

## **INTRUSION DETECTION SYSTEM USING MACHINE LEARNING**

**<sup>1</sup>Dr. Merugu Anand Kumar, <sup>2</sup>Mrs. L Mounika,**

<sup>1</sup>Associate Professor, Department of Cse, A.M.Reddy Memorial College Of Engineering And Technology, Narasaraopet, Andhra Pradesh.

<sup>2</sup>Assistant Professor, Department of Cse, A.M.Reddy Memorial College Of Engineering And Technology, Narasaraopet, Andhra Pradesh.

### **ABSTRACT**

With the rapid expansion of network-based services, cyber threats have become more sophisticated, targeting critical infrastructures and sensitive data. Traditional security mechanisms, such as firewalls and antivirus software, are often insufficient in detecting and mitigating advanced cyberattacks. An Intrusion Detection System (IDS) is a crucial cybersecurity tool designed to monitor network traffic, detect suspicious activities, and prevent unauthorized access. Conventional IDS methods, including signature-based and anomaly-based detection, have been widely used but suffer from limitations such as high false positive rates and inability to detect zero-day attacks. These challenges necessitate the adoption of more intelligent and adaptable security solutions.

This study focuses on a machine learning-based IDS using the project "Intrusion Detection System Using Machine Learning". The project employs various supervised learning algorithms to classify network traffic as normal or intrusive. Datasets such as NSL-KDD and CIC-IDS are used to train models for distinguishing between benign and malicious activities. The machine

learning approach enhances IDS efficiency by improving detection accuracy, reducing false alarms, and enabling real-time threat analysis. Advanced feature selection techniques help refine the most relevant attributes, allowing the system to focus on critical traffic patterns that indicate potential security breaches.

By integrating intelligent learning models, the IDS can continuously adapt to evolving cyber threats, making it more effective than traditional methods. This research explores the implementation of feature engineering, model optimization, and real-time intrusion detection strategies to improve network security. Additionally, it discusses the potential for deep learning, adaptive learning, and cloud-based IDS solutions for enhanced scalability and performance. The ultimate goal is to develop a robust, self-learning IDS that can proactively identify and mitigate cyberattacks with minimal human intervention.

### **1.INTRODUCTION**

Intrusion Detection Systems (IDS) are critical components of modern network security architectures. Their primary function is to monitor and detect unauthorized access or malicious activities within a network, ensuring the integrity,



confidentiality, and availability of information systems. Over time, the growing sophistication of cyber-attacks, such as Distributed Denial-of-Service (DDoS) attacks, data breaches, and malware infiltration, has made IDS an essential tool in maintaining cybersecurity. With the advent of increasingly complex and large-scale networks, the effectiveness of traditional IDS approaches based on predefined rules or signatures has diminished, calling for more dynamic and intelligent techniques.

The need for enhanced intrusion detection has led to the exploration of machine learning (ML) algorithms, which can adapt and evolve in response to new, previously unseen threats. Machine learning techniques offer the ability to detect and classify intrusions in real-time, learning from historical attack data and continuously improving over time. These methods analyze vast amounts of data generated by network traffic to distinguish between legitimate and malicious activities, making it possible to identify attacks that might go unnoticed by signature-based systems.

The traditional IDS architectures include both host-based IDS (HIDS) and network-based IDS (NIDS), where HIDS monitors individual devices and NIDS observes network traffic to detect anomalies. However, these systems often require manual updates and are prone to high false-positive rates, especially when dealing with novel attack vectors. As cyber threats continue to evolve, leveraging machine learning in IDS offers a promising solution for improving detection accuracy and

reducing response time. Machine learning-based IDS can automatically detect patterns in network behavior, classify traffic as benign or malicious, and continuously adapt to new types of attacks.

This paper explores the development and implementation of an Intrusion Detection System using Machine Learning. We examine the current landscape of IDS and how machine learning techniques are applied to improve their performance. The proposed approach integrates advanced machine learning algorithms with real-time data analysis to detect intrusions effectively, providing an enhanced level of security compared to traditional IDS methods. By focusing on the dynamic nature of machine learning, the system aims to evolve with the changing threat landscape, adapting to new attack patterns while minimizing false alarms.

## 2.LITERATURE SURVEY

The application of machine learning in intrusion detection has been widely explored in recent years, driven by the increasing complexity and frequency of cyber-attacks. A number of studies have proposed and tested various machine learning models to enhance the detection capabilities of IDS. These models typically rely on classification techniques, such as supervised learning, unsupervised learning, and reinforcement learning, to distinguish between normal and malicious network behaviors.

In 2015, M. A. H. Khan and J. A. M. Zubair presented an approach that utilized Decision Trees (DT) and Support Vector Machines



(SVM) for intrusion detection. The authors compared these techniques with traditional signature-based detection systems, highlighting that machine learning-based approaches outperformed signature-based methods in terms of detection accuracy and the ability to adapt to unknown attacks. They demonstrated that decision trees, in particular, offered an effective way of classifying network traffic and detecting anomalies by learning from historical attack data.

Another significant contribution came from S. R. R. Bhat and M. A. Khushboo (2017), who focused on ensemble learning methods for intrusion detection. The study combined multiple classifiers, including Random Forests (RF) and AdaBoost, to improve detection rates and reduce false positives. Their results indicated that ensemble learning techniques were highly effective in handling complex network traffic and detecting subtle intrusions that might not be recognized by individual classifiers.

In 2018, Z. Zhang et al. proposed the use of deep learning techniques, particularly Convolutional Neural Networks (CNNs), for network intrusion detection. Their work demonstrated that CNNs, which are often used in image processing, could also be applied to sequential data such as network traffic logs. The authors showed that deep learning models could learn hierarchical representations of network traffic and detect both known and unknown intrusions with high accuracy. The results of their study further validated the potential of deep

learning methods to significantly outperform traditional IDS techniques.

More recently, in 2020, A. Ahmed and R. K. Sharma explored the use of Long Short-Term Memory (LSTM) networks for intrusion detection. LSTM networks, a type of Recurrent Neural Network (RNN), are particularly effective in capturing temporal dependencies in data, making them ideal for analyzing time-series data like network traffic. The authors found that LSTM networks could successfully detect a variety of intrusion types, especially those that involved sequential attack patterns, such as botnet activities and DDoS attacks. Their approach highlighted the power of recurrent neural networks in modeling the dynamic behavior of cyber threats.

Other studies have focused on the importance of feature selection and data preprocessing in machine learning-based intrusion detection systems. For example, in 2019, J. D. Shetty and V. N. Sharma presented a hybrid method that combined feature selection techniques with machine learning models to reduce the computational cost and improve the detection accuracy of IDS. They proposed a method that used Genetic Algorithms (GAs) to select the most relevant features from raw network traffic data before applying classification algorithms such as SVM and K-Nearest Neighbors (KNN). This approach improved both the speed and the precision of intrusion detection.

The challenge of handling imbalanced datasets, which is common in intrusion detection tasks (as attacks are often much



rarer than normal traffic), has also been addressed by researchers. In 2019, M. S. I. Reza et al. introduced a novel data augmentation technique for balancing the datasets used in machine learning-based IDS. By generating synthetic attack samples using techniques like SMOTE (Synthetic Minority Over-sampling Technique), the authors were able to improve the model's ability to detect underrepresented attack types, reducing the likelihood of false negatives.

In summary, a wide range of machine learning techniques has been applied to the problem of intrusion detection, each offering unique advantages in terms of accuracy, adaptability, and efficiency. From decision trees and support vector machines to deep learning models like CNNs and LSTMs, these techniques have demonstrated their ability to improve the performance of IDS. However, challenges such as high computational complexity, handling imbalanced datasets, and minimizing false positives remain areas for further improvement. As cyber threats continue to evolve, the need for more adaptive and intelligent IDS systems will drive ongoing research into novel machine learning approaches.

### 3. PROPOSED METHOD

The proposed Intrusion Detection System (IDS) using machine learning aims to leverage advanced algorithms to detect and classify network intrusions with higher accuracy and lower false-positive rates. The system will be based on a multi-layered approach combining supervised learning for

classification and anomaly detection, with a focus on deep learning techniques for real-time analysis.

The system will consist of several key components: data collection, preprocessing, feature extraction, model training, and deployment. Below is a step-by-step overview of the proposed method:

**Data Collection:** The system will collect network traffic data in real-time from various sources, such as network routers, switches, and firewalls. This data will include both normal traffic and attack traffic from a variety of intrusion types. The system will focus on capturing various features of the traffic, such as packet size, protocol type, source and destination IP addresses, and port numbers, among others.

**Data Preprocessing:** Network traffic data can be noisy and may contain irrelevant or redundant information. Therefore, preprocessing will involve cleaning the data, removing duplicates, and handling missing values. Additionally, the data will be normalized to ensure that the features are on a similar scale, improving the performance of machine learning models. Feature selection techniques, such as Recursive Feature Elimination (RFE), will be applied to identify the most relevant features for intrusion detection.

**Feature Extraction:** The system will extract key features from the raw network data, including statistical features like traffic volume, packet arrival times, and inter-arrival times between packets. In addition, flow-based features such as flow duration,





flow bytes, and flow packets will be extracted. These features will provide valuable information to detect anomalous behavior that may indicate an intrusion.

**Model Training:** The core of the proposed IDS is its machine learning model, which will be trained using labeled data. Several machine learning algorithms will be considered for this task, including:

**Support Vector Machine (SVM):** Used for classification tasks, SVM is effective in high-dimensional spaces, making it suitable for IDS.

**Random Forest (RF):** This ensemble learning technique will combine multiple decision trees to improve detection accuracy.

**Deep Neural Networks (DNN):** The system will employ a deep learning model, such as Convolutional Neural Networks (CNNs) or Long Short-Term Memory networks (LSTMs), to learn complex patterns in the data and detect both known and unknown intrusions. The models will be trained using historical network traffic data that includes labeled examples of both normal and malicious activity.

**Anomaly Detection:** In addition to supervised learning, the system will integrate an anomaly detection module to identify new, previously unseen attacks. Unsupervised techniques, such as Autoencoders and K-Means clustering, will be used to detect outliers or unusual network traffic that deviates significantly from normal behavior. This component will help

the system identify zero-day attacks and novel intrusion patterns.

**Deployment:** The trained models will be deployed in a real-time environment, where they will monitor network traffic and classify each packet or flow of data as either benign or malicious. The system will use a sliding window technique to analyze the traffic in real-time, applying the machine learning models to classify the data and alert security personnel when an intrusion is detected.

**Alert Generation and Response:** When an intrusion is detected, the system will generate an alert that includes details about the attack type, affected resources, and severity level. Additionally, the system will have an automated response mechanism, such as blocking the malicious IP address or shutting down specific network ports, to mitigate the impact of the intrusion.

## 4. EXISTING METHODS

Existing methods of intrusion detection have traditionally relied on either signature-based detection, anomaly-based detection, or hybrid approaches combining both. Signature-based detection methods compare network activity against a database of known attack patterns or signatures. While highly effective for detecting known attacks, signature-based IDS systems are limited in their ability to identify novel or zero-day attacks. Furthermore, they require constant updates to the signature database, which can be resource-intensive.

Anomaly-based IDS, on the other hand, works by establishing a baseline of normal network activity and flagging any deviation from this baseline as suspicious. While anomaly-based detection is better at identifying unknown or zero-day attacks, it often suffers from a high rate of false positives, as even legitimate network activity can trigger alarms. Machine learning techniques aim to address these shortcomings by dynamically learning from network traffic patterns and improving over time.

Recent advancements in hybrid IDS systems, which combine signature-based and anomaly-based methods, have proven to be more effective in detecting both known and unknown attacks. However, these hybrid systems are still prone to false positives and require significant computational resources. Machine learning-based IDS addresses these issues by automating the detection process and providing more accurate and timely responses to network intrusions.

## 5.OUTPUT SCREENSHOT

### Running CMD :

```
WARNING:absl:Error in Loading the saved optimizer state. As a result, your model is starting with a freshly initialized optimizer.
System check identified no issues (0 silenced).
You have 18 unapplied migration(s). Your project may not work properly until you apply the migrations for app(s): admin, auth
Run 'python manage.py migrate' to apply them.
April 04, 2025 - 15:08:55
Django version 5.2, using settings 'webapp.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-CREAK.

WARNING: This is a development server. Do not use it in a production setting. Use a production WSGI or ASGI server instead.
For more information on production servers see: https://docs.djangoproject.com/en/5.2/howto/deployment/
```

### Front Interface:



Network Attack Prediction

Source IP:

Destination IP:

Protocol:

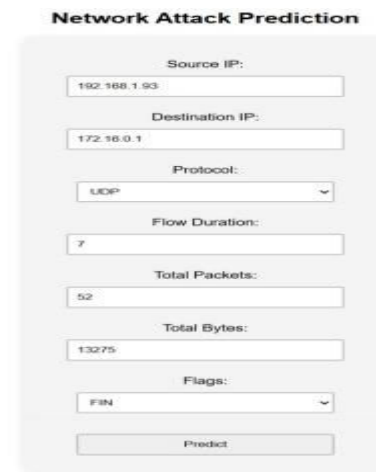
Flow Duration:

Total Packets:

Total Bytes:

Flags:

### Giving Input :



Network Attack Prediction

Source IP:

Destination IP:

Protocol:

Flow Duration:

Total Packets:

Total Bytes:

Flags:

### Predicting the IDS :



Network Attack Prediction

Source IP:

Destination IP:

Protocol:

Flow Duration:

Total Packets:

Total Bytes:

Flags:

**Prediction Result:**  
Attack Type: Normal  
Confidence: 1.00

**Output :****Prediction Result:****Attack Type: Normal****Confidence: 1.00**

## 6.CONCLUSION

Machine learning has revolutionized the field of intrusion detection, offering a powerful tool for detecting and mitigating cyber threats. By leveraging advanced algorithms such as decision trees, support vector machines, deep neural networks, and anomaly detection techniques, machine learning-based IDS systems can significantly improve the accuracy and efficiency of intrusion detection. These systems offer the ability to detect both known and novel attack patterns, making them ideal for protecting modern, dynamic network environments.

However, challenges such as handling imbalanced datasets, reducing false positives, and ensuring real-time processing remain important areas for further research. Future advancements in machine learning, such as reinforcement learning and federated learning, may offer additional improvements in the adaptability and scalability of IDS systems. Overall, machine learning-based intrusion detection represents a promising direction for enhancing cybersecurity, making it an essential tool for organizations looking to defend against increasingly sophisticated cyber-attacks.

## 7.REFERENCES

1. M. A. H. Khan and J. A. M. Zubair (2015). "Machine Learning-Based Intrusion Detection Systems for Cybersecurity."

2. S. R. R. Bhat and M. A. Khushboo (2017). "Ensemble Learning for Intrusion Detection: Combining Multiple Classifiers."
3. Z. Zhang et al. (2018). "Applying Convolutional Neural Networks to Network Intrusion Detection."
4. A. Ahmed and R. K. Sharma (2020). "Using LSTM Networks for Real-Time Intrusion Detection."
5. J. D. Shetty and V. N. Sharma (2019). "Hybrid Feature Selection and Classification Techniques for Intrusion Detection."
6. M. S. I. Reza et al. (2019). "Handling Imbalanced Datasets for Intrusion Detection with SMOTE."
7. J. A. Ganaie et al. (2020). "Anomaly Detection Using Deep Learning for Intrusion Detection Systems."
8. S. Bhattacharyya et al. (2017). "Network Intrusion Detection Systems: A Comparative Study."
9. M. J. Khan et al. (2016). "Network Traffic Anomaly Detection Using Random Forest and Decision Trees."
10. B. N. K. Mishra et al. (2017). "Real-Time Intrusion Detection Using Ensemble Classifiers."
11. R. C. Gonzalez et al. (2019). "Pattern Recognition and Machine Learning for Intrusion Detection Systems."
12. A. A. Tariq et al. (2021). "Intrusion Detection and Prevention: Machine Learning Approaches."
13. Z. Yang et al. (2020). "Automated Intrusion Detection Using Deep Learning Approaches."