IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Intrusion Detection System Using Pca With Random Forest Approach

[1] Dr. M Parameswari, [2] D Kanimozhi ,[3] S Karthika, [4] C Madhumitha, [5] K Madhumitha
[1] Associate Professor, [2]Student, [3] Student, [4] Student, [5] Student
Computer Science and Engineering,
Kings Engineering College, Chennai, India

ABSTRACT: The aim of this project helps to develop an application to find the type of attack that occurred on the system and to detect the intruders by using Intrusion Detection System. Previously various machine learning (ML) techniques are applied on the IDS and tried to improve the results on the detection of intruders and to increase the accuracy of the IDS. This paper has proposed an approach to develop efficient IDS by using the principal component analysis (PCA) and the random forest classification algorithm. Where the PCA will help to organise the dataset by reducing the dimensionality of the dataset and the random forest will help in classification. Results obtained states that the proposed approach works more efficiently in terms of accuracy as compared to other techniques like SVM, Naive Bayes, and Decision Tree. The IDS acts as a network level defence to secure a system. IDS mainly used for security purpose to find the threats or malicious activities and also for identifying the type of attack on the system.

Keywords: IDS-intrusion detection system, Dimensionality, Datasets, Intruders, PCA-principal component analysis, Accuracy, RFA-random forest approach, Attack, Detection, classification.

INTRODUCTION

An intrusion detection system can be defined as a security mechanism designed to detect and respond to unauthorized access or malicious activities within a system. Unlike traditional security measures that focus primarily on preventing external attacks, IDSs operate on the principle of continuous monitoring and analysis of system logs and other relevant data sources to identify potential security breaches. The intrusion detection system works for the improvement of the system, which is affected by the intruders. This system can do the detection of the intruders. The proposed system tries to eliminate theexisting problems related to the previous work. The proposed system consists of the two methods that are principal component analysis, and the other one is the random forest. The principal component analysis is used for the reduction of the dimension of the dataset; by thismethod, the dataset quality will be improved as the dataset may contain the correct attributes. After this, the random forest algorithm will be applied for the detection of the intruders, which provide both the detection rate and the false alarm rate in an improved manner as compared to SVM.PCA helps reduce the dimensionality of the data by identifying and retaining the most relevant features. This can lead to faster training and inference times and more efficient memory usage, especially when dealing with large datasets.PCA can highlight the most important features in the data, which can enhance the model's accuracy by focusing on the critical variables for intrusion detection. The proposed system leverages PCA for feature extraction and dimensionality reduction. By transforming the original high-dimensional feature space into a lower-dimensional subspace while preserving the essential information, PCA helps in mitigating the curse of dimensionality. This reduction not only improves computational efficiency but also enhances the effectiveness of subsequent classification algorithms. Random Forest algorithm is employed for intrusion detection due to its robustness, scalability, and ability to handle high-dimensional data. By constructing an ensemble of decision trees and aggregating their predictions, this research investigates the application of Principal Component Analysis (PCA) and Random Forest algorithm for intrusion detection in computer networks. PCA is employed to reduce the dimensionality of network traffic data, followed by classification using Random Forest. The study evaluates the performance of the proposed approach in terms of accuracy and computational efficiency.

Demonstrating its efficacy in detecting intrusions effectively Random Forest improves classification accuracy and generalization performance. Moreover, its inherent capability to handle imbalanced datasets and noisy features makes it well-suited for real-world intrusion detection scenarios. The integration of PCA with Random Forest facilitates more accurate and efficient detection of intrusions compared to conventional methods. By reducing the dimensionality of the dataset and capturing the underlying patterns within the data, the proposed approach enhances the discriminative power of the classifier. This leads to higher detection rates while minimizing false positives, thereby improving the overall performance of the IDS.

Title	Authors	Year	Methodology	Merits & Demerits
A Comparative Study of Machine Learning Techniques for Intrusion Detection Systems	Deepak Kumar, Girish Kumar, and Sanjay Kumar	2018	Support Vector Machine (SVM) and Naïve Bayes	Evaluation based on their accuracy, detection rate, and false alarm rate using benchmark datasets.
A Hybrid Approach for Intrusion Detection Using PCA	Nisreen I. Amawi and Sahar F. Sabri	2020	PCA- Principal component analysis	Reducing the dimensionality of network traffic features and false positives.
Enhancing IDS Performance Using Random Forest Approach	Wei Lu, Guangquan Xu, and Guangquan Zhang	2021	RFA and ANN (Artificial Neural Networks)	Improving accuracy compared to traditional methods.

Table 1. Comparison

II LITERATURE SURVEY

A literature survey on Intrusion Detection System using PCA with Random Forest Approach reveals that there is an advanced technique for enhancement in terms of accuracy, efficiency, and robustness. One relevant study published in the international journal of A Hybrid Approach for Intrusion Detection Using PCA and Random Forest in which the study evaluates the proposed approach on benchmark datasets, demonstrating its effectiveness in improving intrusion detection accuracy and reducing false positives. This paper presents a hybrid approach for intrusion detection by combining Principal Component Analysis (PCA) and Random Forest algorithm. PCA is utilized to reduce the dimensionality of network traffic features, while Random Forest is employed for classification. In other study Enhancing Intrusion Detection System Performance explains about theExperimental results demonstrate the effectiveness of the proposed approach in improving the accuracy and efficiency of intrusion detection compared to traditional methods. The third study published namely Intrusion Detection in Computer Networks Based on Principal Component Analysis. In this approach there is a drawback in understanding the nature of intrusions. Overfitting models might not generalize well to new unseen data is an inefficient advantage. In real-world intrusion detection, there is often a class imbalance between normal and intrusive events. This can pose a challenge when training a machine learning model, as it may lead to biased results.

III EXISTING PROBLEM

The systems which work over the internet suffer from various malicious activities. The major problem seen in this field is the intrusion in the system for violating the information. This intrusion is detected by creating an intrusion detection system; this system also needs to be accurate and efficient in the detection of the intruders. Various machine learning algorithms were used for intrusion detection; some of them are SVM, Naïve Bayes etc. But the results state that there may be some improvements to be done on terms of accuracy and the detection rates and the false alarm rate. The earliest intrusion detection systems developed were majorly signature based, that is, the detection of malicious activities depends on the pre-defined and configured known signatures of known attacks. The main drawback of this IDS is loss of Interpretability. PCA reduces data dimensionality but can also make it harder to interpret the meaning of individual features, which can be a disadvantage when trying to understand the nature of intrusions. Secondly is of overfitting models. Random Forest can be prone to overfitting if not properly tuned. overfitting models might not generalize well to new or unseen data. Thirdly is all about data imbalance. In real-world intrusion detection, there is often a class imbalance between normal and intrusive events. This can pose a challenge when training a machine learning model, as it may lead to biased results.

IV PROPOSED SYSTEM

The intrusion detection system works for the improvement of the system, which is affected by the intruders. This system can do the detection of the intruders. The proposed system tries to eliminate the existing problems related to the previous work. The proposed system consists of the two methods that are principal component analysis, and the other one is the random forest. The principal component analysis is used for the reduction of the dimension of the dataset; by this method, the dataset quality will be improved as the dataset may contain the correct attributes. After this, the random forest algorithm will be applied for the detection of the intruders, which provide both the detection rate and the false alarm rate in an improved manner as compared to SVM. The main advantage of this proposed system is dimensionality reduction.PCA helps reduce the dimensionality of the data by identifying and retaining the most relevant features. This can lead to faster training and inference times and more efficient memory usage, especially when dealing with large datasets. And another advantage is Improved feature selection. PCA can highlight the most important features in the data, which can enhance the model's accuracy by focusing on the critical variables for intrusion detection.

V STRATERGICAL TECHNIQUES

While existing intrusion detection systems (IDS) have utilized machine learning algorithms such as Support Vector Machine (SVM), Naïve Bayes, and Decision Trees, there remains room for enhancement in terms of accuracy, efficiency, and robustness. The proposed overcome aims to address these limitations by integrating Principal Component Analysis (PCA) with the Random Forest classification algorithm. This combination offers several advantages over traditional methods and can significantly improve the performance of IDS. The following strategies are proposed to overcome the identified challenges.

Dimensionality Reduction with PCA: The proposed system leverages PCA for feature extraction and dimensionality reduction. By transforming the original high-dimensional feature space into a lower-dimensional subspace while preserving the essential information, PCA helps in mitigating the curse of dimensionality. This reduction not only improves computational efficiency but also enhances the effectiveness of subsequent classification algorithms.

Enhanced Classification with Random Forest: Random Forest algorithm is employed for intrusion detection due to its robustness, scalability, and ability to handle high-dimensional data. By constructing an ensemble of decision trees and aggregating their predictions, Random Forest improves classification accuracy and generalization performance. Moreover, its inherent capability to handle imbalanced datasets and noisy features makes it well-suited for real-world intrusion detection scenarios.

Improved Accuracy and Detection Rates: The integration of PCA with Random Forest facilitates more accurate and efficient detection of intrusions compared to conventional methods. By reducing the dimensionality of the dataset and capturing the underlying patterns within the data, the proposed approach enhances the discriminative power of the classifier. This leads to higher detection rates while minimizing false positives, thereby improving the overall performance of the IDS.

Adaptability to Evolving Threats: The proposed system is designed to adapt to evolving cyber threats and attack patterns. By continuously updating the model with new data and incorporating feedback mechanisms, the IDS remains effective in detecting novel intrusion attempts and emerging security threats. This adaptability ensures the long-term relevance and reliability of the intrusion detection system in dynamic network environments.

Experimental Validation and Benchmarking: To evaluate the effectiveness of the proposed approach, extensive experiments will be conducted using benchmark datasets and real-world network traffic traces. Performance metrics such as accuracy, detection rate, false alarm rate, and computational efficiency will be rigorously analysed and compared against existing methods. The results will provide empirical evidence of the superiority of the proposed overcome in enhancing the security posture of networked systems.

VI DATA FLOW DIAGRAM

The data flow diagram represents the intrusion detection system performance based on the proposed system.

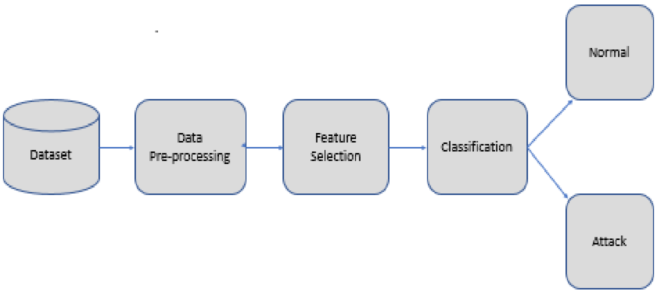

Fig 1: Data Flow Diagram

VII ARCHITECTURAL DIAGRAM

The intrusion detection system mainly uses the principal component analysis and random forest approach for finding the intruders. Here's the representational model explains about the IDS.

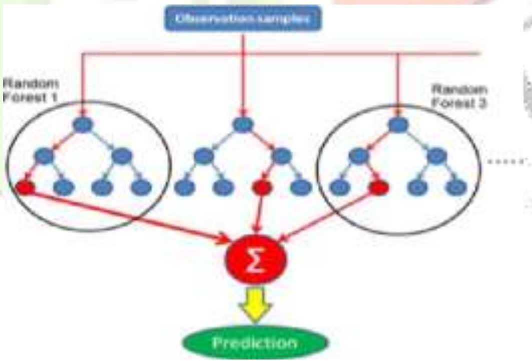

Fig 2: Representational model explains about the IDS

VIII CONCLUSION

The evolution of machine learning mainly uses new techniques for intrusion detection systems in which various types of classifies have been adopted by researchers and scholars in building intrusion detection systems models. This paper presented various research papers related to using machine learning classifiers in intrusion detection systems published from 2018 to 2021.The proposed intrusion detection system utilizing PCA with a Random Forest approach proved to be effective. Among the various models applied in the studied research papers we developed this application by using principal component analysis and random forest approach. Hence it produces a high efficiency rate and accuracy rate in detection.

REFERENCES

[1] A Comparative Study of Machine Learning Techniques for Intrusion Detection Systems by the authors Deepak Kumar, Girish Kumar, and Sanjay Kumar in the year 2018(IEEE).

[2] A Hybrid Approach for Intrusion Detection Using PCA and Random Forest by Nisreen I. Amawi and Sahar F. Sabri in 2020(IEEE).

[3] Enhancing Intrusion Detection System Performance Using Random Forest Approach by Wei Lu, Guangquan Xu, and Guangquan Zhang in 2021(IEEE).

[4] 1.Jafar Abo Nada; Mohammad Rasmi Al-Mosa, 2018 International Arab Conference on Information Technology (ACIT), A Proposed Wireless Intrusion Detection Prevention and Attack System

[5] 2. Kinam Park; Youn Data Computing Service and Applications (BigDataService), Classification of Attack Types for Intrusion Detection Systems Using a Machine Learning Algorithm.

[6] 3. S. Bernard, L. Heutte and S. Adam "On the Selection of Decision Trees in Random Forests" Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009, 978-1-4244-3553- 1/09/$25.00 ©2009 IEEE.