Phoenix: International Multidisciplinary Research Journal
Vol 3, No. 4, Oct-Dec, 2025
ISSN:2583-6897

INTRUSION DETECTION SYSTEM USING WSN NOVEL DT, RF, AND MLP ALGORITHMS

Dr Syed Umar
Professor, Department of Computer Science & Engineering, Malla Reddy (MR) deemed to be University, Hyderabad, syedumar@mrec.ac.in
Goli Madhuri
Assistant Professor, Department of computer science and engineering, Malla Reddy (MR) Deemed to be University, Hyderabad, golimadhurireddy120@gmail.com
Dikshendra Daulat Sarpate
Professor, Department of Artificial Intelligence & Data Science, ZEAL College of Engineering & Research, Pune, Email id -dikshendra@gmail.com
Gopala Soujanya
Assistant Professor, Department of computer science and engineering, Malla Reddy (MR) Deemed to be University, Hyderabad, Soujanyachiluka52@gmail.com
B.Rani
Assistant Professor, Department of computer science and engineering, Malla Reddy (MR) Deemed to be University, Hyderabad, rani@mrec.ac.in

Abstract—**The security of computer networks has emerged as a critical issue in the current digital era, as information is readily shared, and connection is pervasive. Organisations, governments, and people are all at danger from the enhancing frequency as well as complexity of cyber assaults. The need for reliable cybersecurity solutions has never been more pressing as criminal actors attempt to infiltrate sensitive data and exploit vulnerabilities on a constant basis. Network security leaders and a key line of defence against cyberattacks are intrusion detection systems (IDS). These advanced tools are designed to continuously scan for any indications of suspect or unauthorised behaviour while keeping a close eye on system operations and network traffic in real time. IDS are important for stopping unauthorized access attempts, service outages, and data breaches by quickly identifying and reacting to possible threats. The field of intrusion detection systems is observed in this study along with its underlying theories, methodology, and real-world uses. We'll examine how IDS may assistorganizations in safeguarding their digital assets and maintaining network integrity by promptly detecting and addressing various cyberattacks. The introduction will provide a general overview of the growing cybersecurity problems that organisations and people throughout the globe are facing.**

We investigate the components, sensor location, and data collecting of both Network-based Intrusion DetectionSystems (NIDS) and Host-based IntrusionDetection Systems (HIDS). NIDS stands for Network-based IntrusionDetection System. HIDS is for Host-based Intrusion Detection System. In addition, we went further into the complexities of signature-based, AnomalyBased, and BehaviorBased detection approaches, as well as the incorporation of machine learning approaches for improved intrusion detection capabilities. The difficulties associated with implementing an IDS were investigated; these difficulties included evasion methods used by attackers, concerns about privacy, challenges with scalability, and resource limits.

Index Terms—Wireless Sensor Network, Intrusion Detection, Machine Learning, Deep Learning, Energy Efficiency.

I. INTRODUCTION

Security and privacy in WSNs have become more pressing concerns as their use has spread across a wide range of industries. To detect and prevent security issues and intrusion attempts in WSNs, a powerful and effective tool is needed, and such a tool is the Wireless Sensor Network-Based Intrusion Detection System (WSN-IDS). This paper provides an in-depth analysis of the design and deployment of the WSN-IDS, covering everything from data collection through intrusion detection. The system uses the accumulated

knowledge of several sensor nodes to keep tabs on the network's hardware, data flow, and conversational habits. By analysing the gathered data, the WSN-IDS can efficiently distinguish between legitimate network activities and potential security threats, including intrusion attempts, data tampering, and denial-of-service attacks. To handle and interpret sensor data, the WSNIDS relies on cutting-edge machine learning algorithms and data fusion methods.

Protecting the integrity and security of these assets has become a top priority in today's digitally linked world as information travels across networks and systems. Intrusion DetectionSystems (IDS), created to identify and counteract unauthorised or hostile actions inside computer networks and systems, have become essential elements of cybersecurity measures. IDS concentrates on locating and reacting to possible threats that have gotten past the perimeter defences, in contrast to conventional perimeter security systems that put their attention on preventing unauthorised access [1,2].

A software or hardware solution known as an Intrusion DetectionSystem (IDS) analyses network and system activity to look for indications of unauthorised or malicious behaviour. By continually monitoring network traffic,system logs, and user activity for any aberrant patterns that could point to a security breach, an IDS primarily serves to offer another layer of defence. IDS can be broadly divided into two primary categories: Signature-Based Detection (Misuse Detection): This method entails comparing incoming data with termed patterns or signatures of well-known attacks. An alert is produced if a match is found. While signature-based detection performsproperly against known assaults, it may not be able to handle novel or undiscovered threats [3].

II. Traditional vs. Machine Learning-based IDS

The topic of cybersecurity is always evolving, and with it come new approaches for identifying threats and reducing their effects. conventional rule-based Intrusion Detection Systems (IDS) have been successful to a certain degree; however, the growing complexity of cyber threats has led to the study of more sophisticated techniques, including machine learning-based IDS [4]. This is because those conventional IDS have been able to sustain with the evolving nature of cyber threats.

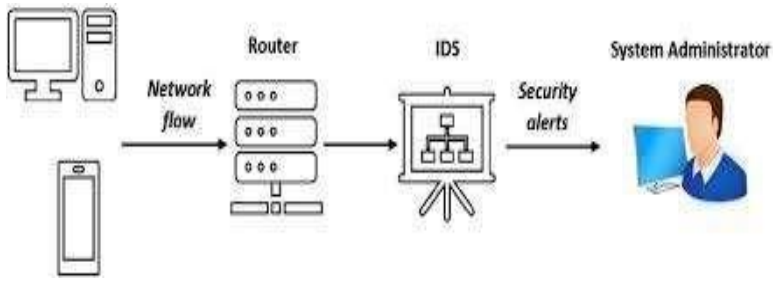

Figure 1: Graphical Design of IDS [5]

Limitations of rule-based systems: Rule-based IDS rely on predefined patterns, signatures, or rules that characterize known attack patterns. While they have their merits, they come with several limitations:

Limited to Known Signatures: Rule-based systems are efficient at detecting well-known attacks for which signatures have been defined. However, they struggle to detect new or evolving attacks that don't match any predefined rules.

False negatives: If an attack does not match any known rule. Rule-based IDS may fail to detect it, leading to false negatives.

Inflexibility: Maintaining and updating rule sets can be time- consuming and requires constant monitoring of emerging threats. This approach struggles to adapt to rapidly evolving attack techniques.

False Positives: Overly complex or broad rules can result in false positives, flagging legitimate activities as threats and inundating security teams with alerts [6,7].

Limited Context: Rule-based systems often lack the ability to consider the broader context of activities and connections, potentially leading to misinterpretation of benign activities.

Advantages of Machine Learning-Based Approaches

Machine learning-focused IDS leverage the capabilities of artificial intelligence to eliminate a few of the restrictions of traditional rule- based systems. Here are some advantages [8]:

Adaptability to New Threats: Models for machine learning are able to learn from data and adjust to novel and undiscovered assault patterns, making them more effective at detecting novel threats.
Anomaly Detection: Machine learning can excel at identifying anomalies in data, making it suitable for detecting sophisticated attacks that deviate from normal behavior.
Reduced False Positives: ML models can analyse vast amounts of data and discem complex patterns, leading to better detection accuracy and fewer false positives.
Contextual Analysis: Machine learning models can consider various attributes and contextual information when making decisions, improving the accuracy of threat detection.
Continuous Learning: Machine learning models can be modified with recent data, allowing them to evolve and adapt to changing threat landscapes without requiring manual rule updates.
Efficient Resource Utilization: Machine learning models can optimize resource u sage, leading to improved performance and reduced impact on network and system operations [9].
Behavioral Profiling: ML-based IDS can create profiles of normal behavior, allowing them to identify deviations over time rather than relying solely on predefined rules

III. Data Collection and Preprocessing

Preprocessing and data collection are crucial stages in creating a successful intrusion detection system (IDS). When data is handled correctly, machine learning models can learn from relevant information and make accurate predictions [10].

A. *Types of Data Sources*

Network Traffic Logs:These logs document network activity, including protocols utilized, source and destination IP addresses, port numbers, and incoming and outgoing packets.
System Logs: These logs record system-level events, likefile access,login attempts, process executions, and system configuration changes.
Application Logs: Application-specific logs deliver insights into the behavior of specific software or services running on the network or system.
User Activities**:** User-related data, such as login patterns, authentication attempts, and user actions, can be important for identifying anomalies.

B. *Data Cleaning and Transformation*

Data Cleaning: Remove duplicate entries, handle missing values, and address inconsistencies to ensure data quality.
Data Transformation: Numerical features should be scaled or normalized to a common scale. Create numerical representations of categorical features by employing methods such as one-hot encoding.
Feature Extraction and Selection Feature extraction includes converting raw data into meaningful features that may be utilized by machine learning models. Feature selection seeks to determine the most pertinent features to improve the model's efficiency and performance.

C. *Important Features for IDS*

IP Addresses and Ports: The IP addresses and port numbers of source and destinationcan help identify unusual communication patterns [11,12].
Protocol Types: Different protocols (TCP, UDP, ICMP) exhibit different behaviors, and analysing protocol distributions can reveal anomalies.
Packet Sizes: Anomalies in packet sizes might indicate data exfiltration or denial of service attacks.
Time Intervals: Detecting deviations in time intervals between actions can help identify brute force attacks or unauthorized access.
User Behavior: Monitoring user activity patterns can help spot suspicious login attempts or unusual behaviors.
System Calls: For host-based IDS, tracking system calls and their frequencies can provide insights into potentially malicious activities.

D. *Techniques for Feature Engineering*

Dimensionality Reduction: Principal ComponentAnalysis (PCA) is one technique that can decrease the degree of features while keeping important information [13,14].

Aggregation: Group related data together to form aggregated features, such as counting failed login attempts within a time window.

Feature Scaling: To avoid some features taking precedence over others during model training, normalize features to the same range.

Feature Interactions: Create new features by combining existing ones, such as the ratio of successful logins to failed logins.

IV. Data Labeling for Training

Labeling data involves categorizing instances as either normal or representing some form of intrusion. This labelled data is utilized to train machine learning models to differentiate between normal and malicious activities within an Intrusion Detection System (IDS) [15].

Anomaly Detection vs. Signature-Based Detection

1) Anomaly Detection**:** Finding departures from typical behavior is the goal of anomaly detection without relying on predefined attack signatures. Instances that significantly differ from the established normal baseline are flagged as anomalies. Effective for detecting previously unseen attacks, but may lead to false positives due to legitimate variations.

2) Signature-Based Detection**:** Predetermined patterns or signatures of known assaults are the foundation of signature-based detection. Instances matching these signatures are labelled as malicious. Effective for detecting well-known attacks, but limited to known threats and vulnerable to new attack variants [16].

Creating Labelled Datasets

1) Historical Data**:** Utilize historical intrusion data that has already been identitled and labelled. This data is utilized to train algorithms for machine learning to find similar trends.

2) Manual Labeling**:** Manually label instances in a dataset as normal or intrusive. Requires domain expertise to accurately identify intrusions and normal activities.

3) Simulation**:** Set up controlled environments to simulate at tacks and normal behaviors. This approach provides labelled data for both normal and intrusive instances.

4) Hybrid Approaches**:** Combine historical data with manual labelling to ensure a comprehensive dataset. This technique helps bridge the gap amongst known and novel threats. Unlabelled Data with Semi-Supervised Learning:Use a limited set of labelled data alongside a large of unlabelled data Semi-supervised learning leverages the labelled data to guide the model's learning from the unlabelled data.

V. Machine Learning Algorithms

The first step in creating a successful intrusion detection system (IDS) is choosing the appropriate machine learning methods. The choice of methods depends on the kind of learning (supervised, unsupervised, or semi-supervised) and the characteristics of the data [17].

Supervised. Unsupervisedand Semi-Supervised Learning

1) SupervisedLearning**:** In supervisedLearning, the technique is trained on labelled data, where each instance is associated with its corresponding class or label. Common for signature-based detection and certain types of anomaly detection where labelled data is available.

2) Unsupervised Learning**:** Unsupervised learning doesn't use labelled data. Instead, it recognizes patterns and structures within the data with no predefined categories Useful for anomaly detection when the types of intrusions are unknown.

3) SemiSupervised Learning**:** SemiSupervised learning joins labelled and unlabelled data for training. Can be effective when labelled intrusion data is limited but can provide valuable guidance.

VI. Algorithms for IDS

RandomForest

The RandomForest ensemble learning approach uses many decision trees to reduce overfitting and improve accuracy, efficient for detection based on anomalies as well as signatures. In practical decision-

making applications, the most common classifiers or regression are used. The decision tree calculation is enlarged into RF calculation. The structure known as the "forest" or RF is made up of a variety of trees. The main difference between both directed learning models is that RF has a minimal level of complexity and is easy to implement by selecting the best arrangements based on the sacking strategy [18,19].

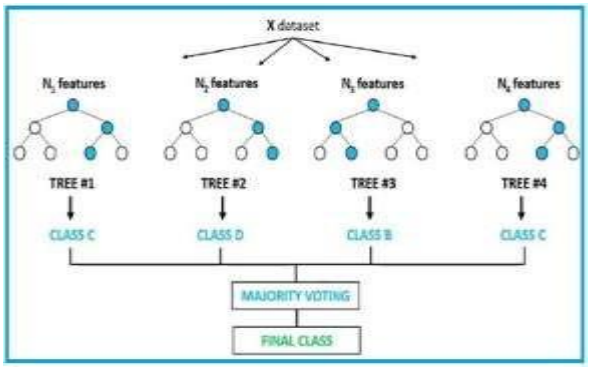

Figure 2: RF Classifier Algorithm

The main goal of using the trees in combination as a sorting approach is to achieve more accurate results as prediction or misfortune accuracy. According to the image below, the classifier that uses the X-dataset should be divided into three different classes using RF. The RF classification algorithm is displayed in Figure 2 [20,21].

Decision Trees

Decision trees, which are frequently employed for classification and regression, are another supervised learning model. In contrast to the SVM approach, the DT is a non-parametric
model. The DT structure that is best appropriate for business, risk management analysis is framed by the parent node with several offspring nodes (i.e. branches). Time complexity, which is quite high compared to other ML algorithms' preparation and testing phases,is DT's main drawback. Figure 34 depicts an example decision tree classification structure.

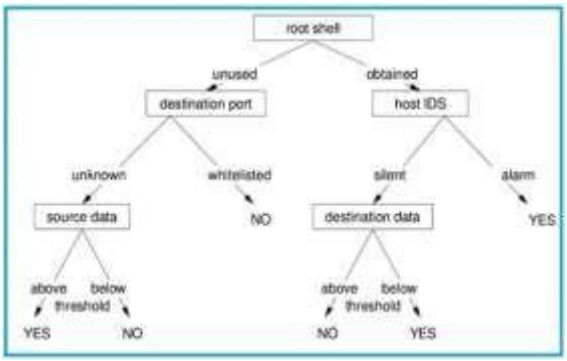

Figure 3: Decision tree sample classification structure

Records in decision trees are composed of attribute vectors, which include a set of classification characteristics that characterise the vector and a class attribute that associates the input data with a particular class. By constantly searching the database for the attribute that best separates the data into the several subclasses, up until a halting condition is satisfied, a decision tree is created. Customers can quickly summarise the facts in the presentation form since decision trees may be shown in an understandable tree-structured way. The parent and child nodes that are linked to the root node of DT are known as leafs.

VII. RESULTS

In the CyberRange Lab of the Australina Center for CyberSecurity (ACCS), the tshark program was used to create the raw packets of network (Pcap files) of the BoT-IoT dataset. These packets include both regular and aberrant traffic. Using the Ostinato tool and Node-red (non-IoT as well as IoT), simulation network traffic has been produced. The source files for the dataset are offered in a variety of formats, including the generated argus files, the original PCAP files, and lastly the CSV file. To aid in the classification procedure, the files were divided into assault categories and subcategories.
Only 0.8% and 0.2% of the enormous database is used for training as well as testing purposes, respectively. The estimate of several metrics is shown in Figure 4 , Figure 5 and Figure 6.

- Accuracy**:** It displays the chosen classifiers' accuracy percentage attributes. It is evident from the accuracy inquiry report that the constructed model has provided an optimal performance %. The instance-based classifier is the one that has a greater accuracy rate.
- Precision: The constructed model's performance quality, which has made accurate predictions. It has been misclassified if the 100% position value is stated by 1 or if the 100% percentage depreciation value obtained is smaller. The chosen classifiers' position levels vary from 0.969914 to 0.971816 for various classifiers.

- Recall**:** The ROC values and the recall value are comparable as well. For an ideal model, a 10% false negative classification is permitted. It adheres to the equation listed below. The investigative report's recall span for several classifiers ranges from 0.965254 to 0.969218.
- F1-Score**:** The recall values and the F-Measure value are comparable. For an ideal model, a 10% false negative classification is permitted. It adheres to the equation listed below. The range for F-Measure in the report is 0.968047 to 0.969688 for different classifiers.

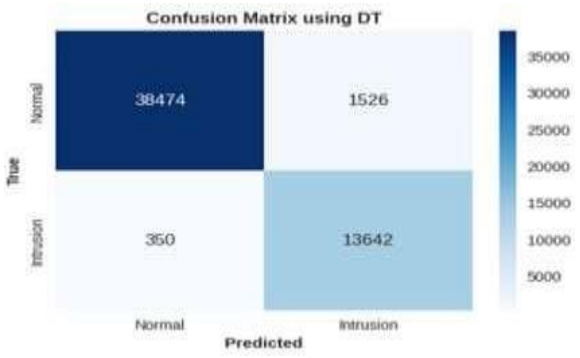

Figure 4: Confusion matrix for decision tree classifier

The confusion matrix for Decision Tree classifier, as represented in Figure 4, shows that the MSE of the classifier is 3.48%, the accuracy of the classifier is 96.53%, the precision value is 96.73%, the recall is 96.53%, and the F1-score is 96.57%.

Figure 5: Confusion matrix for random forest classifier

The confusion matrix for RF classifier, as represented in Figure 5, shows that the MSE of the classifier is 3.24%, the accuracy of the classifier is 96.73%, the precision value is 96.99%, the recall is 96.77%, and the F1-score is 96.80%.

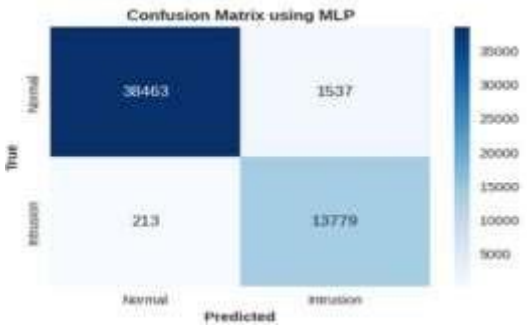

Figure 6: Confusion matrix for multi-layer perceptron classifier

The confusion matrix for MLP classifier, as illustrated in Figure 37, shows that the MSE of the classifier is 3.24%, the accuracy of the classifier is 96.77%, the precision value is 96.99%, the recall is 96.76%, and the F1-score is 96.80%.

References

[1] Adnan, A, Muhammed, A, Abd Ghani, AAA, Abdullah, A & Hakim, F2021, 'An intrusion detection system for the internet of things based on machine learning: review and challenges. Symmetry', vol. 13, no. 6, pp. 1-13.

[2] Kasongo, SM & Sun, Y 2021, 'A Deep Gated Recurrent Unit based model for wireless intrusion detection system. ICT Express, vol. 7, no. 1, pp. 81-87.

[3] Kayode Saheed, Y, Idris Abiodun, A, Misra, S, Kristiansen Holone, M & Colomo-Palacios, R 2022, 'A machine learning-based intrusion detection for detecting internet of things network attacks. Journal of Alexandria Engineering vol. 6, no. 12, pp. 9395-9409.

[4] D. Agrawal, C. Agrawal and H. Yadav. "A Machine Learning Based Intrusion Detection Framework Using KDDCUP 99 Dataset", vol. 4. no. 6, pp. 11.

[5] Umar, Syed, Bommina Naveen Sai, Nagineni Sai Lasya,Doppalapudi Asutosh, and LohithaRani. "Machine Learning based Sentiment Analysis of Product Reviews Using DeepEmbedding." Journal of Optoelectronics Laser 41, no. 6(2022): 108-113.

[6] Lyngdoh, J, Hussain, MI, Majaw, S & Kalita, HK 2019, 'An Intrusion detection method using artificial immune system approach', Communications in Computer and Information Science International conference on advanced informatics for computing research, pp. 379-387.

[7] Saheed, YK 2022, 'Performance improvement of intrusion detection system for detecting attacks on Internet of things and edge of things. In: Misra S, TKA, Piuri V, Garg L, editors, Artificial intelligence for cloud and edge computing'. Internet of things (technology, communications and computing). Cham: Springer; pp. 321-39.

[8] Zhong, W, Yu, N & Ai, C 2020, 'Applying big data based deep learning system to intrusion detection' , Journal of Big Data Min Anal, vol. 3, no. 3, pp. 181-195.

[9] Nagalalli, G & Ravi, G 2022, 'A Novel Megabat Optimized Intelligent Intrusion Detection System In Wireless Sensor Networks', Journal of Intelligent Automation and Soft Computing, Tech Science Press, vol. 35, no. 1, pp. 475-490.

[10] Naveen Sai Bommina , Nandipati Sai Akash, Uppu Lokesh , Dr. Hussain Syed , Dr. Syed Umar, "Multi-Objective Genetic Algorithms for Secure Routing and Data Privacy in IoT Networks", International Journal of Communication Networks and Information Security (IJCNIS), (2020), 12(3), 632–643.

[11] D. E. Baraneetharan, "Role of machine learning algorithms intrusion detection in WSNs: A survey," Journal of Information Technology and Digital World, vol. 2, no. 3, pp. 161–173, 2020.

[12] F. A. Khan, A. Gumaci, A. Derhab and A. Hussain, "TSDL: A Two- Stage Deep Learning Model for Efficient Network Intrusion Detection", IEEE Access, vol. 7. pp. 30373-30385, 2019.

[13] S. Jiang, J. Zhao and X. Xu, "SLGBM: An intrusion detection mechanism for wireless sensor networks in smart environments," IEEE Access, vol. 8, pp. 169548–169558, 2020.

[14] Habeeb, M. S., & Babu, T. R. (2022). Network intrusion detection system: a survey on artificial intelligence-based techniques. Expert Systems, 39(9), e13066.

[15] Naveen Sai Bommina, Nandipati Sai Akash, Uppu Lokesh, Dr. Hussain Syed, Dr. Syed Umar, "A Hybrid Optimization Framework for Enhancing IoT Security via AI-based Anomaly Detection", International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169 Volume: 11 Issue: 3.

[16] Uppu Lokesh , Naveen Sai Bommina , Nandipati Sai Akash , Dr. Hussain Syed , Dr. Syed Umar. (2021). Deep Reinforcement Learning with Genetic Algorithm Tuning for Intrusion Detection in IoT Systems. International Journal of Communication Networks and Information Security (IJCNIS), 13(3), 582–595.

[17] Nandipati Sai Akash, Uppu Lokesh, Naveen Sai Bommina, Hussain Syed, Syed Umar, "Swarm Intelligence-Based Hyperparameter Optimization for AI-Powered IoT Threat Detection", International Journal of Intelligent Systems and Applications in Engineering, (2024), 12(17s), 941.

[18] Uppu Lokesh, Naveen Sai Bommina, Nandipati Sai Akash, Dr. Hussain Syed, Dr. Syed Umar, "Designing Energy-Efficient and Secure IoT Architectures Using Evolutionary Optimization Algorithms", International Journal of Applied Engineering & Technology, Vol. 4 No.2, September, 2022.

[19] Habeeb, M. S. (2024). Predictive analytics and cybersecurity. Intelligent Techniques for Predictive Data Analytics, 151-169.

[20] hakre N, Nimma D, Turukmane AV, Singh AK, Rohatgi D, Bangaru B (2024) Dynamic path planning for autonomous robots in forest fire scenarios using hybrid deep reinforcement learning and particle swarm optimization. Int J Adv Comput Sci Appl 15(9).

[21] M. Mukhedkar, D. Rohatgi, V.A. Vuyyuru, K.V.S.S. Ramakrishna, Y.A. Baker El-Ebiary, V.A. Asir Daniel, "Feline wolf net: A hybrid lion-grey wolf optimization deep learning model for ovarian cancer detection", Int. J. Adv. Comput. Sci. Appl., 14 (9) (2023)