Anomaly Network Traffic Detection of wireless Network System

D. Deepika
Computer Science And Engineering
B V Raju Institute of Technology
Narsapur, India
deepika.d@bvrit.ac.in

Deepika pogiri
Computer Science And Engineering
B V Raju Institute of Technology
Narsapur, India
pogirideepika2111@gmail.com

Lokesh Raj Pandravisham
Computer Science And Engineering
B V Raju Institute of Technology
Narsapur, India
pandravishamlokeshraj@gmail.com

Yashwanth Kumar Prudvi
Computer Science And Engineering
B V Raju Institute of Technology
Narsapur, India
yashu.prudhvi@gmail.com

Sathvik Reddy Ramannagari
Computer Science And Engineering
B V Raju Institute of Technology
Narsapur, India
reddysathvik72@gmail.com

***Abstract*—Network traffic anomaly detection is a current topic in network security. Based on unsupervised learning, this paper constructs a Model for Detecting Irregularities in Network Data to solve the problems of high dimensions of abnormal traffic. Among the prevalent threats to network security, anomalies stand out as a significant menace, capable of causing system malfunctions and impeding proper network functionality. Detecting these anomalies is imperative for ensuring the uninterrupted operation of networks. While DL and ML algorithms have showcased their potential in detecting network anomalies, their efficacy remains relatively uncertain. This paper conducts a detailed inspection of the algorithms (Isolation Forest and Local Outlier Factor) using the datasets to assess their proficiency in identifying network anomalies. The study is committed to offering a thorough examination of various applications of deep learning and machine learning, with the overarching goal of fortifying network security. We will closely inspect how well they fare in practice based on accuracy. In addition, the study delves into the influence of algorithmic feature selection on overall performance. The insights derived from this investigation are programmed to direct the development of innovative techniques focused on boosting network security. Using datasets, we can improve the detection.**

***Index Terms*—Network traffic anomaly detection, Unsuper vised learning, Anomalies, System malfunctions, Efficacy, Com prehensive analysis, Fortifying network security, Isolation Forest and Local Outlier Factor.**

I. Introduction

Anomaly detection in Wireless networks, vital for security and performance, utilizes Isolation Forest and DBSCAN. Isolation Forest isolates anomalies by partitioning data space recursively, while DBSCAN identifies outliers based on density clustering. The process involves data collection, preprocessing, feature extraction, model training, anomaly identification, and response. Challenges include feature selection, model tuning, scalability, and false positives. Applications range from security monitoring to performance optimization and resource allocation.

II. Related Work

The literature surrounding anomaly detection and its application in various domains, including finance, emphasizes the challenge posed by sudden deviations in metrics and the critical need for timely attention to both spikes and dips.This project adopts an unsupervised approach, employing algorithms like Isolation Forest, One-Class SVM, and LSTM to model and address the anomaly detection problem. The focus here is on identifying anomalies using the Isolation Forest algorithm. The literature review explores the intricacies of stock market investing, highlighting the dynamic nature of share prices and the challenges investors face in making timely decisions to mitigate losses and gain profits. The study delves into the significance of comprehensive financial analysis, market history, business tendencies, and external factors in predicting stock market fluctuations, emphasizing the pivotal role of anomaly detection in enhancing decision making processes for investors in this complex and unpredictable field [1].

The landscape of Industrial Control Systems (ICS) security is marked by increasing cyber threats due to the lack of encryption and authentication in protocols. Notable incidents such as Stuxnet and Triton underscore the urgency for robust security measures. Intrusion Detection Systems (IDSs) are piv otal in mitigating these threats, leveraging Machine Learning and Deep Learning for enhanced detection capabilities. However, testing IDSs on operational systems poses challenges, necessitating the development of realistic datasets for evaluation. Existing datasets are limited, often unrealistic, and lack labeled data crucial for effective model training. This study addresses these challenges by proposing the ICS Flow dataset, facilitating ML-based intrusion detection model evaluation. The comprehensive dataset encompasses diverse network flow features and realistic attack scenarios, enabling rigorous assessment of intrusion detection methods for ICS

security enhancement[2].

The evolving landscape of computer networks necessitates robust intrusion detection systems (IDS) to counter diverse threats. Signature-based methods rely on predefined attack patterns, limiting their efficacy against unknown attacks. Anomaly-based IDS, leveraging machine learning, addresses dynamic threats but requires extensive feature engineering and struggles with large-scale data. The emergence of deep learning offers promise, with graph neural networks (GNNs) capturing network topology for improved detection. However, existing GNNs lack dynamic aggregation capabilities vital for network flow analysis. This study introduces Edge-Directed Graph Multi-Head Attention Networks (EDGMAT), employ ing a multi-head attention mechanism to enhance correlation exploration and topological feature utilization [3].

The integration of Network Function Virtualization (NFV) in sensor and IoT networks offers numerous benefits, includ ing enhanced resource utilization and network management. However, it also introduces security challenges that necessitate effective mitigation strategies. This survey examines NFV's se curity implications and proposes anomaly detection techniques to counter potential cyber threats. By evaluating machine learning algorithms for anomaly detection, the research aims to bolster NFV network security. NFV enhances IoT network f lexibility and scalability, enabling dynamic deployment of network functions and improved resource allocation. Addition ally, virtualized security functions enhance the security posture of sensor and IoT networks. The study underscores NFV's importance in optimizing resource utilization and improving network security, offering valuable insights for network ad ministrators and security professionals [4].

This survey examines the use of anomaly detection in blockchain networks, focusing on analyzing network traffic traces for identifying malicious activities. By employing a semi-supervised learning approach, the study proposes an anomaly detection engine based on AutoEncoder (AE) tech nology. Evaluation of the proposed mechanism demonstrates promising results in terms of accuracy and time complexity, supporting its effectiveness in real-time threat detection. The survey emphasizes the importance of securing blockchain networks and offers insights into the development of robust anomaly detection systems [5].

Traditional security measures often fall short in identifying evolving threats, prompting a shift towards machine learning based anomaly detection. By analyzing network traffic data, machine learning algorithms can accurately classify normal and malicious behavior, offering enhanced threat detection capabilities. These techniques adapt to new data and minimize false positives, providing organizations with proactive security measures. Integrating machine learning into network security strategies enables organizations to safeguard against potential threats and protect valuable digital assets effectively [6].

Traditional anomaly detection methods like firewalls and intrusion detection systems are inadequate, given the evolving network landscape and proliferation of cyber threats. Machine learning (ML) offers a promising solution, leveraging statistical analysis to swiftly identify anomalies and enhance network security. ML's adaptability and quick response to abnormal behavior make it a valuable asset in early threat detection. However, effective ML implementation hinges on meticulous data preparation. This paper explores ML algorithms and their application in anomaly detection across diverse network envi ronments, addressing this critical aspect of network security [7].

The escalating complexity of network environments has underscored the importance of anomaly detection in network traffic for robust security management. Conventional rule based methods exhibit limitations in adaptability and scal ability, prompting the adoption of machine learning techniques. Supervised models, such as neural networks and support vector machines, leverage labeled datasets for classification, while unsupervised and semi-supervised approaches uncover latent patterns in unlabeled data. Feature selection and dimensionality reduction enhance model efficiency. Novel methods, like SVM-C, integrate statistical laws and linear coding for anomaly detection, exhibiting superior performance [8].

Rule-based systems require frequent updates, making them less efficient. Using anomaly detection systems to define normal behavior enables the detection of unseen anomalies. In collaboration with Trimma, a decision support company, machine learning models are applied to analyze event logs for abnormal execution times. Various approaches are evaluated to detect anomalies efficiently. This dissertation explores the effectiveness of machine learning techniques in identifying abnormal execution times, aiding in the timely detection of issues and root cause analysis in Trimma's data processing operations [9].

As the proliferation of networking devices accelerates, ensuring secure access to sensitive data has become paramount. Intrusion Detection Systems (IDS) play a vital role in safe guarding networks, particularly anomaly-based IDS, which identify abnormal behavior as potentially malicious. By train ing machine learning models on normal behavior, anomalies can be detected and alerts raised accordingly. While IDS have evolved from heuristic-based systems to more sophisticated data-driven approaches, challenges persist, such as the risk of false positives. Despite advancements, detecting benign instances as false positives remains a concern, leading to resource wastage. With the exponential rise in digital attacks, companies increasingly rely on IDS for network security, emphasizing the need for effective anomaly detection methods to combat evolving threats and ensure data integrity [10].

III. Problem statement

• The increasing complexity and diversity of wireless net work infrastructures pose significant challenges in effectively detecting anomalies, leading to potential security breaches and performance degradation.

• The scalability and adaptability of anomaly detection methods are crucial for addressing the growing complexity and scale of wireless networks.

IV. Objectives

1) Implementing the classification (isolation forest) and clustering (DB scan) algorithms for anomaly detection of wireless network system .

2) Ensemble method(boosting) is used to combine both classification and clustering algorithms to improve the accuracy of the result.

V. Existing system

The existing model focuses on anomaly detection using classification(Isolation Forest) algorithm in unsupervised ma chine learning. It emphasizes early classification of anomalies during data partitioning. The model aims to predict stock market trends efficiently, crucial for investors. It incorporates verification and validation processes for software testing. The anomaly score calculation is based on the average path length in a Binary search tree. Future enhancements include combining machine learning algorithms for improved accuracy, feature normalization, and real-time anomaly detection in network security.

VI. Proposed Work

The proposed study aims to investigate anomaly detection in wireless networks, focusing on the efficacy ensemble method of Isolation Forest and DB scan clustering compared to other established methods. Wireless networks are prone to various anomalies, making robust detection mechanisms crucial for ensuring security and performance. Isolation Forest has emerged as a promising technique in anomaly detection due to its ability to isolate outliers efficiently.DB scan form clusters and seperate outliers efficiently. The study will commence with a comprehensive review of existing anomaly detection methods in wireless networks, highlighting the principles of Isolation Forest , DB Scan and there application context. Methodologically, the study will utilize relevant datasets, preprocess them by cleaning and selecting features, and implement Isolation Forest ,DB Scan clustering. Evaluation metrics like accuracy, precision, recall, and F1-score will be employed to assess the performance of each method.

VII. Methods

A. Dataset

We extracted WSN-DS dataset from [14] . The dataset consists of 477,426 entries of network traffic data, each representing an individual packet with 14 attributes. Key attributes include `frame.number` (unique frame identifier), `frame.time` (capture timestamp), `frame.len` (frame size in bytes), `eth.src` and `eth.dst` (source and destination Ethernet addresses), `ip.src` and `ip.dst` (source and destination IP addresses), `ip.proto` (IP protocol type), `ip.len` (IP packet length), `tcp.len` (TCP segment length), `tcp.srcport` and `tcp.dstport` (source and destination TCP port numbers), `Value` (a custom value), and `normality` (normality indicator). This dataset provides comprehensive details necessary for network performance monitoring, security threat detection, and traffic pattern analysis.

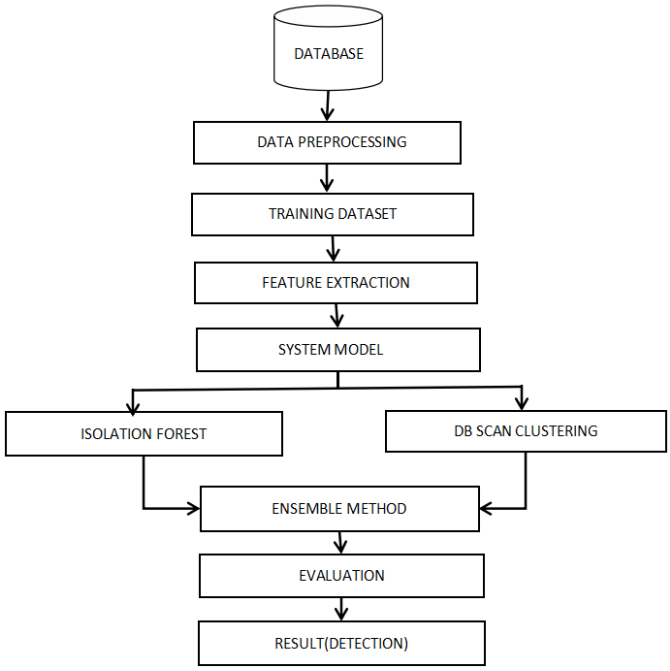

Fig. 1. Architecture Diagram

B. Frame Work

As shown in the fig:1. The anomaly detection in network traffic data begins with storing raw data in a database. This data is then preprocessed to clean and format it, resulting in a training dataset. Key features are extracted from this dataset to train the system model. Two techniques, Isolation Forest and DB Scan Clustering, are applied to the model for anomaly detection. The ensemble method combines these two algorithms using boosting to enhance accuracy. The combined model is evaluated on a separate dataset to ensure its effectiveness. Finally, anomalies in the network traffic data are detected and reported based on the evaluated model.

C. Data Preprocessing

Real-world data often contains noise, missing values, and unusable formats, making it unsuitable for direct use in machine learning models. Data preprocessing, which enhances model accuracy and efficiency, involves several steps: importing necessary libraries, loading the data, and checking for missing values. If missing values are found, they can be handled by either removing the affected rows or estimating the values using mean, median, or mode. The data is then arranged in numerical form and converted into shorter ranges.

D. Feature Selection

Feature extraction for anomaly network traffic detection using wrapper methods involves selecting relevant features that improve model performance. This approach begins with collecting and preprocessing network traffic data, including cleaning, encoding categorical features, and normalizing numerical features. A suitable machine learning model, such as a

decision tree, is then selected. Recursive Feature Elimination (RFE), a wrapper method, is used to iteratively train the model, removing the least important features until the optimal subset is found. This process ensures that the selected features contribute significantly to the model's accuracy. The refined feature set is then used to train the final model, which is evaluated for its performance, leading to improved detection of network traffic anomalies.

E. Isolation Forest

Isolation Forest is a robust unsupervised machine learning algorithm widely used for anomaly detection in wireless networks. Unlike traditional methods that model normal behavior, it directly isolates anomalies by recursively partitioning the data with randomly chosen features and split values. This approach makes it highly efficient in detecting rare anomalies. In wireless networks, where anomalies can indicate security breaches, performance issues, or unexpected behavior, Isolation Forest excels by swiftly identifying and isolating these irregularities. Its ability to handle high-dimensional data and minimal assumptions about the data distribution make it particularly suited for maintaining the security and reliability of network infrastructures.

F. DB Scan Clustering

DB SCAN (**Density-Based Spatial Clustering** is a popular clustering algorithm that can also be effectively utilized for anomaly detection in wireless networks. Unlike traditional clustering algorithms, DBSCAN identifies clusters based on the density of data points rather than predefined shapes. This makes it robust in detecting outliers, which often indicate anomalies in wireless network behavior. By defining parameters such as the minimum points in a neighborhood and a maximum distance threshold, DBSCAN efficiently distinguishes between normal network behavior and anomalous activities. In wireless networks, anomalies may manifest as sudden spikes in traffic, unusual patterns in signal strength, or unexpected device behaviors. DBSCAN excels in identifying such anomalies by isolating data points that do not conform to the established clusters, providing valuable insights into potential security breaches, performance issues, and other irregularities. Compared to other clustering algorithms, DBSCAN does not require the number of clusters to be specified beforehand and can identify clusters of arbitrary shapes, making it particularly effective for the dynamic and unpredictable nature of wireless network data.

G. Ensemble Methods

Ensemble methodin ML involve the integration of multiple base models and algorithms to achieve the best possible outcome for a particular task. The principle behind these methods is to harness the advantages of different models to create a more accurate and reliable overall model, surpassing the performance of any individual model. Various popular ensemble methods exist, each employing a unique approach to model combination.

VIII. Result

The analysis of the ensemble method of classification method(Isolation Forest) and clustering method (DB scan clustering) produced a result as shown in the below graph:

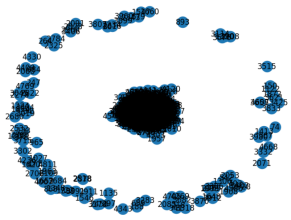

Fig. 2. Graphical representation of data points

The fig:2. displays graphical results for the anomaly detection of 5000 data points . the graph is plotting using DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm for anomaly detection. DBSCAN groups closely packed points into clusters and labels isolated points in low-density areas as anomalies. The graph features a central dense cluster surrounded by less dense clusters forming a shape resembling a smiley face, with scattered points representing noise or outliers. This visualization demonstrates DBSCAN's ability to identify dense clusters and distinguish anomalies within a dataset.

Tabular representation of result containing the details of accuracy, precision, F1 score , recall, specificity and error rate . The result generated as shown in the below table:

PERFORMANCE METRICS	ENSEMBLE METHOD [ISOLATION FOREST+DB SCANCLUSTERING]
ACCURACY	**1.0** [100%]
PRECISION	**1.0** [100%]
RECALL	**1.0** [100%]
F1 SCORE	**1.0** [100%]
SPECIFICITY	**1.0** [100%]
ERROR RATE	**0.00** [0%]

Fig. 3. Tabular representation of result

The fig. 3. represents the performance metrics of an ensemble method combining Isolation Forest and DBSCAN clustering. The metrics include accuracy, precision, recall, F1 score, specificity, and error rate. The results indicate perfect performance, with all metrics showing a value of 1.0 (100%) and an error rate of 0.00% (0%). This suggests the method is highly effective in identifying anomalies.

A. Equations

The equations we used in our project are as follows :

$$\text{Accuracy} = \frac{TN + TP}{TN + FP + TP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$F1 \text{ Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{5}$$

$$\text{Classification Error Rate} = \frac{FP + FN}{TP + FP + TN + FN} \tag{6}$$

DISSCUSSION

This project demonstrates effective anomaly detection in wireless networks using feature extraction, preprocessing, and an ensemble of DBSCAN and Isolation Forest algorithms. This combination achieves high accuracy, precision, recall, F1 score, and specificity, with a low error rate. Ensuring high precision and recall minimizes false alarms and correctly identifies anomalies, while balanced specificity and sensitivity reduce false positives and negatives. The security and efficiency of anomaly detection are enhanced through machine learning algorithms, effective preprocessing, and feature selection. Ensemble methods improve robustness, while scalability, real-time processing, advanced IDS, and continuous monitoring with updates ensure robust network security. The accuracy and other performance metrics of the existing model were low, so this model improved the efficiency and performance of the anomaly detection.

CONCLUSION

By using the ensemble method for both classification and clustering algorithms, we can detect anomalies in wireless network systems. When solving any problem using classification and clustering, the dataset plays a crucial role. The project successfully demonstrated by machine learning algorithms for anomaly detection in network systems. The ensemble methods proved to be highly effective, with robust performance metrics which gave highest performance metrics values . The results highlight the importance of combining different techniques to leverage their individual strengths, leading to more accurate and reliable anomaly detection. Future improvements could include hyperparameter tuning, additional ensemble techniques, and incorporating more features or external data sources for better detection accuracy.

ACKNOWLEDGEMENT

We thank Mrs. **D. Deepika**- Assistant Professor(B V Raju Institute of Technology) for guiding and helping us by providing useful content that helped in performing our project.

REFERENCES

[1] Prathamesh Kulkarni, Himanshu Samariya,Akash Sitoke,Aman Chandre,Prof. Sagar Dhanake, al. "Anomaly Detection in Network Traffic Using Unsupervised Machine Learning Approach." IJIREEICE, ijireeice.com/papers/anomaly-detection-in-network-traffic-using- unsupervised-machine-learning-approach/. 2022

[2] Dehlaghi-Ghadim, Alireza, Mahshid Helali Moghadam, Ali Balador, and Hans Hansson. "AnomalyDetection Dataset for Industrial Control Systems." arXiv, May 11, 2023. http://arxiv.org/abs/2305.09678.

[3] Li, Xiang, Jing Zhang, Yali Yuan, and Cangqi Zhou. "Network Intrusion Detection with Edge-Directed GraphMulti-Head Attention Networks." arXiv,2023. http://arxiv.org/abs/2310.17348

[4] Zehra, Sehar, Ummay Faseeha, Hassan Jamil Syed, Fahad Samad, Ashraf Osman Ibrahim, Anas W. Abulfaraj, and Wamda Nag- meldin. "Machine Learning-Based Anomaly Detection in NFV: AComprehensive Survey." Sensors 23, no. 11 (June 5, 2023): 5340. https://doi.org/10.3390/s23115340

[5] Kim, Jinoh, Makiya Nakashima, Wenjun Fan, Simeon Wuthier, Xi aobo Zhou, Ikkyun Kim, and Sang-YoonChang. "A Machine Learn ing Approach to Anomaly Detection Based on Traffic Monitor ing for SecureBlockchain Networking." IEEE Transactions on Net work and Service Management 19, no. 3(September 2022): 3619–32. https://doi.org/10.1109/TNSM.2022.3173598

[6] Thwaini, Mohammed Hussein. "Anomaly Detection in Network Traffic Using Machine Learning for EarlyThreat Detection." Data and Metadata 1 (December 23, 2022): 34. https://doi.org/10.56294/dm202272..

[7] Wang, Song, Juan Fernando Balarezo, Sithamparanathan Kandeepan, Akram Al-Hourani, Karina GomezChavez, and Benjamin Rubinstein. "Machine Learning in Network Anomaly Detection: A Survey."IEEE Access 9 (2021): 152379–96. https://doi.org/10.1109/ACCESS.2021.3126834

[8] Ma, Qian, Cong Sun, and Baojiang Cui. "A Novel Model for Anomaly Detection in Network Traffic Basedon Support Vector Machine and Clustering." Edited by Marimuthu Karuppiah. Secu rity andCommunication Networks 2021 (November 20, 2021): 1–11. https://doi.org/10.1155/2021/2170788.

[9] Iivari, Albin. "ANOMALY DETECTION TECHNIQUES FOR UNSU- PERVISED MACHINELEARNING," n.d.

[10] Vikram, Aditya and Mohana. "Anomaly Detection in Network Traffic Using Unsupervised Machine LearningApproach." In 2020 5th International Conference on Communication and Electron ics Systems (ICCES),476–79. Coimbatore, India: IEEE, 2020. https://doi.org/10.1109/ICCES48766.2020.9137987

[11] Rana, Samir. "Anomaly Detection in Network Traffic Using Machine Learning and Deep LearningTechniques." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 10, no. 2(September 10, 2019): 1063–67 https://doi.org/10.17762/turcomat.v10i2.13626.

[12] Suman, Chanchal, Somanath Tripathy, and Sriparna Saha. “Building an Effective Intrusion Detection SystemUsing Unsupervised Feature Selection in Multi-Objective Optimization Framework.” arXiv, May 16,2019.http://arxiv.org/abs/1905.06562.

[13] Mazel, Johan. “Unsupervised Network Anomaly Detection” .

[14] Almomani I, Al-Kasasbeh B and Al-Akhras M 2016 ”Journal of Sensors”.

[15] Syarif, Iwan, Adam Prugel-Bennett, and Gary Wills. “Unsupervised Clustering Approach for NetworkAnomaly Detection.” In Networked Digital Technologies, edited by Rachid Benlamri, 293:135–45.Com munications in Computer and Information Science. Berlin, Heidel berg: Springer Berlin Heidelberg,2012 https://doi.org/10.1007/978-3- 642-30507-8-13.

[16] Simmross-Wattenberg, Federico, Juan Ignacio Asensio-Perez, PabloCasaseca-de-la-Higuera, Marcos MartinFernandez, Ioannis A Dimitri-adis, and Carlos Alberola-Lopez. “Anomaly Detection in NetworkTrafficBased on Statistical Inference and alpha-Stable Modeling.” IEEE Transactions on Dependable andSecure Computing 8, no. 4 (July 2011):494–509 https://doi.org/10.1109/TDSC.2011.14.