International Journal of Engineering Research and Science & Technology

NETWORK SECURITY ENHANCEMENT THROUGH MACHINE LEARNING BASED CYBER ATTACK DETECTION

Dr.M.Venkateswara Rao 1,Kapakayala Jahnavi 2, Malladi Narendra3, Chappidi Hema Sai4
#1 Professor , In Department Of Information Technology,NRI Institute Of Technology,Agiripalli
#2#3#4 B.Tech With Department Of Information Technology In NRI Institute Of Technology,Agiripalli

Abstract: In today's digital world, cyber-attacks such as DDOS, ransomware, and botnet attacks are increasing rapidly, creating a strong need for intelligent network security. An Intrusion Detection System (IDS) helps monitor network traffic and identify suspicious activities. Recent research shows that machine learning (ML) techniques can greatly improve intrusion detection because they learn patterns from data and accurately classify attacks. Different ML models—such as Naïve Bayes, SVM, KNN, Decision Tree, Random Forest, XGBoost, and Deep Learning—have been tested on modern datasets like UNSW-NB15 to compare their performance. Studies highlight that ensemble models and advanced algorithms like Random Forest and XGBoost often achieve high accuracy, precision, recall, and detection rates for modern cyber-attacks. Overall, ML-based IDS provides faster, more reliable detection of both known and unknown attacks, making it suitable for real-time network protection.

KEYWORDS — Cyber attacks, Machine Learning, Network Security, Intrusion detection, Support Vector Machine (SVM), Random Forest, Decision Trees, Anomalous activity detection

I.INTRODUCTION

The growing digitization of global activities demands strong security measures to protect the integrity, availability, and confidentiality of networked data from rising cyber threats. Modern network systems deal with constant challenges from advanced and rapidly changing malicious activities, including large-scale Distributed Denial of Service (DDOS) attacks and ransomware. In response, Intrusion Detection Systems (IDSs) have become a key security tool, designed to monitor network traffic, spot suspicious patterns, and provide timely alerts. However, the large volume and complexity of today's network data, along with the smarter nature of advanced persistent threats (APTs), require a change from traditional detection methods. As a result, this work includes several comparative studies and reviews that look into the use of advanced Machine Learning (ML) and Deep Learning (DL) techniques to improve network security. This research specifically aims to develop high-performance, reliable Network Intrusion Detection Systems (NIDS). Through detailed experimentation on modern, state-of-the-art datasets like UNSW-NB15 and CICIDS-2017, this study implements and evaluates various ML algorithms, including Random Forest (RF), Support Vector Machine (SVM), Linear SVC, XGBoost, and K-Nearest Neighbour (KNN). A key goal is to compare the effectiveness, strengths, and weaknesses of these models while also tackling practical challenges, particularly the widespread class imbalance issue, to achieve high accuracy and an exceptional detection rate for modern network intrusions..

.In this project, machine learning techniques such as Decision Tree, Support Vector Machine (SVM), and Deep Belief Network (DBN) are employed to create an effective intrusion detection model. These algorithms classify network traffic as either "normal" or "malicious" based on learned patterns. Machine learning boosts accuracy, cuts down on false alarms, and improves the ability to detect new and evolving cyber threats.

The primary goal of the project "Detection of Cyber Attacks in a Network Using Machine Learning Algorithms" is to create a smarter intrusion detection system that can secure networks better. By using modern ML techniques, the system can provide quicker detection, improved prediction of attacks, and stronger defence against cybercrimes. Ultimately, this project aims to enhance network safety, protect sensitive data, and ensure a secure digital environment for users.

II. LITERATURE SURVEY

Many researchers have used the NSL-KDD dataset as a standard for testing intrusion detection systems because it enables fair comparison of different methods. Early studies mainly focused on Artificial Neural Networks (ANNs) with improved back propagation techniques. However, these models often showed lower performance when tested on unseen or unlabelled data, which highlighted the issue of over fitting. Later, decision-tree-based models like J48 gained popularity and were tested using 10-fold cross-validation on the training data.

Some researchers reduced the original 41 features to about 22 to speed up processing without losing accuracy. Among tree-based methods, Random Tree and Random Forest frequently performed better, achieving higher accuracy and lower false-positive rates. Other studies explored probabilistic models like Discriminative Multinomial Naive Bayes (DMNB) in two-level classification frameworks. These frameworks used a fast base classifier followed by a stronger second-level model like Random Forest, which improved detection and reduced false alarms. Feature selection techniques such as information gain, PCA, and behaviour-based ranking were also used frequently.

These methods reduced the feature set to about 23 or even 20 features while improving accuracy for several attack categories. Some researchers looked into Support Vector Machines (SVM), particularly with the RBF kernel. They found that using PCA for feature reduction combined with SVM classification led to high detection accuracy.

Unsupervised clustering techniques like k-means were also tested, but their performance dropped when evaluated on real test data rather than just training data. Hybrid approaches like fuzzy classification with genetic algorithms achieved over 80% accuracy with low false positives. Lesser-known models like Optimal Path Forest (OPF) provided fast computation with competitive accuracy compared to SVM. Deep learning techniques such as Deep Belief Networks were also tried and showed strong learning capability when properly tuned. Overall, the literature indicates that feature selection, ensemble models, and multi-level classifiers generally provide better results. Additionally, the effective use of training and testing datasets is crucial for realistic performance evaluation.

III. DETECTION OF CYBER ATTACKS IN NETWORKS USING MACHINE LEARNING ALGORITHMS

Most methods used in today's intrusion detection systems (IDS) struggle to handle the changing and complex nature of cyber-attacks on computer networks. Recently, machine learning for cyber security has gained significant attention because of its effectiveness in addressing these issues. An IDS typically doesn't detect intrusions until they have progressed deeper into the network. This leaves your systems exposed until the intrusion is found. This is a major concern, especially as encryption becomes more common to protect our data.

One big problem with IDS is that they often generate false alerts. In many cases, these false alerts outnumber genuine threats. If not monitored carefully, real attacks can be overlooked or ignored.

When an IDS identifies suspicious activity, it usually sends a report to a security information and event management (SIEM) system. This is where actual threats are identified among normal traffic fluctuations or other false alarms. However, the longer it takes to recognize a real threat, the more damage can occur. An IDS is very useful for monitoring the network, but its effectiveness relies on how you use the information it provides. Because detection tools do not block or fix potential problems, they are not effective at increasing security unless you have the right team and policies in place to manage them and respond to any threats. Additionally, an IDS cannot inspect encrypted packets, allowing intruders to exploit them to enter the network.

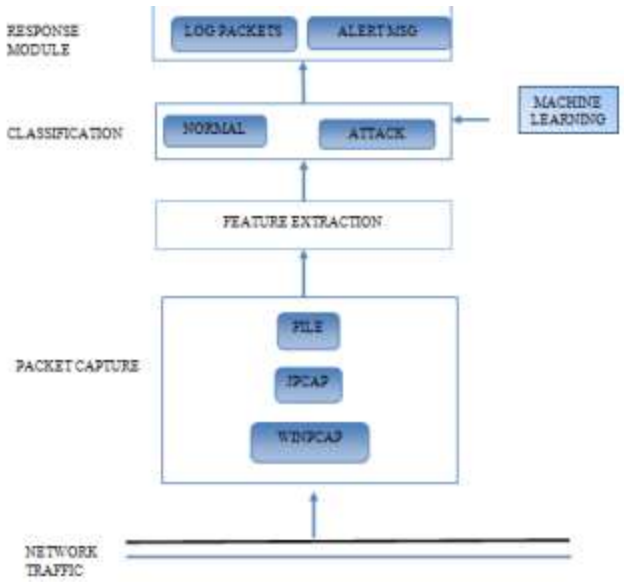

FIG1:SHOWS THE SYSTEM ARCHITECTURE

Network Traffic: Network traffic is the flow of data across a network at any moment, usually in the form of packets.

Looking at network traffic helps organizations find users or applications that use a lot of bandwidth. It also helps spot unusual activity that could indicate security threats and plan for future network needs effectively.

Packet Capture: Packet capture means intercepting and storing data packets that travel through a specific point in the network. These captured packets are analysed later to find security threats, troubleshoot abnormal network behaviour, detect congestion, uncover packet loss, and assist with forensic network investigations.

Classification: Classification is a supervised machine learning technique often used in cyber security. It helps sort data by identifying whether an email is spam. Common machine learning algorithms used for classification include Logistic Regression, Random Forest, Naive Bayes, Support Vector Machine, and Decision Trees.

Securing confidential information Algorithms:

Artificial Neural Network (ANN): The idea behind an ANN is to mimic how the human brain works. An ANN has an input layer, several hidden layers, and an output layer. The units in neighbouring layers are fully connected. An ANN consists of a large number of units and can theoretically estimate complex functions; thus, it has strong fitting ability, especially for nonlinear functions. However, training ANNs takes a lot of time due to the complex model structure.

Support Vector Machine (SVM): The goal of SVMs is to find a maximum margin hyperplane in the n-dimensional feature space. SVMs can produce good results even with small training sets because the hyperplane is determined by a few support vectors. However, SVMs are sensitive to noise near the hyperplane.

K-Nearest Neighbour (KNN): The main idea of KNN is based on a simple theory. If most of an example's neighbours belong to the same class, the example is likely to belong to that class too. Therefore, the classification result depends solely on the top-k closest neighbours. The parameter k significantly affects the performance of KNN models. A smaller k makes the model more complex, increasing the risk of over fitting. Conversely, a larger k makes the model simpler, reducing its fitting ability.

Proposed System: Machine learning algorithms can help train and identify cyber attacks. Once an attack is detected, an email can be sent to security engineers or users. Any classification algorithm can determine if an attack is a DOS or DDOS. One example is Support Vector Machine (SVM), a supervised learning method that analyses data and detects patterns. We cannot control when, where, or how an attack might occur, and we cannot guarantee total prevention. Therefore, early detection is our best option, as it can reduce the risk of severe damage from such incidents. Organizations can use existing solutions or create their own to spot cyber attacks at a very early stage and lessen the impact. A system that requires minimal human involvement would be ideal.

IV. SYSTEM ANALYSIS AND IMPLEMENTATIONS

A system is a structured group of interrelated components connected according to a plan to achieve a specific goal. Its main features include organization, interaction, dependence on one another, integration, and a central purpose. System analysis and design apply the system approach to solving problems, often using computers. To rebuild a system, the analyst must look at its elements, including inputs and outputs, processors, controls, feedback, and environment.

IMPLEMENTATION: The system is implemented using ANACONDA software. Anaconda is the most popular data science platform and the foundation of modern machine learning.

Support vector machine (SVM) is a commonly used supervised machine learning model. SVM works by finding a hyperplane that best separates the dataset into two classes on either side. Each side of the hyperplane represents a different class. The class of each data point depends on which side of the hyperplane it falls on. SVM requires a lot of space and time to work with larger and noisier datasets. The computational complexity of SVM is O(n2), where n represents the number of instances. A confusion matrix is used to evaluate the performance of a machine learning classifier.

Cyber attack detection techniques are divided into two categories: signature-based and inconsistency-based. Both methods use machine learning techniques. The authors improved the detection of Denial-of-Service (DOS) attacks. They created a Naive Bayes classifier based on feature vectors that included different User Datagram Protocol (UDP) and Transmission Control Protocol (TCP) packets and their sizes. It has also been demonstrated that Discrete Wavelet Transform and Matching Pursuit can effectively calculate features based on various organizational boundaries.

Error Rate: The error rate (E Rate) is a percentage of the total number of misclassified instances to all instances of the dataset. E Rate = (F Positive + F Negative)/ (T Negative + F Positive + F Negative + T Positive)

Recall: The recall is a percentage of correctly classified positive instances to the total number of positive instances classified in the dataset. Recall = T Positive / (T Positive + F Negative)

Precision: The precision is a percentage of the total number of positive instances classified to the total number of positive instances. Precision= T Positive / (T Positive + F Positive)

Fig 2:Home Page

Fig 3: input url testing

Fi 4: Predicting the given URL is safe or unsafe with accuracy

V. CONCLUSION

Network-based Intrusion Detection Systems (NIDS) are very effective for addressing modern network security threats, especially with the use of machine learning (ML) models. As cyber-attacks become more complex, especially with the growth of IOT devices, effective detection systems have become crucial. Researchers have tested various ML algorithms such as Linear SVC, Logistic Regression, Random Forest (RF), Decision Trees, and XG Boost using the UNSW-NB15 dataset. Their results indicate that RF and XG Boost offer better detection performance than other models.

In a similar way, studies on the CICIDS2017 dataset assessed SVM, ANN, CNN, and RF models. The results showed that deep learning models, particularly neural networks, achieved higher accuracy than traditional ML models like SVM and RF. These methods use historical attack data stored in datasets to find patterns that signal cyber-attacks. Overall, the research confirms that both ML and deep learning techniques can effectively detect network intrusions. Future work may look into large-scale applications using technologies like HadOop and Spark for real-time analysis.

VI. REFERENCES

[1] Intrusion Detection: Comparision of Machine Learning models on unbalanced datasets, SN Computer Science, Sept-2024.

[2] Network Intrusion Detection using Machine Learning, Deep Learning, **2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT).**

[3] Cyber attack detection in Networks using ML Algorithms, South Asian Journal of Engineering and Technology, July 2022.

[4] Predicting the System failures using Machine Learning Algorithms, International Journal of Advanced Scientific Innovation, volume 1, no 1, Dec 2022.

[5] A Classification Model for Intrusion Detection Based Machine Learning, Amity School of Engineering and Technology.

Author Profiles

KAPAKAYALA JAHANAVI is a final-year B.Tech student in Information Technology with a strong interest in emerging technologies and intelligent systems. Her academic interests include Artificial Intelligence, Machine Learning, Computer Vision, and Data Analytics. She has been actively involved in research and project development related to AI-based applications. One of his notable works includes contributing to a facial emotion detection system designed for retail environments to analyze customer perception using computer vision techniques. Through this project, he gained practical experience in image processing, machine learning models, and data analysis. She is passionate about applying technological solutions to real-world problems. Bharath Kumar is also interested in exploring advancements in deep learning and smart automation systems. She aims to further develop his skills in AI-driven technologies and innovative software solutions. Her goal is to contribute to impactful research and technological development in the field of Information Technology.

MALLADI NARENDRA is a final-year B.Tech student in Information Technology at NRI Institute of Technology, Agiripalli, Andhra Pradesh. He has a strong interest in advanced technologies and intelligent computing systems. His main areas of interest include Artificial Intelligence, Machine Learning, and Deep Learning.He has completed a Machine Learning certification from NPTEL, which helped him gain a solid foundation in machine learning concepts and algorithms. In addition, he completed an internship with Blackbox, where he gained practical exposure to deep learning technologies and AI-based applications.Through his learning and practical experience, he developed skills in data analysis, machine learning models, and AI-driven solutions. He is highly motivated to explore innovative technologies and apply them to solve real-world problems.He is passionate about building his career in AI, ML, and Deep Learning–based roles. His goal is to continuously improve his technical skills and contribute to innovative developments in the field of Information Technology and intelligent systems.

CHAPPIDI HEMA SAI is an aspiring Information Technology professional with a strong interest in software development, Artificial Intelligence, and Machine learning. He is currently pursuing his undergraduate studies in the field of Information Technology and continuously works on improving his technical and problem-solving skills. He has completed professional learning programs such as the Accenture Software Engineering Job Simulation and HP Life – AI for Beginners, which helped him gain practical knowledge about real-world software development practices and emerging technologies.He has also worked on machine learning and programming projects where he explored concepts such as data analysis, model training, and basic AI applications. Through these projects, he developed hands-on experience with programming and analytical thinking. He passionate about learning new technologies, building innovative solutions, and applying his knowledge to solve real-world problems. His goal is to grow as a software developer and contribute to impactful technological advancements in the IT industry.