International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

Cyber Attack Detection in Network Utilising Machine Learning Approach

Y.Nagamalleswararao [1],K.Pavani[2],D.Roshini[3]

#1 Assistant Professor in the Department Of MCA, SRK Institute Of Technology,Vijayawada.

#2 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada

Abstract: In contrast to earlier times, advancements in communication and computer technology have brought about significant and all-encompassing changes. People, businesses, and governments all profit greatly from the use of new innovations, even when some of them work against them. For example, information accessibility, security of stored data, and protection of critical information stages. Among these issues, digital despotism based on fear is now one of the most significant. Digital anxiousness, which has caused many issues for people and organizations, has escalated to a point where many occurrences, such as criminal associations, competent persons, and digital activists, might jeopardize open and national security. In this regard, intrusion detection systems (IDS) have been created to strategically avoid cyberattacks.

Index terms - *— Cyber Security, Intrusion Detection System (IDS), Machine Learning, Support Vector Machine (SVM), Decision Tree, Random Forest, CICIDS2017 Dataset, Cyber Terrorism, DoS/DDoS Attack Detection, Network Security.*

1. INTRODUCTION

Digital infrastructure has advanced significantly as a result of the quick development of computer and communication technology. Although these advancements have enormous advantages for people, businesses, and governments, they also present new security risks. Cyberterrorism, which targets private information, interferes with systems, and puts public and national security at risk, is one of the biggest concerns of our day. Hackers, cybercriminals, and hacktivists are examples of malicious actors who take use of system flaws to execute assaults ranging from large-scale denial-of-service (DoS) and distributed denial-of-service (DDoS) attacks to data breaches.

Intrusion Detection Systems (IDS) are designed to detect and react to malicious activity and unauthorized access in order to counter these threats. By allowing automated cyber threat identification and

categorization, machine learning (ML) has demonstrated promise in improving intrusion detection systems (IDS) in recent years. Using the CICIDS2017 dataset, this study investigates the use of machine learning (ML) methods, particularly Support Vector Machine (SVM), Decision Tree, and Random Forest, for precise cyberattack detection. These methods seek to increase overall network security, minimize manual involvement, and identify assaults early.

2. LITERATURE SURVEY

a) **Detecting cyber attacks on web applications by applying Different Machine learning techniques:**

Growing usage of cloud services, an increase in the number of users of online applications, modifications to network architecture connecting devices running mobile operating systems, and constantly evolving network technology all contribute to new cyber security challenges. Network security methods, sensors, and protection systems must change to accommodate evolving threats and user needs. Since they are recognized as the biggest risks and the primary concern for network and cyber security, we will focus on thwarting growing application layer cyberthreats in this essay. The paper's primary contribution is the suggestion of a machine learning technique for mimicking common application behavior and detecting cyberattacks. The model is created by deriving patterns in the form of Perl compatible regular expressions (PCRE) using a graph-based segmentation technique and dynamic programming. The model is based on data obtained from client-generated HTTP requests to a web server. Using the CSIC 2010 HTTP Dataset, we tested our approach and discovered that it worked well.

b) **Anomaly detection methods in wired networks: a survey and taxonomy**

Therefore, despite advancements over the past 20 years, anomaly detection in network behavior is still an underdeveloped technology, as seen by the absence of commercial solutions. However, the benefits—particularly in network security—that might arise from both improving existing techniques and gaining a better understanding of the issue itself justify the need for more research in this area.

This paper presents an analysis of contemporary anomaly detection methods for network intrusion detection in conventional wired systems. A taxonomy of current solutions is provided once the topic is introduced and its significance is clarified. We can examine the many facets of the issue and systematically classify current detection methods thanks to the approach that has been described. The key concepts are then examined and illustrated through several case studies of specific systems developed in the area. This makes clear the problems that each of them addresses as well as their areas of weakness. Finally, a review of the outstanding concerns concludes this work. This discussion aids in identifying possible study paths.

c) **Combined analysis of support vector machine and principle component analysis for IDS:**

A variety of subscribers are managed by energy users, who also utilize equipment to measure, bill, disconnect, and connect subscribers from the connection management. These days, we use smart gadgets to store, retrieve, and process data on the cloud. Since information transfer is essential to the operation of these intelligent systems when storing large amounts of data, reported data from networks

should be managed to prevent fraudulent activities and issues that might affect the system's potential quality of service. This paper proposes the use of an intrusion detection system based on principle component analysis (PCA) and support vector machines to detect and identify intrusions and attacks in the smart grid in order to control reported wireless data and ensure the accuracy of the information obtained. Here, we examine how intrusion detection systems function for different SVM kernels when PCA and support vector machines (SVM) are used together. Five potential intrusion detection system kernels are simulated numerically using support vector machines and PCA, based on data KDD99, in order to evaluate the method. Comparative analysis is also looked at in terms of time-response, the pace at which network efficiency and system error are increasing, and differences in the usage or non-use of PCA for the particular intrusion detection technique. The results demonstrate that when PCA is used, both the correct detection rate and the rate of attack error detection have the best values; when the algorithm's core is radial type, the SVM approach reduces data processing time and enhances intrusion detection performance.

d) **Building an efficient intrusion detection system based on feature selection and ensemble classifier**

Intrusion detection systems (IDS) are one of the frequently used techniques in a network architecture to guarantee the availability and integrity of sensitive data in the protected systems. Even though several machine learning-based supervised and unsupervised learning approaches have been used to increase the efficacy of intrusion detection systems, contemporary intrusion detection systems still struggle to perform well. First, a large amount of redundant and useless data in high-dimensional datasets hinders an IDS's categorization process. Second, a single classifier might not be able to identify every type of attack. Third, because many models are built on outdated datasets, they are less adaptable to new threats. Therefore, we provide a novel intrusion detection framework based on ensemble learning and feature selection techniques in this paper. First, a dimensionality reduction heuristic called CFS-BA is introduced, which selects the optimal subset based on feature correlation. Next, we provide an ensemble approach that combines the Random Forest (RF), Forest by Penalising Attributes (Forest PA), and C4.5 algorithms. Finally, a voting method for attack recognition is used to integrate the probability distributions of the fundamental learners. The experimental results using the NSL-KDD, AWID, and CIC-IDS2017 datasets demonstrate that the suggested CFS-BA-Ensemble method can outperform other related and state-of-the-art approaches under a variety of criteria.

e) **An efficient XGBoost–DNN-based classification model for network intrusion detection system:**

The trend of Internet technology's usefulness is rising dramatically every day. A lot of data is generated and maintained as a result of this significant increase. For obvious reasons, network security requires full attention. An intrusion detection system is crucial in the aforementioned security domain. The suggested XGBoost–DNN model uses a deep neural network (DNN) for network intrusion classification after using the XGBoost method for feature selection. The XGBoost–DNN

model consists of three stages: normalization, feature selection, and classification. The Adam optimizer is used to optimize learning rate during DNN training, whereas the softmax classifier is used to classify network intrusions. The tests were conducted appropriately utilizing Tensor flow and Python on the benchmark NSL-KDD dataset. The proposed model is validated by cross-evaluation and comparison with existing shallow machine learning methods including logistic regression, SVM, and naive Bayes. The generated classification assessment parameters, such as accuracy, precision, recall, and F1-score, are used to compare current shallow techniques. The proposed approach outperformed the existing shallow methods used for the dataset.

3. METHODOLOGY

i) Proposed Work:

By examining trends and abnormalities in network traffic data, the suggested solution use machine learning techniques to identify cyberattacks. It is more appropriate than previous datasets like KDD99 since it makes use of the CICIDS2017 dataset, which contains current attack scenarios. The system uses supervised learning techniques to categorize different kinds of assaults, including DoS and DDoS. Models that can reliably anticipate and identify harmful activity based on network behavior are constructed using algorithms like Support Vector Machine (SVM), Decision Tree, and Random Forest.

The system may send email notifications to network managers as soon as it detects a possible assault, guaranteeing prompt action to stop or lessen harm. This system's primary benefit is its capacity to identify dangers with little assistance from humans. Organizations may improve their security posture and protect sensitive data from disruption or unauthorized access by automating threat detection and speeding up reaction times. Adaptability and efficacy against new cyberthreats are guaranteed by the ability to include different ML models.

ii) System Architecture:

The system architecture is made up of a number of essential parts that cooperate to use machine learning to identify cyberattacks. First, the CICIDS2017 dataset is used to gather and feed network traffic statistics into the system. To get ready for analysis, this raw data goes through preprocessing procedures such feature extraction, normalization, and noise reduction. Training and testing sets are then created from the preprocessed data. The labeled data is used to train a variety of machine learning algorithms, including SVM, Decision Tree, and Random Forest, to identify patterns linked to both benign and malevolent conduct. After being trained, these models identify possible threats by classifying incoming traffic in real-time. Administrators get alarm notifications from the system upon detection. In order to deliver reliable, automated intrusion detection with the least amount of human involvement, this design guarantees a smooth transition from data collecting to real-time danger alerting.

iii) Modules:

a. Dataset Collection

The CICIDS2017 dataset is collected, which contains both normal and various types of attack traffic for training and evaluation.

b. Data Preprocessing

Raw data is cleaned and transformed to a suitable format by removing missing values, normalizing features, and converting categorical data to numerical.

c. Feature Extraction and Selection

Important features relevant to detecting attacks are extracted to reduce dimensionality and improve model efficiency.

d. Data Splitting

The processed dataset is split into training and testing sets to evaluate the machine learning models.

e. Model Training

Machine learning algorithms like Support Vector Machine (SVM), Decision Tree, and Random Forest are trained using the training dataset.

f. Attack Detection

The trained models analyze incoming traffic data and classify it as normal or attack (e.g., DoS, DDoS, etc.).

g. Alert Notification System

Upon detecting an attack, the system sends real-time alerts via email or dashboard to network administrators.

h. Performance Evaluation

The model's performance is measured using metrics such as accuracy, precision, recall, and F1-score to determine detection effectiveness.

iv) Algorithms:

a. Support Vector Machine (SVM)

For classification problems, SVM is a supervised learning algorithm. Finding the ideal hyperplane that divides data points from several classes with the greatest margin is how it operates. In this study, SVM learns from labeled data and finds patterns linked to cyberattacks to assist differentiate between malicious and benign network traffic.

b. Random Forest

In order to increase accuracy and decrease overfitting, Random Forest is an ensemble learning technique that constructs many decision trees and aggregates their outputs. A random subset of the data is used to train each tree. Random Forest enhances resilience and offers higher generalization in identifying intricate attack patterns for cyberattack detection.

c. Decision Tree Classification Algorithm

An easy-to-understand algorithm that divides data according to decision rules generated from feature values is called a decision tree. It creates a structure like a tree, with each leaf node representing a class name and each inside node representing a test on a feature. Based on logical criteria, this approach aids in categorizing network behavior as either malicious or legitimate.

4. EXPERIMENTAL RESULTS

The CICIDS2017 dataset, which contains a range of actual attack scenarios, was used to assess the suggested solution. The dataset was divided into training and testing sets following preprocessing and feature extraction. Performance criteria such as accuracy, precision, recall, and F1-score were used to apply and evaluate the machine learning algorithms SVM, Decision Tree, and Random Forest. Random Forest outperformed Decision Tree in terms of accuracy, whereas SVM had respectable but lower detection rates. The system correctly identifies threats and promptly generates notifications for network security, demonstrating the efficacy of machine learning algorithms in detecting cyberattacks.

Accuracy: How well a test can differentiate between healthy and sick individuals is a good indicator of its reliability. Compare the number of true positives and negatives to get the reliability of the test. Following mathematical:

Accuracy = TP + TN /(TP + TN + FP + FN)

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

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{TP + FN}$$

mAP: Mean Average Precision (MAP) is a ranking quality metric. It considers the number of relevant recommendations and their position in the list. MAP at K is calculated as an arithmetic mean of the Average Precision (AP) at K across all users or queries.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic.

$$\text{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

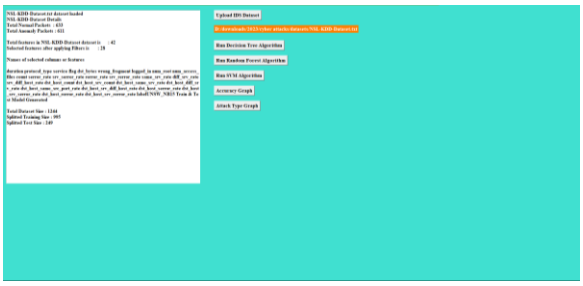


Fig: **Data set Loaded**

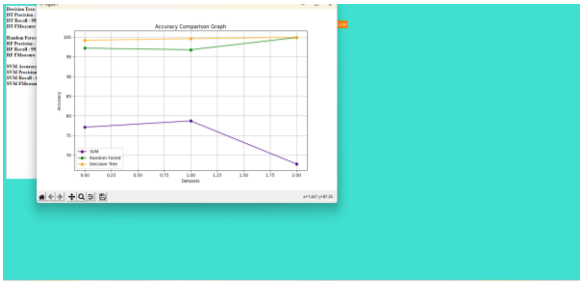


Fig: Accuracy Comparison Graph

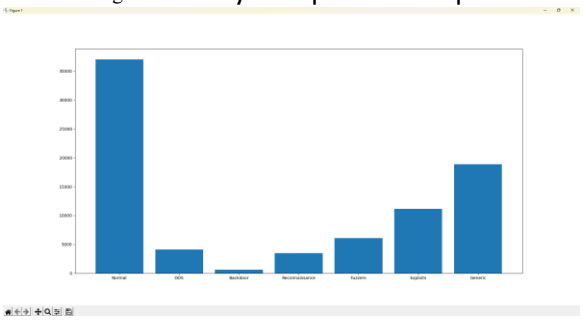


Fig: Predict Attacks In Graph

5. CONCLUSION

Assist vector machines, decision trees, SVM, Random Forest, and deep learning computations are now introduced in a rather basic manner based on the current dataset. The results show that the great learning computation yielded essentially better results than We will later combine port sweep activities and other attack types using artificial intelligence and significant learning computations with Apache Hadoop and Sparkle innovations based on this dataset. We can detect the attacks in a network thanks to these calculations. It happens in such a manner that, when we go back many years, there may be a lot of attacks taking place; thus, when these attacks are found, the features of the values from whence they originated will be maintained in some datasets. Therefore, our goal is to predict whether or not a cyberattack occurs using these facts. These predictions may be made using three different algorithms. In order to guide the prediction of ideal outcomes to signal whether or not cyberattacks have happened, this study aids in identifying the approach that anticipates the best accuracy rates.

6. FUTURE SCOPE

Deep learning models like CNN and LSTM can be included into this system in the future to improve accuracy and identify more sophisticated threats in real time. Additionally, the technology may be expanded to handle encrypted communications and use unsupervised learning techniques to find zero-day vulnerabilities. Additionally, the system may be made more scalable and effective for big enterprises by implementing cloud-based deployment and automatic response systems. Models for continuous learning that adjust to changing attack patterns would enhance detection even more and lower false positives.

REFERENCES

[1] K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.

[2] R. Christopher, “Port scanning techniques and the defense against them,” SANS Institute, 2001.

[3] M. Baykara, R. Das¸, and I. Karado ˘gan, "Bilgi g ¨uvenli ˘gi sistemlerinde kullanilan arac¸larin incelenmesi," in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp. 231–239.

[4] S. Staniford, J. A. Hoagland, and J. M. McAlerney, "Practical automated detection of stealthy portscans," Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.

[5] S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, "Surveillance detection in high bandwidth environments," in DARPA Information Survivability Conference and Exposition, 2003. Proceedings, vol. 1. IEEE, 2003, pp. 130–138.

[6] K. Ibrahimi and M. Ouaddane, "Management of intrusion detection systems based-kdd99: Analysis with lda and pca," in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1–6.

[7] N. Moustafa and J. Slay, "The significant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems," in Building Analysis Datasets and Gathering

Experience Returns for Security (BADGERS), 2015 4th International Workshop on. IEEE, 2015, pp. 25–31.

[8] L. Sun, T. Anthony, H. Z. Xia, J. Chen, X. Huang, and Y. Zhang, "Detection and classification of malicious patterns in network traffic using benford's law," in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864–872.

[9] S. M. Almansob and S. S. Lomte, "Addressing challenges for intrusion detection system using naive bayes and pca algorithm," in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.

[10] M. C. Raja and M. M. A. Rabbani, "Combined analysis of support vector machine and principle component analysis for ids," in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5.

[11] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," Journal of Computational Science, vol. 25, pp. 152–160, 2018.

[12] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization." in ICISSP, 2018, pp. 108–116.

[13] D. Aksu, S. Ustebay, M. A. Aydin, and T. Atmaca, "Intrusion detection with comparative analysis of supervised learning techniques and fisher score feature selection algorithm," in International Symposium on Computer and Information Sciences. Springer, 2018, pp. 141–149.

[14] N. Marir, H. Wang, G. Feng, B. Li, and M. Jia, "Distributed abnormal behavior detection approach based on deep belief network and ensemble svm using spark," IEEE Access, 2018.

[15] P. A. A. Resende and A. C. Drummond, "Adaptive anomaly-based intrusion detection system using genetic algorithm and profiling," Security and Privacy, vol. 1, no. 4, p. e36, 2018.

[16] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[17] R. Shouval, O. Bondi, H. Mishan, A. Shimoni, R. Unger, and A. Nagler, "Application of machine learning algorithms for clinical predictive modeling:

a data-mining approach in sct," Bone marrow transplantation, vol. 49, no. 3, p. 332, 2014.

Author Profiles



Mr. Y. Nagamalleswararao completed his Master of Technology (M.Tech) from JNTUK, M.Sc (IS) from ANU, and BCA from ANU. He has expertise in System Administration, Network Administration, and Oracle Administration. He is also a Web Developer and Python Developer. Currently, he is working as an Assistant Professor in the Department of MCA at SRK Institute of Technology, Enikepadu, NTR District. His areas of interest include Artificial Intelligence and Machine Learning.



Ms. K. Pavani is working as an Assistant Professor and Head of the Department of MCA at SRK Institute of Technology, Vijayawada. She completed her M.Tech and MCA in Computer Science. She has 10 years of teaching experience at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. Her areas of interest include Machine Learning with Python and DBMS.



Mrs. D.Roshini is an MCA student in the Department of Computer Applications (MCA) at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. He completed his degree in B.Sc. (Computers) from Sri Kakatiya Degree College, Vijayawada. His areas of interest include DBMS and Machine Learning with Python.