IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

INTRUSION DETECTION SYSTEM USING MACHINE LEARNING

[1]Deshpande G. R, [2] Patil Shravani , Sondare Priya , Shinde Akanksha

[1]Assistant Prof Department of Computer Engineering ,1st Student,2Student,3Student Of Computer Engineering Department

Gramin Technical & Management Campus Vishnupuri ,Nanded, India

Abstract: A machine learning-based intrusion detection system is a security tool that utilizes advanced algorithms to automatically detect and respond to suspicious activities within a computer network. It employs machine learning techniques to analyze network traffic patterns, identify anomalies, and distinguish between normal and abnormal behavior, thereby enhancing the network's security and threat detection capabilities. A machine learning-based intrusion detection system is a security tool that utilizes advanced algorithms to automatically detect and respond to suspicious activities within a computer network. It employs machine learning techniques to analyze network traffic patterns, identify anomalies, and distinguish between normal and abnormal behavior, thereby enhancing the network's security and threat detection .

Index Terms – Machine learning, Support Vector, Detection ,Naïve Bayes .

I. INTRODUCTION

A machine learning-based intrusion detection system (ML-IDS) is a security mechanism that employs algorithms to analyze network traffic patterns, identify anomalies, and detect potential security threats or intrusions. By leveraging historical data, ML-IDS models can learn to distinguish between normal and malicious behavior, enhancing the system's ability to detect emerging and sophisticated attacks. This proactive approach allows for real-time threat detection, response, and mitigation, contributing to the overall cybersecurity posture of a network or system. This study introduces a robust machine learning-based intrusion detection system designed to enhance the cybersecurity of computer networks. Leveraging advanced algorithms, the system learns from historical data to identify and alert on abnormal patterns, effectively detecting potential security threats. Through real-time analysis and adaptive learning, the system provides a proactive defense against evolving cyber threats, contributing to the overall resilience of network security infrastructure."

INTRUSION DETECTION SYSTEM

An intrusion detection system (IDS) is like a security alarm for computer networks. It monitors network traffic and activities, looking for any signs of unauthorized access, misuse, or policy violations, and alerts administrators or users when it detects suspicious behavior. An intrusion detection system (IDS) is a device or software application that monitors a network for malicious activity or policy violations. Any malicious activity or violation is typically reported or collected centrally using a security information and event management system.

MACHINE LEARNING IN INTRUSION DETECTION

The role of machine learning in intrusion detection systems is pivotal for enhancing the capabilities of traditional security measures.

Here are key roles played by machine learning

Anomaly Detection:

Machine learning enables the detection of unusual patterns or behaviors that may indicate a security threat. It can identify deviations from normal network or system activities, allowing for the early detection of potential intrusions

.Behavioral Analysis:

ML models can analyze and understand normal behavior patterns within a system. By recognizing deviations from this baseline, they can identify suspicious activities, even those that do not match known attack signatures.

Adaptability to Emerging Threats:

One of the significant advantages is the ability to adapt to new and evolving threats. Machine learning models can learn from new data, allowing the intrusion detection system to recognize and respond to previously unseen attack patterns.

Reducing False Positives:

Traditional systems may generate false positives, triggering alerts for normal behavior that resembles malicious activity. Machine learning can reduce false positives by learning to differentiate between benign and malicious behavior more accurately.

Real-Time Detection:

ML-based intrusion detection systems can operate in real-time, providing timely alerts and responses to security incidents as they occur. This is crucial for minimizing the impact of security breaches.

Continuous Learning:

Machine learning facilitates a continuous learning process. As the system encounters new data, it refines its understanding of normal and malicious behavior, improving its accuracy over time.

Handling Complex Data:

ML models can effectively process and analyze large volumes of complex data, including diverse sources such as network traffic, logs, and user activities. This capability is essential for detecting sophisticated and multifaceted cyber threats.

Customization for Specific Environments:

Machine learning models can be tailored to the specific characteristics and requirements of different environments, providing a more customized and effective intrusion detection solution.

In summary, machine learning enhances intrusion detection systems by bringing adaptability, accuracy, and real-time capabilities to the identification and response to security threats, making them more robust in the face of evolving cyber risks.

MODEL IN MACHINE LEARNING

Support Vector Machine

Support Vector Machines (SVMs) in intrusion detection systems work by classifying data points into different categories, such as normal or malicious, based on the features extracted from network or system activities. Here's a simplified overview of how SVMs work in the context of intrusion detection.

Data Collection:

Gather labeled data representing instances of normal and malicious activities. Features could include various aspects of network traffic, system logs, or user behavior.

Feature Extraction:

Identify relevant features from the collected data that can distinguish between normal and malicious behavior.

Data Preprocessing:
Normalize or scale the features to ensure they have similar ranges. This step helps SVM perform optimally.

Training the SVM:
Split the labeled dataset into a training set and a testing set.
Train the SVM using the training set, where the algorithm learns to create a hyperplane that best separates the data into different classes (normal and malicious)

Kernel Function:
Choose an appropriate kernel function. SVMs use kernels to transform the input features into a higher-dimensional space, making it easier to find a hyperplane that separates the data.

Testing and Evaluation:
Use the testing set to evaluate the performance of the trained SVM. Assess metrics like accuracy, precision, recall, and F1 score to understand how well the model classifies normal and malicious instances.

Deployment:
Implement the trained SVM model into the intrusion detection system to classify real-time or incoming data.

Alert Generation:
When the SVM detects an anomaly or potential intrusion based on the learned hyperplane, it can generate alerts or trigger appropriate responses.

Fine-Tuning and Optimization:
Depending on the performance, fine-tune parameters or consider different kernels to optimize the SVM for the specific requirements of the intrusion detection system.

SVMs excel at finding a hyperplane that maximally separates different classes, making them effective for binary classification tasks like intrusion detection. Their ability to handle high-dimensional data and adapt to complex patterns makes them a valuable tool in cybersecurity applications.

Bayes Theorm

Bayes provides their thoughts in decision theory which is extensively used in important mathematics concepts as Probability. Bayes theorem is also widely used in Machine Learning where we need to predict classes precisely and accurately. An important concept of Bayes theorem named **Bayesian method** is used to calculate conditional probability in Machine Learning application that includes classification tasks. Further, a simplified version of Bayes theorem (Naïve Bayes classification) is also used to reduce computation time and average cost of the projects.

Bayes theorem is also known with some other name such as **Bayes rule or Bayes Law.** ***Bayes theorem helps to determine the probability of an event with random knowledge***. It is used to calculate the probability of occurring one event while other one already occurred. It is a best method to relate the condition probability and marginal probability Naive Bayes is a probabilistic machine learning algorithm commonly used in intrusion detection systems for classifying network or system activities as normal or malicious.

The theorem is stated in mathematics in the manner shown below −

$$P(A|B) = P(B|A) * P(A)/P(B)$$

Where ,
The conditional probability that event A will happen if event B has already happened is known as P(A|B).
P(B|A) is the conditional probability of event B given that event A has already occurred.
P(A) represents the earlier probability that event A will take place.
P(B) represents the probability that event B will occur.
According to the Bayes Theorem, the probability of an event A occurring given evidence B is calculated by multiplying the likelihood of evidence B given the occurrence of event A by the prior probability of A and dividing the result by the prior probability of B.

Naive Bayes is particularly useful for its simplicity, efficiency, and effectiveness in scenarios where the independence assumption is reasonable. It's a well-suited algorithm for text classification and can be adapted for intrusion detection tasks by appropriately selecting and engineering features related to network or system activities.

Implementation

The block diagram of this approach is given in "Fig. 1". Accuracy has been calculated and a graph has been plotted based on the obtained results. From the graph, we have can analyze, SVM outperforms Naïve Bayes.

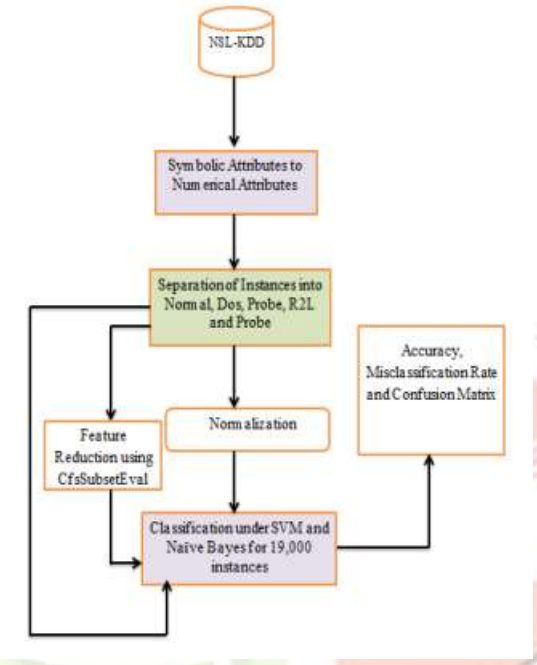

fig.block diagram

Data Collection: Gather a dataset containing both normal and anomalous network behavior. Popular datasets like NSL-KDD or UNSW-NB15 are commonly used

Data Preprocessing: Clean and preprocess the data to handle missing values, normalize features, and convert categorical variables. This step is crucial for the effectiveness of the machine learning model.

Feature Selection: Identify relevant features for the detection task. Feature engineering may be necessary to enhance the model's ability to differentiate between normal and malicious activities.

Model Selection: Choose a suitable machine learning algorithm. Common choices include decision trees, random forests, support vector machines, or neural networks. Experiment with different models to find the one that performs best for your specific dataset.

Model Training: Split the dataset into training and testing sets. Train the chosen model on the training set, adjusting parameters to optimize performance. Use the testing set to evaluate the model's accuracy.

Evaluation: Assess the model's performance using metrics like precision, recall, F1 score, and accuracy. Fine-tune the model as needed.

Deployment: Once satisfied with the model's performance, deploy it in your network environment. Ensure it can handle real-time data and adapt to evolving threats.

Monitoring and Updating: Regularly monitor the IDS and update the model as new data becomes available. This helps the system adapt to changes in network behavior and maintain its effectiveness over time.

Conclusion

machine learning-based intrusion detection systems offer promising capabilities for enhancing cybersecurity. They leverage advanced algorithms to detect patterns and anomalies in network behavior, improving the ability to identify potential threats. However, their effectiveness depends on robust training datasets, continuous updates, and adaptation to evolving attack methods. Despite their potential, these systems should be part of a comprehensive cybersecurity strategy, complemented by traditional security measures, to provide a more resilient defense against cyber threats.

REFERENCES

[1]H.Wang,J.Gu,andS.Wang,‘‘An effective intrusion detection framework based on SVM with feature augmentation,’’ Knowl.-Based Syst., vol. 136, pp. 130–139, Nov. 2017.

[2]Setareh Roshan, Yoan Miche, Anton Akusok, Amaury Lendasse; “Adaptive and Online Network Intrusion Detection System using Clustering and Extreme Learning Machines”, ELSEVIER, Journal of the Franklin Institute, Volume.355, Issue 4,March 2018,pp.1752-1779.

[3]Wathiq Laftah Al-Yaseen , Zulaiha Ali Othman , Mohd Zakree Ahmad Nazri; “Multi-Level Hybrid Support Vector Machine and Extreme Learning Machine Based on Modified K-means for Intrusion Detection System”, ELSEVIER, Expert System with Applications, Volume.66,Jan 2017,pp.296-303.

[4]Iftikhar Ahmad, Mohammad Basheri, Muhammad Javed Iqbal, Aneel Raheem; “Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection”, IEEE ACCESS, Survivability Strategies for Emerging Wireless Networks, Volume.6,May 2018,pp.33789-33795.

https://ieeexplore.ieee.org/document/8862784

https://www.tutorialspoint.com/what-is-bayes-theorem-in-machine-learning

https://chat.openai.com/