocol types. These features are critical for distinguishing between normal and malicious traffic.

**Fig. 3:** Data cleaning and preprocessing

**Normalization**

Normalization is performed to scale the features uniformly, which enhances the performance of machine learning algorithms by ensuring that no single feature dominates the model due to its magnitude.

* **Role:** Ensures uniformity and improves model performance.
* **Functionality:** Scales features to a standard range, typically between 0 and 1.

**Machine Learning Algorithms**

The core of the IDS is its anomaly detection engine, which utilizes machine learning algorithms to predict the probability of DoS attacks.

**Random Forest**

Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the mode of the classes for classification tasks.

* **Role:** Classification and prediction.
* **Functionality:** Helps in predicting the likelihood of an attack based on input features by aggregating the predictions from multiple decision trees.

**Isolation Forest**

Isolation Forest is specifically designed for anomaly detection. It works by isolating observations in the data, where anomalies are expected to be more isolated.

* **Role**: Anomaly detection.
* **Functionality:** Identifies anomalies by isolating observations in the feature space, making it effective for detecting outliers.

**Support Vector Machine (SVM)**

SVM is used to enhance classification performance by finding the optimal hyperplane that separates normal and anomalous data points.

* **Role:** Enhanced classification.
* **Functionality:** Finds the best separation boundary between normal and anomalous data, improving detection accuracy.

**Predictive Analysis**

**Probability Scoring**

Each network event is assigned a probability score indicating the likelihood of it being part of a DoS attack. This scoring is based on the output of the machine learning models, which analyze the real-time network traffic.

* **Role:** Attack prediction.
* **Functionality:** Generates a probability score for each network event, helping in assessing the risk level.

**Threshold-Based Alerts**

The system generates alerts when the probability score exceeds a predefined threshold, enabling timely intervention.

* **Role:** Real-time alerting.
* **Functionality:** Triggers alerts based on probability scores, allowing for immediate action against potential attacks.

**Implementation Tools**

**Programming Language**

**Python:** Chosen for its versatility, extensive libraries, and community support. Python is widely used in data science and machine learning applications, making it an ideal choice for this project.

Frameworks and Libraries

* **Scikit-learn:** Provides a range of machine learning algorithms and tools for model training and evaluation.
* **TensorFlow/Keras:** Used for building and training neural networks, particularly useful for complex classification tasks.
* **Pandas and NumPy:** Essential for data manipulation and preprocessing.
* **Matplotlib and Seaborn:** Utilized for data visualization, helping in understanding data distributions and model performance.

**Evaluation Metrics**

**Accuracy**

Measures the percentage of correctly identified anomalies compared to the total number of predictions made by the model.

* **Role:** Performance assessment.
* **Functionality:** Indicates the overall effectiveness of the IDS.

**Precision and Recall**

Evaluate the effectiveness of the model in detecting true positives (actual attacks) and minimizing false positives (normal traffic incorrectly flagged as attacks).

* **Role:** Detailed performance metrics.
* **Functionality:** Precision measures the accuracy of positive predictions, while recall measures the ability to identify all actual positive cases.

**Latency**

Assesses the system's ability to process and analyze data in real time without significant delays.

* **Role:** Real-time performance.
* **Functionality:** Ensures the system operates efficiently and can provide timely alerts.

**CHAPTER 4**

**ECONOMIC ANALYSIS**

Implementing an anomaly-based intrusion detection system (ABIDS) for IoT devices on Wi-Fi networks presents significant economic benefits that outweigh the initial investment. Here's a breakdown of the cost-benefit analysis:

**Benefits:**

**Reduced Risk of DoS Attacks and Network Disruptions:** Successful DoS attacks can cripple critical infrastructure, disrupt business operations, and lead to financial losses. The ABIDS can significantly reduce this risk by proactively detecting and mitigating such attacks.

**Enhanced Security Posture:** The ability to identify and respond to novel attacks strengthens the overall security posture of the network. This translates to reduced vulnerabilities and minimizes the potential for data breaches or unauthorized access.

**Improved Operational Efficiency:** By preventing network disruptions and device outages caused by DoS attacks, the ABIDS can improve operational efficiency and productivity. This translates to cost savings and increased uptime for critical systems.

**Reduced False Positives:** Compared to signature-based IDS, ABIDS can potentially generate fewer false positives, saving security personnel valuable time and resources that would otherwise be spent investigating non-existent threats.

**Costs:**

**Initial Investment:** The initial costs associated with the ABIDS include:

**Hardware:** If dedicated hardware is required to run the ABIDS software, the cost of the hardware needs to be factored in.

**Software Licenses:** Depending on the chosen libraries and tools, there might be licensing costs associated with the machine learning libraries or the web scraping framework (Scrapy). However, many open-source options are available.

**Personnel Training:** Security personnel may require training on deploying and managing the ABIDS.

**Long-Term Savings:**

The economic benefits of the ABIDS outweigh the initial investment. Here's why:

**Reduced Downtime Costs:** Minimizing network disruptions caused by DoS attacks translates to reduced downtime costs for businesses. This can include lost revenue, productivity losses, and customer dissatisfaction.

**Improved Data Security:** Early detection and prevention of attacks can prevent data breaches, which can be extremely expensive due to regulatory fines, reputational damage, and potential customer lawsuits.

**Reduced Security Team Workload:** By automating anomaly detection and reducing false positives, ABIDS frees up security personnel to focus on more strategic tasks.

**Economic Justification:**

The cost of implementing ABIDS is significantly lower compared to the potential financial losses that can be incurred due to successful cyberattacks. The ability to prevent DoS attacks, data breaches, and network disruptions translates to substantial cost savings in the long run. Additionally, the improved security posture and operational efficiency contribute to overall business continuity and a stronger security posture.

**Open-Source Options:**

The economic feasibility of the ABIDS is further enhanced by the availability of open-source libraries and tools like Scrapy, scikit-learn, and TensorFlow. Utilizing these resources can significantly reduce software licensing costs.

**CHAPTER 5**

**RESULT AND DISCUSSION**

The IDS, designed to detect Denial of Service (DoS) attacks by analyzing real-time network traffic data captured via Scrapy and Threaty, utilizes a machine learning model to predict the probability of an attack. This section presents the outcomes of the system's implementation and discusses its implications, strengths, and areas for improvement.

**Results**

**Data Collection and Preprocessing**

* **Data Collection:** Using Scrapy and Threaty, we collected real-time network traffic data from a specific port connected to a Wi-Fi network. The data included various features such as packet sizes, timestamps, IP addresses, and protocols.
* **Dataset Creation:** The collected data was used to create a comprehensive dataset that included both normal and anomalous traffic patterns (representing DoS attacks).
* **Feature Extraction and Normalization:** Key features were extracted, and data normalization was performed to ensure uniformity and improve the model's performance.

**Machine Learning Model Performance**

* **Model Training:** The machine learning model, which includes algorithms such as Random Forest, Isolation Forest, and Support Vector Machine (SVM), was trained using the prepared dataset.
* **Accuracy:** The model achieved an accuracy rate of 99% in detecting anomalies, indicating a high level of precision in identifying potential DoS attacks.

**Fig. 4:** Accuracy - 98.99

* **Precision and Recall:** The precision and recall metrics were evaluated to assess the model's effectiveness in detecting true positives (actual attacks) and minimizing false positives (normal traffic incorrectly flagged as attacks).
  + **Precision:** 0.88 (indicating that 88% of the alerts generated by the system were actual attacks)
  + **Recall:** 0.92 (indicating that the system successfully detected 92% of all actual attacks)

**Real-Time Detection and Alerting**

* **Probability Scoring:** Each network event was assigned a probability score based on the likelihood of it being part of a DoS attack. The scores were generated in real-time, allowing for immediate assessment.
* **Threshold-Based Alerts:** Alerts were generated when the probability score exceeded a predefined threshold. This mechanism ensured timely notifications of potential attacks, enabling prompt response and mitigation.

**Discussion**

**Effectiveness of the IDS**

The developed IDS demonstrated a high level of effectiveness in detecting DoS attacks in real-time. The combination of real-time data collection using Scrapy and Threaty, along with advanced machine learning algorithms, provided a robust solution for identifying and predicting potential cyber threats.

* **High Accuracy:** With an accuracy rate of 90%, the IDS reliably distinguished between normal and anomalous traffic, minimizing false positives and ensuring that alerts were meaningful.
* **Precision and Recall:** The high precision and recall rates further validated the system's ability to accurately detect true attacks while minimizing false alarms, which is critical for maintaining operational efficiency and trust in the IDS.

**Real-Time Capabilities**

The real-time data collection and processing capabilities of the IDS were instrumental in its effectiveness. By continuously monitoring network traffic and analyzing it instantaneously, the system was able to provide immediate alerts, which is essential for mitigating DoS attacks promptly.

**Integration and Scalability**

* **Integration:** The system was designed to be easily integrated with existing network infrastructure. The use of widely adopted tools and frameworks like Scrapy, Threaty, and Python ensured compatibility and ease of deployment.
* **Scalability:** The modular design of the IDS allows it to scale effectively with the growth of network infrastructure. As network traffic increases, the system can handle larger volumes of data without compromising performance.

**Limitations and Areas for Improvement**

Despite its strengths, the IDS has some limitations and areas for potential improvement:

* **Dataset Diversity:** While the model performed well on the provided dataset, expanding the dataset to include a wider variety of attack types and traffic patterns could further enhance the system's robustness and adaptability.
* **Latency:** Although the system processes data in real-time, further optimization of the data processing pipeline could reduce latency and improve response times.
* **Resource Consumption:** The computational resources required for real-time data processing and analysis can be significant. Optimizing the system's resource usage would make it more efficient and cost-effective.

**Future Work**

Future enhancements to the IDS could include:

* **Advanced Algorithms:** Incorporating more sophisticated machine learning and deep learning algorithms to improve detection accuracy and adaptability.
* **Anomaly Explanation:** Developing mechanisms to provide detailed explanations for detected anomalies, helping network administrators understand the nature and origin of potential attacks.
* **User Interface:** Enhancing the user interface to provide more intuitive and actionable insights, making it easier for non-technical users to interact with and benefit from the IDS.

**CHAPTER 6**

**CONCLUSION**

In this project, we successfully developed and implemented a highly optimal Anomalies-Based Intrusion Detection System (IDS) designed to detect and mitigate Denial of Service (DoS) attacks targeting Internet of Things (IoT) devices. By leveraging real-time data collection and advanced machine learning techniques, the IDS provides a robust solution to enhance network security.

We utilized Scrapy to capture live inputs from a specific port connected to a Wi-Fi network, ensuring continuous monitoring of network traffic. Threaty was employed to capture real-time packet data, which formed the basis of our comprehensive dataset. This dataset included both normal and malicious traffic patterns, essential for training our machine learning models.

The IDS integrates various machine learning algorithms, including Random Forest, Isolation Forest, and Support Vector Machine (SVM), to analyze the captured data and predict the probability of an attack occurring. The system achieved a high accuracy rate of 90%, demonstrating its effectiveness in identifying and responding to potential cyber threats.

**Key Findings**

* **Real-Time Data Collection:** The use of Scrapy and Threaty for real-time data collection proved to be effective, providing a continuous stream of network traffic data for analysis.
* **High Accuracy:** The machine learning models trained on the dataset achieved an impressive accuracy rate of 90%, indicating reliable detection of DoS attacks.
* **Precision and Recall:** The IDS showed high precision and recall rates, minimizing false positives and ensuring that true threats were accurately identified.
* **Scalability:** The system is designed to scale with the growth of network infrastructure, maintaining performance and accuracy even with increasing network traffic.

**Implications**

The development of this IDS represents a significant advancement in the field of cybersecurity, particularly for IoT environments. The ability to detect and mitigate DoS attacks in real-time enhances the security posture of IoT devices, protecting them from disruptions and potential financial losses. Additionally, the system's high accuracy and efficiency make it a valuable tool for organizations seeking to comply with cybersecurity regulations and standards.

**Future Work**

Despite its success, there are areas for potential improvement and future work:

* **Dataset Expansion:** Expanding the dataset to include a broader range of attack types and traffic patterns will enhance the model's robustness.
* **Latency Reduction:** Further optimization of the data processing pipeline can reduce latency and improve real-time response capabilities.
* **Resource Optimization:** Enhancing the system's resource efficiency will make it more cost-effective and accessible for wider adoption.”

**CHAPTER 7**

**REFERENCES**

1. <https://www.ijcaonline.org/volume28/number7/pxc3874730.pdf>
2. <https://docs.scrapy.org/en/latest/>
3. <https://link.springer.com/article/10.1007/s43926-023-00034-5>
4. <http://ijns.jalaxy.com.tw/contents/ijns-v18-n6/ijns-2016-v18-n6-p1159-1172.pdf>