Machine Learning Driven Cyber Intrusion Detection and Visualization System Using NSL-KDD Dataset

Mr. Laxman Rajendra Ganpule.
Department of Information Technology, Sindhudurg Sub Campus, Sawantwadi

***Abstract*—The rapid expansion of digital communication and services over the Internet has led to a notable rise in the frequency and sophistication of cyberattacks. Traditional Intrusion Detection Systems (IDS) frequently fail to detect modern sophisticated threats due to their reliance on static signatures and limited adaptability. This study proposes a Machine Learning Driven Cyber Intrusion Detection and Visualisation System using NSL-KDD dataset for intelligent network attack detection and real-time monitoring. The proposed system integrates supervised machine learning algorithms (Decision Tree and Random Forest) with a real-time detection engine and an interactive visualisation dashboard developed using Flask and Chart.js. The system performs data pre-processing, encoding of categorical features, attack classification, confidence aware prediction, live logging and service level attack analysis. Intrusions are classified into various attack classes such as Denial of Service (DoS), Probe, Remote-to-Local (R2L), User-to-Root (U2R), Unknown and Normal traffic. The experimental results show that the Random Forest algorithm is better than Decision Tree in detection performance. The dashboard provides real-time monitoring, graphical visualisation, and threat analytics to enhance cybersecurity operations.**

I. INTRODUCTION

The rapid growth of digital technologies including cloud computing, e-commerce platforms, online banking and IoT devices makes cybersecurity one of the most critical challenges in modern computing environments. Organisations are always under threat from Denial of Service (DoS) attacks, unauthorised access attempts, malware infection and network intrusions. Such attacks can result in data breaches, financial damages, loss of system availability and privacy violations.

IDS (Intrusion Detection Systems) are an important security mechanism to monitor network traffic and detect malicious activities. Two major approaches to conventional IDS systems are signature-based and anomaly-based. Signature-based systems look for known attacks using predefined rules and anomaly-based systems look for abnormal network activity. However, conventional IDS solutions are often unable to detect unknown or evolving attacks and generate high false-positive rates.

Machine Learning (ML) is a promising approach for intelligent intrusion detection, where systems can learn patterns from network traffic data and automatically detect malicious activities. ML algorithms can analyse large datasets and classify network traffic more efficiently and adaptably. The integration of real-time visualisation dashboards with machine learning further enhances the monitoring and analysis capabilities of IDS systems.

This research focuses on the design and implementation of a Machine Learning Driven Cyber Intrusion Detection and Visualisation System using NSL-KDD dataset. The proposed system incorporates machine learning based classification and real time visualisation dashboard for continuous attack monitoring and threat analysis.

II. PROBLEM STATEMENT

Some limitations of traditional intrusion detection systems are as follows:

1. Ignorance of modern and evolving cyber-attacks.
2. High false alarm rate.
3. No intelligence learning capability.
4. No real-time monitoring and visualisation.
5. Doesn't have service-level attack analytics.
6. Bad handling of uncertain or low-confidence predictions.

Many current implementations of machine learning-based IDS are only in classification
accuracy without live monitoring, visualisation, or confidence-aware detection.
Therefore, an intelligent IDS framework capable of real-time detection is needed,
multi-class attack classification, visualisation and analytics of threats.

III. METHODOLOGY

3.1 Proposed System Overview

The proposed system is an intrusion detection system based on machine learning techniques, which detect cyber-attacks and visualize attack data through real-time dashboards. This system comprises modules for data pre-processing, training, intrusion detection, log generation, and visualization.

3.2 System Architecture

The workflow of the system includes:
1. Dataset Collection
2. Data Preprocessing
3. Feature Encoding
4. Model Training
5. Attack Prediction
6. Confidence Evaluation
7. Threat Categorization
8. Logging and Visualization

3.3 Description of the Dataset

The NSL-KDD dataset will be used in both training and testing the IDS model. The dataset has many records that are divided into:

Attack Category	Description
Normal	Legal network traffic without malicious activity
DoS (Denial of Service)	Attacks intended to disturb network services by overwhelming resources
Probe	Attacks used to gather information about target systems and network vulnerabilities
R2L (Remote to Local)	Unauthorized access tys from a remote machine to a local system
U2R (User to Root)	Attacks where a normal user gains root or administrative privileges

3.4 Pre-processing of the Dataset

Step No.	Preprocessing Activity	Description
1	Dataset Loading	NSL-KDD dataset loaded using Panda's library
2	Data Inspection	Dataset structure and records analysed
3	Feature Encoding	Categorical variables converted into numerical values using Label-Encoder
4	Feature Selection	Features separated from target labels
5	Train-Test Split	Dataset divided into training and testing datasets
6	Model Preparation	Processed dataset prepared for machine learning algorithms

Protocol type

Value	Description
tcp	Transmission Control Protocol
udp	User Datagram Protocol
icmp	Internet Control Message Protocol

service

Contains many categories such as:

Service Type	Description
http	Hypertext Transfer Protocol
ftp data	File Transfer Protocol Data Service
smtp	Simple Mail Transfer Protocol
telnet	Remote Terminal Service
Domain u	Domain Name Service
Eco i	Echo Service
private	Private Network Service
finger	User Information Service
ftp	File Transfer Protocol
ssh	Secure Shell Service
pop_3	Post Office Protocol Version 3
imap4	Internet Message Access Protocol
others	Additional network services available in the dataset

flag

Flag Value	Description
SF	Normal connection completion
S0	Connection attempt seen, no reply
REJ	Connection rejected
RSTR	Connection reset by responder
SH	Connection established and terminated
S1	Connection established but not terminated
S2	Connection established with partial

	completion
S3	Connection fully established
OTH	Other connection states
RSTO	Connection reset by originator
RSTOS0	Connection reset without response

3.5 Machine Learning Algorithms

Decision Tree

A supervised learning model that uses feature values to generate decision rules. It is simple,
easy to interpret, and computationally efficient.

Random Forest

An ensemble machine learning technique that utilizes a multitude of decision trees. It was chosen as the final model due to its superior performance.

3.6 Model Training

The dataset is partitioned into training and testing data following the 80/20 rule. Training and evaluation of both the Decision Tree and Random Forest models are carried out based on accuracy. The optimal model is saved using the Joblib library.

3.7 Real-time Detection Engine

The real-time detection engine consists of:

1. Loading the trained model.
2. Loading the encoders.
3. Picking random packets.
4. Determining attack types.
5. Computing confidence values.
6. Categorizing attacks.
7. Logging attacks.
8. Updating dashboard.

3.8 Confidence-aware Classification

Threshold set at 0.60. Any prediction below this threshold will be labled as "unknown," eliminating the need for uncertainty in attack classification.

3.9 Dashboard Visualization

Features offered by the dashboard include:

- Real-time attacks statistics
- Real-time pie-chart visualization
- Service-specific attack statistics
- Real-time logging
- Dynamic update every 2 seconds

Chart.js is used for visualizations.

IV. CONCLUSION

The suggested Machine Learning Driven Cyber Intrusion Detection and Visualization System effectively combines the use of machine learning in intrusion detection together with
real-time monitoring and visual representation. Using the NSL-KDD Dataset and applying
Decision Tree and Random Forest algorithms, this system is able to accurately classify
network traffic into different types of attacks. Random Forest proved to be more efficient
in detecting attacks compared to Decision Tree. The inclusion of confidence-based
classification increases the accuracy of predictions, while the dashboard implemented using
Flask allows real-time analysis and visualization of attacks.

REFERENCES

[1] D. E. Denning, “An intrusion-detection model,” *IEEE Transactions on Software Engineering*, vol. SE-13, no. 2, pp. 222–232, Feb. 1987.

[2] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 dataset,” in *Proc. IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, Ottawa, ON, Canada, 2009, pp. 1–6.

[3] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[4] Scikit-learn Developers, *Scikit-learn: Machine Learning in Python*. [Online]. Available: https://scikit-learn.org

[5] Flask Developers, *Flask Documentation*. [Online]. Available: https://flask.palletsprojects.com

[6] G. Kim, S. Lee, and S. Kim, “A novel hybrid intrusion detection method integrating anomaly detection with misuse detection,” *Expert Systems with Applications*, vol. 41, no. 4, pp. 1690–1700, 2014.

[7] Javaid, Q. Niyaz, W. Sun, and M. Alam, “A deep learning approach for network intrusion detection system,” in *Proc. EAI International Conference on Bio-inspired Information and Communications Technologies (BICT)*, New York, NY, USA, 2016, pp. 21–26.

[8] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *Proc. IEEE Symposium on Security and Privacy*, Oakland, CA, USA, 2010, pp. 305–316.

[9] W. Stallings, *Network Security Essentials: Applications and Standards*, 6th ed. Upper Saddle River, NJ, USA: Pearson Education, 2017.

[10] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," *Journal of Computational Science*, vol. 25, pp. 152–160, 2018.