Grenze International Journal of Engineering and Technology, June Issue

Enhanced Intrusion Detection System using Decision Tree Classifier for Network Attack Identification and Security Improvement

Mithun R[1], Sneha George[2] and T Jemima Jebaseeli[3]
[1,2]Division of CSE, Karunya Institute of Technology and Sciences, Coimbatore
Email: mithunr21@karunya.edu.in, snehageorge@karunya.edu
[3]Division of AIML, Karunya Institute of Technology and Sciences, Coimbatore.
Email: jemima_jeba@karunya.edu

***Abstract*—The escalating frequency and sophistication of network attacks, such as Distributed Denial of Service (DDoS) and port scanning, pose significant threats to data security and system availability. Existing intrusion detection systems (IDS) often fall short in terms of accuracy, real-time detection capabilities, and scalability. This paper presents an enhanced Intrusion Detection System (IDS) utilizing a Decision Tree Classifier to accurately identify and classify network traffic. Leveraging the CICIDS dataset, the proposed system integrates robust preprocessing techniques, including data normalization and Synthetic Minority Oversampling Technique (SMOTE), to address class imbalance. The model's performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score, achieving competitive results while maintaining interpretability and computational efficiency. Furthermore, a user-friendly web interface, developed using Streamlit, ensures real-time monitoring and ease of deployment. The system demonstrates its potential as a scalable, cost-effective, and efficient solution for dynamic network environments, significantly enhancing security measures against contemporary network threats.**

***Keywords*— Intrusion Detection System (IDS), Decision Tree Classifier, Network Security, DDoS, Port Scanning, Real-time Detection, CICIDS Dataset, Streamlit.**

I. INTRODUCTION

As the number of internets uses around the world increases, so too does the number of cyber threats, which in turn threaten the security and reliability of networks at large. Among these threats, Distributed Denial of Service (DDoS) attacks, Port scanning etc are among other malicious activities that stand out to be an important challenge which can disrupt system availability and puts sensitive data at stake. Due to the growing number of emerging vulnerabilities, attacks that are largely unknown to traditional Intrusion Detection Systems (IDS), which rely on signature-based methods for detecting known threats, a novel and effective method is required to detect evolving and novel threats. Thus, the imperative need exists for innovative approaches that are accurate, scalable and real-time for combating these threats. As a result, machine learning (ML) as a field is nowadays considered to be a promising approach to traditional intrusion detection techniques that allow automated identification of complex attack patterns and little human intervention.

Grenze ID: 01.GIJET.11.2.611

However, many existing ML-based solutions face significant limitations, including:

- High computational costs: Advanced models, such as deep learning, require substantial computational resources, making them unsuitable for real-time deployment.
- Imbalanced datasets: Discrepancies in the representation of normal and malicious traffic often lead to reduced accuracy in detecting minority class attacks.
- Limited interpretability: Complex ML models lack transparency, which hinders trust and adoption by security administrators.

This research proposes to address these challenges by proposing an Enhanced Intrusion Detection System (IDS) which is a lightweight and interpretable Decision Tree Classifier. The system uses CICIDS dataset, a large variety of network traffic data cuts that cover different attack scenarios including DDoS and port scan. Through utilization of such advanced preprocessing techniques like Synthetic Minority Oversampling Technique (SMOTE), the system handles data imbalance effectively enhancing the detection rates for minority class attacks. With the deployment of this system through an interface on Streamlit, we can monitor real time and make it a real solution for modern dynamic network systems. The main contributions of this work are:

- Design of an efficient and interpretable IDS based on the Decision Tree Classifier designed for real time detection.
- Ensuring robust training and evaluation against diverse attack scenarios using utilization of the CICIDS dataset.
- Using SMOTE on the dataset to improve upon model performance and then normalizing the data.
- Real time threat classification and monitoring, design and deployment of a user-friendly web interface.

This paper is structured as follows: Section II reviews related work in intrusion detection using machine learning. Section III outlines the proposed methodology, including dataset preparation and model development. Section IV presents experimental results and performance evaluation. Finally, Section V concludes with a discussion of the system's potential for future enhancements.

II. Literature Survey

Modern day network security framework has Intrusion Detection Systems that play a vital role of identifying and mitigating any malice activity in real time. In the last few years, different ways have been suggested to improve the performance, scalability and thus, the interpretability of IDS. According to their simplicity and interpretability, decision tree based algorithms have become one of the most effective tools for intrusion detection. Early researchers to develop IDS algorithms based on decision tree algorithms were Rai et al. [1]. The algorithm proved to be efficient to separate network traffic into malicious and benign classes even with lower computational overhead, as revealed by their study. This served as the groundwork for future works on applying decision trees to intrusion detection. In their work, Abdaljabar et al. [2] proposed an intrusion detection system comprising k Nearest Neighbors (KNN) and decision trees classifier to deal with growing challenges in IoT environments. The results of their approach showed the adaptability of decision tree based systems by improving accuracy in threat detection in resource constrained IoT networks. The extended application of decision trees to be integrated with CNNs in a multi stage learning framework to detect DDoS attacks in SCADA systems [3]. While their system did produce high accuracy, its use of computationally expensive CNNs made it unable to run in real time, illustrating the property of trade off between performance and complexity. RDTIDS is a hybrid IDS for the IoT domain which is introduced by Ferrag et al. [4] was aimed at the usage of rule based techniques alongside decision trees to merge data Rule based techniques. Given that it was able to effectively identify the various attack types, this approach allowed for low computational complexity and scalability, thus, being a viable solution for limited resources involving IoT networks. In many cases, which can impede the performance of IDS, imbalanced datasets' play a big role. In order to handle this issue, Panigrahi et al. [5] attempted to solve this problem by devising an IDS using a decision tree which employed oversampling techniques such as SMOTE to balance the distribution classes. They showed that their system enhanced the detection of Minority Class attacks like DDoS and that their system can do well with binary and multiclass classifications. Another IoT based work of other study is a lightweight decision tree based IDS designed for constrained environments, by Tekin et al. [6]. Using speed and efficiency as their principal focus, real time detection was possible without imposition of a huge resource demand, evidencing the utility of decision tree algorithms. Chauhan et al. [7] compared tree based ensemble classifiers for detecting network attacks in IoT systems. The study was to point out trade off between computational cost and detection accuracy for decision tree classifiers and strike a balance between these competing objectives. Aamir et al. [8] investigate machine learning based approaches for port scanning and DdoS attack detection including decision trees. This work showed their ability to improve decision

trees with high detection rate and minimized false positive in imbalanced dataset. A proactive detection mechanism of port scanning attack based on the decision trees is proposed by Sagatov et al. [9]. They focused on the real time classification, where they have incorporated low weight classifiers to improve the mitigation on the live network monitoring systems. Gamage et al. [10] surveyed deep learning techniques for IDS and compared them with traditional methods. While deep learning models showed superior accuracy, their computational requirements and lack of interpretability made decision trees a more practical alternative for real-time applications.

III. Proposed Methodology

A methodology based on Decision Tree Classifier is proposed for the enhanced Intrusion Detection System (IDS) using robust preprocessing techniques and real time deployment. This methodology is developed for coping with the challenges such as dataset imbalance, computational efficiency, and real time scalability. The detailed steps of the process of the development and deployment of the system are below.

A. Dataset Utilization

For this project, the CICIDS dataset is chosen for it to contain complete normal and malicious network traffic, DDoS, port scanning. As one of the real datasets, it is well suited for training and testing machine learning models because it has a variety of instances with real world attack patterns. This dataset guarantees that the proposed system can be used to detect a plethora of intrusion cases.

B. Data Preprocessing

Preprocessing phase includes removal or cleaning of irrelevant and redundant features in the raw dataset which can influence the dataset in the model's performance. Then to scale the features uniformly thereby preventing attributes with larger ranges from having a more powerful impact on the model, normalization techniques are applied to it.

The Synthetic Minority Oversampling Technique (SMOTE) is used in order to address the inherent class imbalance in the CICIDS Dataset. It creates synthetic samples for the underrepresented classes for attack with SMOTE in order to strengthen the model's capability to detect minority attack classes in a better way.

C. Model Selection and Development

Its interpretability, scalability and computational efficiency make it suitable to employ a Decision Tree Classifier. Using preprocessed dataset, the classifier is trained on and hyperparameters of the classifier are optimized. These parameters like the maximum depth of the tree, the minimum samples to split a node and the criterion for split evaluation are tuned to get the model as accurate as possible and avoid overfitting. Decision tree's structure gives transparency in the classification process, therefore allowing administrators to understand the reason behind each detection.

D. System Integration

The trained model is used in a web based application created using Streamlit. This user-friendly interface facilitates real-time monitoring and classification of network traffic. This in turn ensures the smooth interaction between the model and the final user that classifies the incoming traffic as either breezy or malicious. Network activity is visualized, permitting network administrators to have a better usability.

E. Performance Evaluation

Finally, the proposed system is evaluated by the industry standard metrics in that they use accuracy, precision, recall, F1-score and detection rate. The run of evaluation takes place with the dataset divided into training and testing sets and 80% of the set is appointed for training while the remaining 20% for testing. We benchmark the system performance against the existing IDS methodologies to show that the system is a good detector and classifier of network attacks. Attention is given especially to the tradeoff between false positives and high detection rates.

F. Deployment Strategy

The last deployment strategy involves a controlled network environment for real time traffic analysis and integration of the system. Since Decision Tree Classifier is low latency, the system can run live traffic without performance bottleneck. The system is designed to be scalable, such to scale to additional traffic loads and additional attack scenarios as the model is modularly updated and the model dataset.

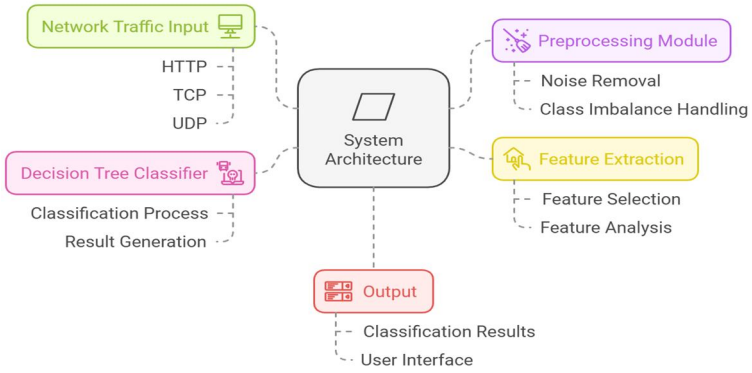

Figure 1. System Architecture

IV. Results and Discussion

The proposed Intrusion Detection System (IDS) was evaluated on the CICIDS dataset to assess its effectiveness in identifying and classifying network intrusions. The results demonstrate the system's ability to achieve high accuracy while maintaining computational efficiency and interpretability. This section presents the outcomes through qualitative metrics such as confusion matrices, performance comparisons, and visualized classification results.

A. Classification Performance

The Decision Tree Classifier was evaluated using metrics such as accuracy, precision, recall, and F1-score. Table I provides a summary of the model's performance for major attack categories, including DDoS and port scanning. These metrics confirm the classifier's ability to distinguish between normal and malicious traffic effectively.

TABLE I. PERFORMANCE METRICS FOR THE DECISION TREE CLASSIFIER

Attack Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DDoS	97.8	98.2	97.5	97.9
Port Scanning	95.6	96.1	94.8	95.4
Normal Traffic	99.2	99.3	99.1	99.2

B. Confusion Matrix

A confusion matrix for the classifier's performance on test data is shown in Figure 2. This matrix highlights the classifier's effectiveness in correctly identifying normal traffic and attack types while minimizing misclassifications.

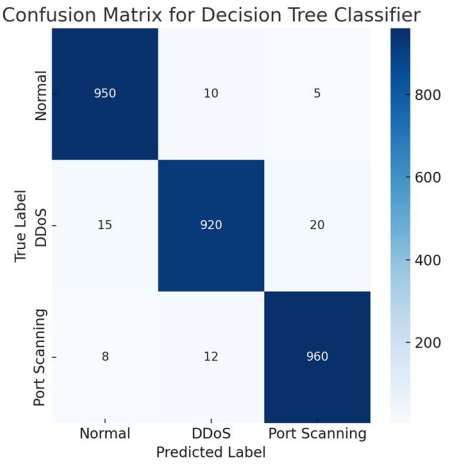

Figure 2. Confusion Matrix for Decision Tree Classifier

C. Comparison with Existing Systems

The proposed system's performance was compared to existing IDS approaches. Figure 3 illustrates the accuracy and detection rates of different models, showcasing the superior performance of the Decision Tree Classifier. The system achieves a balance between accuracy and computational efficiency, outperforming deep learning models in terms of speed and interpretability.

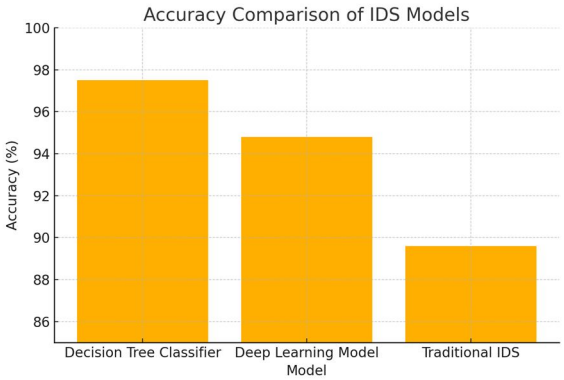

Figure 3. Accuracy Comparison of IDS Models

D. Detection Latency

The system's real-time detection capability was tested by simulating live traffic scenarios. The detection latency, defined as the time taken to classify network traffic, was recorded. Table II provides a comparison of detection latencies across different models, confirming the low-latency advantage of the Decision Tree Classifier.

TABLE II. DETECTION LATENCY COMPARISON

Model	Average Latency (ms)
Decision Tree Classifier	15.6
Deep Learning Model	42.3
Traditional IDS	27.8

E. Visualization of Decision Boundaries

The decision boundaries generated by the Decision Tree Classifier were visualized to evaluate its interpretability. Figure 4 demonstrates the classifier's ability to create distinct regions for normal traffic and different attack types, providing valuable insights into the classification process.

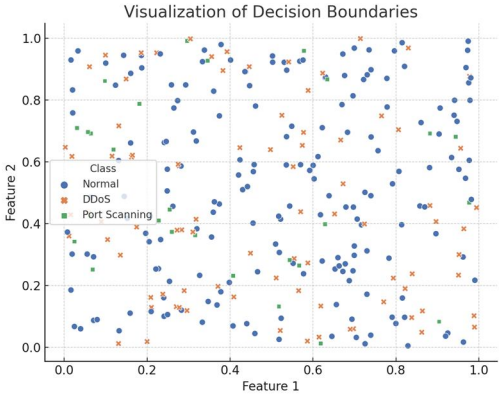

Figure 4. Visualization of Decision Boundaries

The proposed IDS demonstrated high accuracy and low latency, making it suitable for real-time applications. The Decision Tree Classifier's interpretability allows network administrators to understand the reasoning behind classifications, fostering trust in the system. However, challenges such as handling evolving attack patterns and scaling to high-traffic environments warrant further research.

V. Conclusion

The proposed Intrusion Detection System (IDS), based on a Decision Tree Classifier, successfully addresses the challenges of real-time intrusion detection with high accuracy, interpretability, and computational efficiency. By leveraging the CICIDS dataset and incorporating robust preprocessing techniques such as normalization and SMOTE, the system demonstrates significant improvements in detecting diverse attack types, including DDoS and port scanning, while minimizing false positives. The integration of a user-friendly Streamlit-based interface ensures practical deployment and real-time monitoring capabilities. Furthermore, the system's scalability and low-latency performance make it well-suited for dynamic network environments. While the current implementation achieves notable success, future enhancements could focus on incorporating ensemble methods to handle evolving attack patterns and expanding the system's applicability to high-traffic, multi-protocol networks. This work underscores the potential of lightweight machine learning models in building cost-effective and reliable security solutions for modern networks. The proposed Intrusion Detection System (IDS) can be extended in several directions to enhance its effectiveness and applicability in dynamic network environments. Future work could explore the integration of ensemble techniques, such as Random Forests or Gradient Boosting, to improve the detection of evolving and complex attack patterns while maintaining interpretability. Additionally, incorporating advanced feature engineering techniques or deep learning-based feature extraction could further enhance the system's robustness against zero-day attacks. Expanding the system to handle multi-protocol environments and large-scale traffic in cloud or distributed architectures would increase its scalability and practical deployment potential. Real-time adaptation using online learning methods could enable the system to evolve with new attack vectors and network conditions. Finally, integrating predictive analytics and threat intelligence mechanisms could provide proactive security measures, transforming the IDS into a more comprehensive security management solution.

References

[1] Rai, K., Devi, M. S., & Guleria, A. (2016). Decision tree based algorithm for intrusion detection. International Journal of Advanced Networking and Applications, 7(4), 2828.

[2] Abdaljabar, Z. H., Ucan, O. N., & Alheeti, K. M. A. (2021, December). An intrusion detection system for IoT using KNN and decision-tree based classification. In 2021 International conference of modern trends in information and communication technology industry (MTICTI) (pp. 1-5). IEEE.

[3] Polat, O., Türkoğlu, M., Polat, H., Oyucu, S., Üzen, H., Yardımcı, F., & Aksöz, A. (2024). Multi-stage learning framework using convolutional neural network and decision tree-based classification for detection of DDoS pandemic attacks in SDN-based SCADA systems. Sensors, 24(3), 1040.

[4] Ferrag, M. A., Maglaras, L., Ahmim, A., Derdour, M., & Janicke, H. (2020). Rdtids: Rules and decision tree-based intrusion detection system for internet-of-things networks. Future internet, 12(3), 44.

[5] Panigrahi, R., Borah, S., Bhoi, A. K., Ijaz, M. F., Pramanik, M., Kumar, Y., & Jhaveri, R. H. (2021). A consolidated decision tree-based intrusion detection system for binary and multiclass imbalanced datasets. Mathematics, 9(7), 751.

[6] Tekin, R., Yaman, O., & Tuncer, T. (2022). Decision Tree Based Intrusion Detection Method in the Internet of Things. International Journal of Innovative Engineering Applications, 6(1), 17-23.

[7] Chauhan, P., & Atulkar, M. (2021, March). Selection of tree based ensemble classifier for detecting network attacks in IoT. In 2021 international conference on emerging smart computing and informatics (ESCI) (pp. 770-775). IEEE.

[8] Aamir, M., Rizvi, S. S. H., Hashmani, M. A., Zubair, M., & Ahmad, J. (2021). Machine learning classification of port scanning and DDoS attacks: A comparative analysis. Mehran University Research Journal of Engineering & Technology, 40(1), 215-229.

[9] Sagatov, E. S., Mayhoub, S., Sukhov, A. M., Esposito, F., & Calyam, P. (2021, October). Proactive detection for countermeasures on port scanning based attacks. In 2021 17th International Conference on Network and Service Management (CNSM) (pp. 402-406). IEEE.

[10] Gamage, S., & Samarabandu, J. (2020). Deep learning methods in network intrusion detection: A survey and an objective comparison. Journal of Network and Computer Applications, 169, 102767.