International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

An Ensamble Model for Detecting Intranet Threats and Possible Attacks

Sabitha P
Assistant Professor,
Department of Computer Science and Engineering
SRM Institute of Science and Technology, Ramapuram, Chennai.
sabithap@srmist.edu.in

I Poshya Kumar
UG Student,
Department of Computer Science and Engineering
SRM Institute of Science and Technology, Ramapuram, Chennai
ii8221@srmist.edu.in

Sanjay.M
UG Student,
Department of Computer Science and Engineering
SRM Institute of Science and Technology, Ramapuram, Chennai.
mm3798@srmist.edu.in

Vikram Rakshit U
UG Student,
Department of Computer Science and Engineering
SRM Institute of Science and Technology, Ramapuram, Chennai.
up4259@srmist.edu.in

Abstract- Detecting intranet attacks is tough because attackers keep changing their methods. This paper suggests a new way to catch these attacks using machine learning. The idea is to study how these attacks behave and use that knowledge to spot them early. By looking at things like network traffic and system logs, the system learns what's normal and what's not. This helps it flag any strange activity before it becomes a serious threat. The approach aims to boost intranet security by offering real-time detection and flexible ways to defend against attacks.

I. INTRODUCTION

Securing the intranet is vital for modern organizational networks, where internal threats are often underestimated due to their evolving and low-key nature. Traditional rule-based intrusion detection systems often fall short against sophisticated cyberattacks such as DDoS, brute-force attempts, phishing campaigns, malware infections, and insider threats.. A more dynamic and intelligent strategy of threat discovery is needed for this.

In this study, we introduce a machine learning-based IDS that utilizes an ensemble learning strategy, combining Support Vector Classifier (SVC), Gradient Boosting (GB), and Random Forest (RF). Our objective, unlike traditional approaches, not only classifies threats but also estimates the probability of each attack type, enabling risk-based threat assessment instead of strict classification.The system analyzes network traffic data through the extraction of significant behavioural attributes and the utilization of supervised as well as unsupervised learning methods. The two-layered approach allows real-time anomaly identification with high-confidence classification. Additionally, our methodology presents an adaptive threat intensity score system, dividing attacks into low, medium, and high-threat levels, permitting security teams to effectively prioritize threats.

Our research seeks to bridge the gap between signature-based detection and a threat detection, providing a scalable and solid solution that further secures intranets. The system suggested here offers real-time detection, probabilistic threat assessment, and an automated response system, hence a better solution compared to classical IDS models.

II. LITERATURE SURVEY

Jang and Lee (2024) had proposed an advanced behaviour-based system for detecting intranet attacks by using feature engineering on Zeek IDS logs. They used six supervised machine learning algorithms to experiment on improving attack detection accuracy. Similarly, Zhao (2024) created a system for abnormal traffic detection for intranets using a whitelist-based approach, ensuring secure network traffic monitoring.

Yang and colleagues (2022) carried out a through review of different anomal-based methods for network intrusion detection,along with tyhe datasets used,pinpointing trends and shortcomings in current security models,In a fresh take,jiao and team (2024) presented a deep transfer learning model that transforms user behaviour into image and utilizes oversampling techniques such as SMOTE to enhance the detection of intranet threats.

Wang et al. (2022) took a deep dive into how reinforcement learning can enhance cybersecurity,creating a framework aimed at making smart decisions for cyber defence.Meanwhile,Chen and their team (2022) focused on modelling cyber attacks,using behaviour trees and ant colony optimization (AOC) to identify possible attack routes within a network.

Liu et al. (2018) introduced MaliceScript, a threat detection system that operates right in your browser and represents user activity as grayscale images.By using a deep learning approach,it effectively spots malicious behaviour in web environments.Meanwhile,Chen and their team(2020) turned their attention to analyzing Linux logs, employing supervised machine learning techniques to identify advanced persistent threats (APts) and differentiate between normal and harmful activities.

Wang et al. (2018) took a close look at the operating data from enterprise intranets,including logs and security alerts,to identify possible risks and vulnerabilities.Their research highlighted how crucial it is to use data-driven approaches for evaluating risks in enterprise security.

These studies come together to showcase the progress we've made in areas like anomaly detection,behavioural modelling,and the use of machine learning in intranet security.This sets the stage for better intrusion detection systems moving forward.

Building on the groundwork laid by earlier studies in intranet security, recent research has brought forth some innovative methods to boost intrusion detection systems (IDS). For instance, Ansari et al. (2022) introduced a deep learning model based on "Gated Recurrent Units (GRU)" that can effectively learn the patterns in security alert sequences, which in turn enhances the accuracy of predicting network intrusion alerts.

In a similar vein, Talukder and colleagues (2024) created a "machine learning-driven network intrusion detection system specifically designed for handling large and imbalanced datasets". Their method cleverly combines oversampling techniques, stacking feature embeddings, and feature extraction to improve the identification of those rare attack patterns.

Uppal et al. (2025) tackled the tough issues of intrusion detection and privacy protection in smart city networks by introducing a machine learning framework that uses an ensemble approach. This innovative framework combines several classifiers to boost detection accuracy while ensuring that data privacy is upheld.

Sohail et al. (2023) came up with a cutting-edge meta-learning approach that uses deep neural network for network intrusion detection. They really honed in on how adaptable intrusion detection systems (IDS) can be when it comes to tackling new and evolving threats.Their model significantly boosts the system's capability to spot attack vectors that it hasn't encountered befor.

Gyimah et al. (2024) took a deep dive into how automated machine learning (AutoML) can be used for network intrusion detection. Their research showed that autoML is quite capable of picking and fine-tuning manual adjustments we often see.

Amouri et al. (2024) iintroduced a cutting-edge ensamble method that leverages kolmogorov-Arnold networks to improve intrusion detection in IOT settings.This innovative approach tackles the specific security challanges that come with IoT devices.

Bibers et al. (2024) carried out an in-depth comparative study on different machine learnig models and ensamble strategies specifically for network intrusion detection systems.Their research sheds light on the various strengths and weakness of these approaches,offering valuable insights into the field.

Khan et al. (2021) introduced an innovative deep learning framework designed for network intrusion detection.This approach cleverly combines convolutional and recurrent neural networks to effectively capture both the spatial and temporal characteristics of network traffic.

Shone et al. (2018) explored the use of deep learning methods for network intrusion detection,showcasing how unsupervised feature learning can effectively identify intricate attack patterns.

Kim et al. (2016) used Long Short-Term Memory (LSTM) recurrent neural networks to tackle intrusion detection,showcasing how well the model can grasp the temporal patterns in network traffic data.

Vinayakumar et al. (2017) explored deep learning techniques for predicting network traffic.This is crucial for staying ahead for potential intrusions and effectively managing networks.

Yuan et al. (2017) introduced DeepDefense, a system that uses deep learning to identify Distributed Denial of Service (DDoS) attacks, showcasing the effectiveness of deep neural networks in detecting complex attack patterns.

These studies together showcase the continuous progress being made using these multiple types of algorithms and novel implementations that paved the way for the sope to create our instigation

III. METHODOLOGY

The Intrusion Detection System (IDS) follows a clear and step-by-step method. It first checks the network, firewall, and system logs to collect important data. After that, it finds the key features from that data. Then, it uses an ensemble model to detect threats. The same data is looked at again, and the system carefully checks for anything that could be dangerous. This process helps in sorting out safe actions from harmful ones by using a group of models instead of just one.The dataset For training and evaluating our intranet security focused intrusion detection system (IDS), we use the CICIDS 2017 dataset, developed by the Canadian Institute for Cybersecurity.

The Key Features of the Dataset are 80+ key_flow features, basic features- source and IP, port numbers, protocol type- traffic-based features- total number of packets, number of bytes, flow duration, time-based data- inter-arrival times, timestamp-based metrics mean, variance, standard deviation of all flow metrics

Data Preprocessing is Necessary for cleaning and preparing raw data beforeFeature extraction and model training. This step includes practices such as

DataCleaning (Erasing duplicate or irrelevant records), Filtering (Reducing noise by removing log entries which are not useful eve at debugging time), Data Augmentation (Oversampling or synthetic data generation for preparing balance dataset), Normalization (Scale numeric features to a uniform range to maintain uniformity between various models), Sorting (Organizing logs chronologically for sequential analysis.) The data is separated into training and test data once it has been pre-processed.

Model-Specific Processing and possibly more preprocessing on per model base, and choices of models are SVM, as it seeks to create the optimal hyperplane. For classification, it is necessary to normalize the data to ensure that features have the same scale. Gradient Boosting This is generally less sensitive to feature scaling but may require appropriate handling of missing values and outliers for robustness. Random Forest as the third model is a collective of decision trees, the decision tree handles unscaled data and we know it is relatively robust to outliers. However, more performant results can be achieved if categorical variables are well encoded.

However, in the case of our IDS, we have purposefully combined several models to scrutinize networks and detect intruders on them. This ensemble learning model merges the individual classifiers to combine their individual strengths for improved detection performance and resilience.

Support Vector Classifiers (SVCs) are used to determine the best hyperplane that divides the various classes in the dataset. SVMs use kernel functions to map input data points to more complex places, capture linearly complex data relationships, thus enabling accurate classification of normal and anomalous network activities.

Gradient Boosting combines weak learners sequentially so as to improve on the errors of prior models. Through this cycle, the model is explicitly made aware of cases that it struggles to classify correctly, so the model's performance improves overall. Each class is assigned a probability score based on predictions made from individual models, which are combined cumulatively to provide a more accurate detection of the threat.

Random Forests help us instill multiple decision trees based on arbitary subsets of features and data samples to form the ensemble. They work independently, and their outputs are aggregated through majority voting or averaging. This approach is useful for improved generalization and

prevention of overfitting, making the intrusion detection more reliable.

The use of such diverse models enables our IDS to unravel a more detailed picture of the network traffic to identify and retaliate to threats with better accuracy and less false positive rate

IV. IMPLEMENTATION & RESULTS

In this endeavor we suggested a system to detect intrusion based on three supervised machine learning models such as Support Vector classifier (SVC), Gradient Boosting and Random Forest. Data processing was optimized for performance of each model employed.

i. Support Vector Classifier (SVC):

Feature Scaling: The features were meant no unit variance to ensure that all input variables had the same scale and contributed equally to the model.

Parameters:

Kernel: Radial Basis Function (RBF)

Regularization parameter (c) : 1.0

kernel coeff (gamma)–penalty (C): 'scale'=

Sklearn SVC inpy:

The SVC class from sklearn uses decision_function to find the optimal hyperplane The decision function is:

$$f(x)= f(x)=(\sum i=1n\alpha iyiK(xi,x)+b)$$

αi is for the Lagrange multipliers.

The class labels are yi

- $K(xi,x)$ is the kernel f(n),

such as the RBF kernel

$$K(x_i,x)=exp(-\gamma \|x_i-x\|^2).$$

ii. Gradient boosting

Data Preprocessing — As we used tree based methods, we did not need to apply scaling.

Parameters:Number of estimators: 100

Learning rate: 0.1

Maximum depth of trees: 3

Model Update: Gradient Boosting produces a sequential ensaamble of trees , where each new tree hm(x) is fitted to the – ve gradient (residuals) of loss function L with respect to the current model

$$F_m(x)=F_{m-1}(x)+v \cdot h_m(x)$$

Fm(x) is the updated model at iteration m, ν is the learning rate.

hm(x) is the new base learner fitted to residuals.

Ensemble Prediction: Random Forest aggregates the predictions from multiple decision trees $h_m(x)$ to make a final prediction:

$$\hat{y} = \frac{1}{M}\sum_{m=1}^{M} hm(x)$$

where:

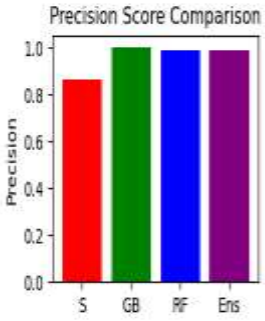
Mis the total number of trees.

$h_m(x)$is the prediction from the m-th tree. The model's performance analysis reveals its efficacy in predicting threat levels with high accuracy. The confusion matrix indicates that Class 0 (Benign) had 171 correct predictions, with only 2 instances misclassified as Class 1. Class 1 (Attack) saw 371 accurate predictions, and just 2 misclassifications into Class 2. For Class 2 (Severe Attack), there were 50 correct predictions, with 4 instances misclassified as Class 1. This results in an overall accuracy of 99%, highlighting the model's

precision in classification tasks.

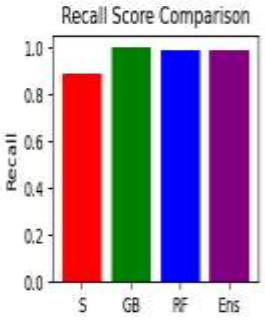
```
[162]: plt.subplot(2, 2, 2)
       plt.bar(models, precision_scores, color=['red', 'green', 'blue', 'purple'])
       plt.title('Precision Score Comparison')
       plt.ylabel('Precision')

[162]: Text(0, 0.5, 'Precision')
```

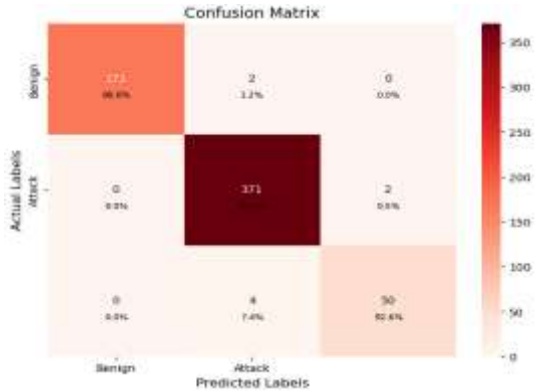



```
[163]: plt.subplot(2, 2, 3)
       plt.bar(models, recall_scores, color=['red', 'green', 'blue', 'purple'])
       plt.title('Recall Score Comparison')
       plt.ylabel('Recall')

[163]: Text(0, 0.5, 'Recall')
```

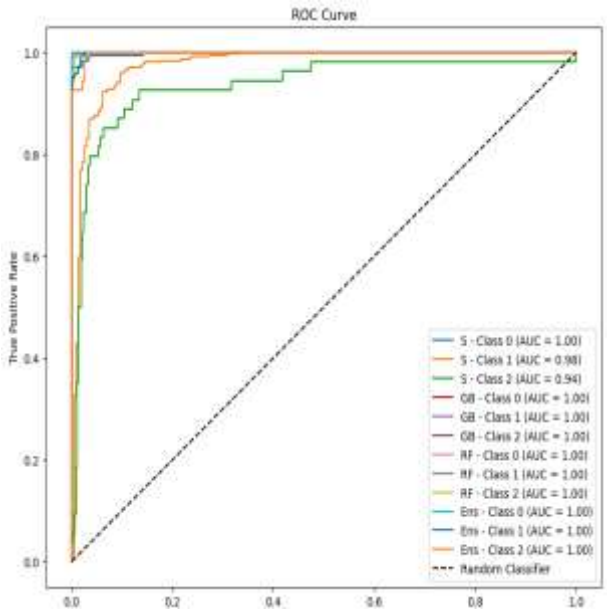


The report of the classification further underscores the model's robustness. Precision stands at 100% for Class 0, 98% for Class 1, and 96% for Class 2. Recall rates are 99% for Classes 0 and 1, and 93% for Class 2. The F1-Scores are 99% for Classes 0 and 1, and 94% for Class 2. Both weighted and macro averages range between 98–99%, indicating a balanced and effective performance across all classes.



When comparing F1 scores among different models, Gradient Boosting , Random Forest, and the Ensemble method exhibit the highest scores, nearing 1.00. The Support Vector Classifier (SVC) shows a slightly lower F1 score, approximately 0.85, yet it remains commendable. This comparison suggests that while all models perform admirably, GB, RF, and the Ensemble method offer a slight edge in terms of F1 score.

In summary, the model demonstrates exceptional proficiency in threat level classification, with minimal misclassifications and precision, recalls and f1 scores turn high. The comparative analysis indicates that GB, RF, and Ensemble models slightly outperform the SVC, making them preferable choices for this specific classification task.



V. CONCLUSION

The experimental results clearly show that all three classifiers—SVC, GB, and RF—perform exceptionally well in identifying threat levels. However, Gradient Boosting and Random Forest, especially when combined in an ensemble model, achieve superior results with nearly perfect classification metrics. The ensemble approach particularly excels in precision, recall, F1 score, and ROC-AUC values, ensuring robust and reliable threat detection with minimal false positives or negatives.

VI. REFERENCES

[1] M. Jang and K. Lee, "An Advanced Approach for Detecting Behavior-Based Intranet Attacks by Machine Learning," *IEEE Access*, vol. 12, pp. 52480-52495, 2024, doi: 10.1109/ACCESS.2024.3387016.

[2] A. Zhao, "Design and Implementation of Intranet Abnormal Traffic Security Detection System," *2024 2nd International Conference on Mechatronics, IoT and Industrial Informatics (ICMIII)*, Melbourne, Australia, 2024, pp. 782-785, doi: 10.1109/ICMIII62623.2024.00152.

[3] Z. Yang, X. Liu, T. Li, D. Wu, J. Wang, Y. Zhao, and H. Han, "A Systematic Literature Review of Methods and Datasets for Anomaly-Based Network Intrusion Detection," *Computers & Security*, vol. 116, 2022, Art. no. 102675, ISSN 0167-4048, doi: 10.1016/j.cose.2022.102675.

[4] J. Jiao, Z. Liu, and L. Li, "Intranet Security Detection Based on Image and Deep Transfer Learning," in *Proceedings of the 2023 13th International Conference on Communication and Network Security (ICCNS '23)*, ACM, New York, NY, USA, 2024, pp. 196–202, doi: 10.1145/3638782.3638812.

[5] W. Wang, D. Sun, F. Jiang, X. Chen, and C. Zhu, "Research and Challenges of Reinforcement Learning in Cyber Defense Decision-Making for Intranet Security," *Algorithms*, vol. 15, no. 4, p. 134, 2022, doi: 10.3390/a15040134.

[6] T. Chen et al., "Research on Cyber Attack Modeling and Attack Path Discovery," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1850-1864, 2021, doi: 10.1109/TIFS.2021.3125637

[7] C. Liu et al., "MaliceScript: A Novel Browser-Based Intranet Threat," *IEEE Transactions on Dependable and Secure Computing*, vol. 17, no. 6, pp. 1270-1285, 2018, doi: 10.1109/TDSC.2018.2827041.

[8] L. Chen et al., "Detecting Advanced Attacks Based on Linux Logs," *IEEE Access*, vol. 8, pp. 123456-123470, 2020, doi: 10.1109/ACCESS.2020.3012376.

[9] H. Wang et al., "Research on Operating Data Analysis for Enterprise Intranet Information Security Risk Assessment," *IEEE Access*, vol. 7, pp. 102400-102415, 2019, doi: 10.1109/ACCESS.2019.2938758.

[10] M. S. Ansari, V. Bartoš, and B. Lee, "GRU-based Deep Learning Approach for Network Intrusion Alert Prediction," *Future Generation Computer Systems*, vol. 128, pp. 235–247, 2022, doi: 10.1016/j.future.2021.09.040.

[11] M. A. Talukder et al., "Machine Learning-Based Network Intrusion Detection for Big and Imbalanced Data Using Oversampling, Stacking Feature Embedding, and Feature Extraction," *Journal of Big Data*, vol. 11, no. 33, 2024, doi: 10.1186/s40537-024-00886-w.

[12] M. Uppal et al., "Enhancing Accuracy through Ensemble-Based Machine Learning for Intrusion Detection and Privacy Preservation over the Network of Smart Cities," *Discover Internet of Things*, vol. 5, article number 11, 2025, doi: 10.1007/s43926-025-00101-z.

[13] A. Sohail et al., "Deep Neural Networks Based Meta-Learning for Network Intrusion Detection," *arXiv preprint arXiv:2302.09394*, 2023.

[14] N. K. Gyimah et al., "An AutoML-Based Approach for Network Intrusion Detection," *arXiv preprint arXiv:2411.15920*, 2024.

[15] A. Amouri et al., "Enhancing Intrusion Detection in IoT Environments: An Advanced Ensemble Approach Using Kolmogorov-Arnold Networks," *arXiv preprint arXiv:2408.15886*, 2024.

[16] I. Bibers, O. Arreche, and M. Abdallah, "A Comprehensive Comparative Study of Individual ML Models and Ensemble Strategies for Network Intrusion Detection Systems," *arXiv preprint arXiv:2410.15597*, 2024.

[17] M. A. Khan et al., "A Novel Deep Learning-Based Framework for Network Intrusion

Detection," *IEEE Access*, vol. 9, pp. 123565-123574, 2021, doi: 10.1109/ACCESS.2021.3109669.

[18] S. Shone et al., "A Deep Learning Approach to Network Intrusion Detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41-50, 2018, doi: 10.1109/TETCI.2017.2772792.

[19] Y. Kim et al., "Long Short-Term Memory Recurrent Neural Network Classifier for Intrusion Detection," *2016 International Conference on Platform Technology and Service (PlatCon)*, Jeju, South Korea, 2016, pp. 1-5, doi: 10.1109/PlatCon.2016.7456805.

[20] R. Vinayakumar et al., "Applying Deep Learning Approaches for Network Traffic Prediction," *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Udupi, India, 2017, pp. 2353-2358, doi: 10.1109/ICACCI.2017.8126164.

[21] X. Yuan et al., "DeepDefense: Identifying DDoS Attack via Deep Learning," *2017 IEEE International Conference on Smart Computing (SMARTCOMP)*, Hong Kong, China, 2017, pp. 1-8, doi: 10.1109/SMARTCOMP.2017.7946998.