"LEVERAGING SVM FOR ENHANCED DETECTION OF ATTACKS IN WIRELESS SENSOR NETWORKS"

Mr. Pravin U. Chokakkar[1], Dr. Sunil B. Mane[2]
[1]Student, Computer Science and Engineering, COEP Technological University
A Unitary Public University of the Government of Maharashtra,
Shivajinagar, Pune – 5, India.
[2]Department of Computer Science and Engineering, COEP Technological University
A Unitary Public University of the Government of Maharashtra,
Shivajinagar, Pune – 5, India.

Abstract

Machine learning techniques, particularly Support Vector Machines (SVMs), have been supported to enhance IoT security. Within the framework of an Intrusion Detection System (IDS) designed to oversee IoT devices, this study conducts a comprehensive assessment of SVM methodologies trained with both normal and abnormal activity data, exhibits exceptional classification accuracy for network topologies. Using machine learning techniques, like SVMs, to safeguard IoT networks from malicious intrusions and device failures. Establishing secure and reliable IoT networks and threat detection, this study provides effectively utilizing SVMs to achieve this goal. This project employs supervised machine learning to identify security threats particularly like blackhole, flooding, grayhole within IoT network and offers practical experience in safeguarding while enhancing network security.

Keywords:- Internet of Things, Wireless Sensor Networks, Machine Learning, Support Vector Machine, Attack Detection, Data Preprocessing, IoT Security.

1 INTRODUCTION

Internet of Things (IoT) is being integrated into our daily lives through the connectivity of smart devices. However, IoT is vulnerable to malicious attacks and device failures that threaten data integrity and application performance. In this work, we explore the application of support vector machines (SVMs) to enhance IoT security to identify and classify network activities in IoT networks using SVMs and machine learning. The results demonstrate SVM's ability to protect IoT networks and anticipate evolving threats like blackhole,flooding ,grayhole ,TDMA, malicious C&C also protecting with higher accuracy with proposed model and highlighting the importance of security controls in an IoT dependent world.

Wireless networks have become an essential part of modern life, playing a crucial role in enabling internet access in remote areas and powering new technologies like smart homes and IoT devices. It has transformed the way we communicate, access information, and engage with our environment by enabling connectivity between devices and people. Physical cords are no longer required, giving users more flexibility and mobility to access information and communicate from various locations. It is more practical and effective than cable networks, which might limit mobility. It enables rapid and simple setup simultaneous connecting of several devices. Wireless networks are a necessary tool in many different industries, including healthcare, transportation, and logistics. They are also cost and time efficient. A wired network setup can be difficult. It can be rapidly and easily set up. For link several devices at once, enabling more effective data transfer and communication. IoT which just changed the perspective to interact with our homes, appliances, and devices, is one of the most important innovations in recent years. Wireless networks are necessary for IoT devices to connect to the internet and transmit data, which allows for automation, system monitoring, and remote control. For instance, smart thermostats may be managed from a smartphone, and home security systems can notify owners via push notifications of potential dangers. According to a survey by Fortune Business Insights, the global 5G market has also seen rapid expansion. Its valued $ 41.48 billion dollars in 2020 and is anticipated to reach upto $ 414.5 billion dollars by 2027. Wireless networks are anticipated to play an ever-bigger role in linking people and gadgets, spurring economic growth, as they continue to develop and flourish. Wireless networks are becoming a necessary component of modern life.

1.1 Vulnerabilities in IoT

A variety of possible security risks have also been brought on by the expanding use of wireless networks. Hackers can access wireless networks using a variety of tools and methods, including breaking passwords, taking advantage of weaknesses in network protocols, and employing rogue access points. Once they get access, they can intercept and change data, put malicious software on the system, conduct attacks on other computers, or install malware.

A DoS attack involves an attacker overloading a network with traffic so that it can no longer function. Users may experience network outages as a result of this, or in severe circumstances, the network may potentially completely crash. DoS assaults can be launched from a variety of locations, making it challenging to identify the attacker and mount an effective defense.

Wireless networks are susceptible to eavesdropping and interception, when attackers read and intercept data sent over the network. This can be problematic when transmitting private or sensitive information, such as when conducting financial transactions, exchanging medical information, or in other private interactions. Attackers can intercept data using a variety of tools and strategies, such as packet sniffers, man-in-the-middle assaults, and other methods.

Rogue access points, which an attacker puts up as a false access point in order to capture and access data transmitted over the network, can make wireless networks vulnerable. Since rogue access points frequently imitate the design and appearance of authentic access points, they can be extremely challenging to identify.

1.2 SVM Support Vector Machine

1.2.1 Types of SVM

Support Vector Machine (SVM) is a popular and efficient categorization method that divides features into hyperplanes or decision boundaries, allowing for classification. It performs faster and more efficiently when dataset size is smaller than neural networks. SVM is effective in distinguishing between benign and malicious IoT traffic and converts low dimensional features space into high dimensional feature space using intricate kernel functions. It can be useful in classification issues where identifying a good hyperplane is not clear.

There are different types of SVM like Linear, Non-Linear, one-class SVMs in that Linear SVM gives Spam or a detection in that emails can be classified as spam. Otherwise, it is not spam based on their contents. Non-Linear SVM is used for Image classification where images need to be classified into different categories and One-Class SVM is used for Detecting fraudulent detections, in that anomaly detection separates normal data points from anomalies in financial datasets

1.2.2 Feature

Support Vector Machine (SVM) is a powerful and versatile supervised machine learning algorithm that is primarily used for classification but can also be applied to regression and outlier detection. Here are the key features of SVM

1.2.3. Hyperplane

It splits information into hyperplanes for classification. A smaller dataset results in faster and more efficient performance. SVM is helpful for separating malicious from benign IoT traffic because it uses complex kernel operations to transform low dimensional feature space into high-dimensional feature space hyperplane.

1.2.4 Support Vectors

These are the nearest data points to the hyperplane, and they have a direct bearing on its orientation and position. The hyperplane’s location would shift if these points were removed. Thus, support vectors are critical in defining the hyperplane and decision boundary.

1.2.5 Versatility

Because of its kernel functions SVM can handle data that is both linearly and non-linearly separable that’s why its versatile. By selecting the right kernel functions, SVM may be applied to a wide range of problem areas.

2 RESEARCH GAPS

2.1 Limited focus on wireless networks

Wireless network support vector machine research is still in its infancy, compared to traditional enterprise networks, which have received substantial study in this area. More study is required on support vector machine in wireless networks.

2.2 Better dimensional reduction approaches are required

In several research, the problem of big size of data in wireless support vector machine has been addressed using feature selection and feature reduction strategies. To reduce the size of the data more effective algorithms are required.

2.3 Lack of attention to imbalanced data

Many datasets used for support vector machine are imbalanced, implying that there are small number of intrusions relative to regular traffic. Many support vector machine performance has been degraded as a result. To address the issue of uneven data, further research is required.

2.4 Need for more studies on real-world data

There is a need for greater research on real-world data since, even though numerous studies have been done on intrusion detection, support vector machine, machine learning, deep learning the majority of them use simulated datasets. More research on actual data is required to confirm the effectiveness of the suggested strategies.

3 RELATED WORKS

Wireless networks have grown in popularity due to their lower acquisition costs and ease of use compared to wired connections. The usage of a common media and the accessibility of information about wireless networks online, IoT among other features of wireless networks, have made them vulnerable to attacks from potential attackers. wireless intrusion detection systems (WIDS), support vector machine (SVM), One class support vector machine OC-SVM, network intrusion detection system (NIDS), machine learning (ML), deep learning (DL) have been suggested as a solutions.[5]

In this paper the proposals for machine learning models to identify unauthorized activity in computer networks. To find unusual activity within the network, they are employed as anomaly detection techniques. There suggestion is to employ the Support Vector Machine (SVM) learning anomaly detection model for identifying irregularities. The suggested detection model achieves up to 81% accuracy. Using KDD-99 internet traffic dataset.[5]

In this paper Wireless sensor networks are vulnerable to security threats that seek to harm the network, alter its configuration, or intercept network traffic for nefarious purposes. In that to identify whether an attack is occurring within the network. The third line of defense is to quickly neutralize or stop the attack. In order to improve intrusion detection systems (IDS), we address the detection phase in this work and investigate the effects of routing layer attacks in WSNs. IDSs use the information from known attacks to inform their detection decisions. The features and techniques employed by the attacks to influence the network are examined and utilized to inform the development of recovery plans. In that multiple layers are there for monitoring. In this WSP routing protocol is used for monitor.[4]

We compared and contrasted two SVM approaches. OC-SVM, which only detects normal behavioral activities, and C-SVM, which requires two classes: vector values for abnormal activities and vector values for normal activities. Both strategies were applied as components of an intrusion detection system (IDS) to monitor smart node devices and detect unusual activities. We used real network traffic along with the specific network level attacks we implemented to develop and evaluate the SVM detection model. C-SVM is shown to work on unknown topologies with a classification accuracy of 81% and achieves classification accuracy of up to 100% when evaluated with unknown data taken from the same network topology on which it was trained. The OC-SVM generated as a result of the positive activity. [6]

4 METHODOLOGY

The goal of this project is to build a Support Vector Machine (SVM) classifier to detect potential attacks in a Wireless Sensor Network (WSN) dataset. The WSN-DS,IOT23 dataset is utilized, which includes various features indicative of normal and attack behaviors. The dataset should be obtained from a reliable source, such as a research publication or a data repository. The dataset consists of 100,000 samples for this project. Loading the dataset using pandas to facilitate easy manipulation and analysis then identify and select relevant features for the model. Exclude non-informative columns such as ’id’, ’Time’, and ’Attack type’. Normalize the features using standard scaler to standardize the dataset. Rain a Support Vector Machine (SVM) classifier with a linear kernel on the scaled training data. Using the trained SVM model to predict the labels of the test set. Analyze the classification report to understand the model’s performance in terms of precision, recall, and F1-score for each attack type. Identify any potential areas for improvement. Summarize the findings, highlighting the effectiveness of the SVM model in classifying different types of attacks in the dataset.

4.1Proposed System

As proposed system the dataset is IOT 23 that dataset is having various parameters that parameters are Time, Is CH, Who CH, ADV R, JOIN R,,Data R,,Data sent to BS were goes to data processing then load the dataset using pandas to facilitate easy manipulation and analysis After Feature selection identify and select relevant features for the model. Exclude non-informative columns such as ’id’, ’Time’, and’ Attack type’ as target variable. Then Feature Scaling is done on that normalize the features using Standard Scaler to standardize the dataset, ensuring that each feature contributes equally to the model. Model Training is done after that we train a Support Vector Machine (SVM) classifier with a linear kernel on the scaled training data. Then model evaluation is done then after we use the trained SVM model to predict the labels of the test set. and it gives significant amount of increase in accuracy in overall 97.85% with precision recall f1-score.

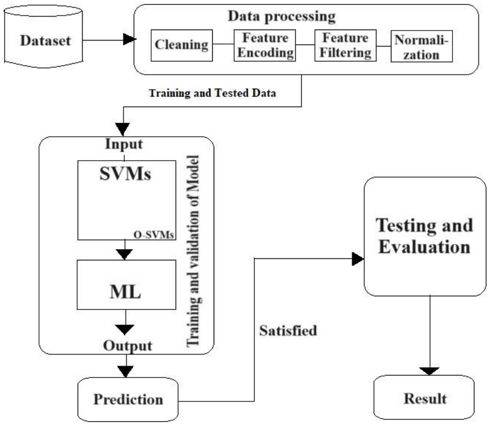

Figure 4.1: Proposed System

4.2Generalized Model

Critical analysis methods can be categorized into applied algorithms, game theory, fuzzy logic, machine learning, and biological methods. But the generalized model shows that more many ways to handle it. The actual measurement itself provides partial information about the distribution of results. It gives the proper input and output for data-set that is filtered and get trained for giving more accuracy of results.

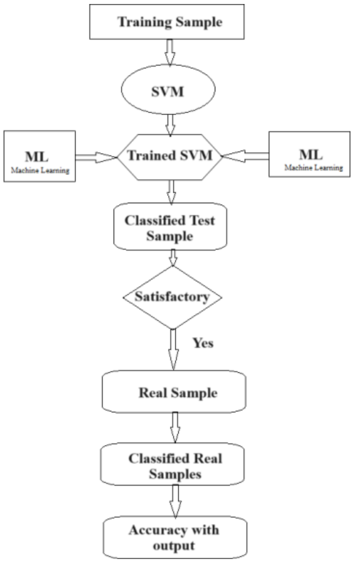

Figure 4.2: Generalized Model

5.DATASET DESCRIPTION

In scalability and efficiency firstly, dataset is got conformed that is IOT23.An annotated dataset of network traffic, the IoT-23 dataset was developed for cybersecurity studies. It includes recordings of both benign and malevolent IoT network traffic. This dataset can be used for a number of tasks, such as researching IoT security, analyzing network behavior, and training machine learning models for intrusion detection systems (IDS). Its ranges of 20 to 50 features common for practical application. The dataset can be downloaded from the official website stratosphere laboratory website. It is a size of 21gb
.rar file and after extracting it ranges upto 150gb.

5.1 Dataset Preprocessing1

The IOT 23 dataset, which is part of the suggested system, has a number of parameters, including time, Is CH, Who CH, ADV R, JOIN R, Data R, and Data sent to BS. this are got selected from the study of attacks in prepossessing. Then after finalizing this attribute, we proceed to data processing. Using Pandas to load the dataset for simple manipulation and analysis. Find and choose pertinent features for the model after feature selection. Leave out columns that don’t provide much information, like” id”,” Time,” and” Attack type”
Subsequently, Feature Scaling is carried out to ensure that every feature contributes equally to the model by standardizing the dataset using Standard Scaler. After completing the model training, we use the scaled training data to train a Support Vector Machine (SVM) classifier with a linear kernel. Following model evaluation, we utilize the SVM model that has been trained to forecast the accuracy.

5.2 Machine Learning

5.2.1 Random Forest

Random Forest is a widely used machine learning algorithm for classification and regression problems. This is an ensemble learning method that improves model accuracy and reliability by combining multiple decision trees. In a random forest, each decision tree is trained on a randomly selected subset of the training data and a random subset of the features. This randomness helps reduce overfitting and increases tree diversity in the forest. Random forests are often used for feature selection as they provide an importance score for each feature in the dataset. This estimate is based on the impurity reduction each feature achieves when used to split data in the decision tree that makes up the random forest.

6. DISCUSSIONS AND EXPERIMENTAL RESULTS

6.1 Performance Metrics

To evaluate the performance of the proposed system and verify its effectiveness and robustness, WSD, IOT23 database is divided into training and testing data. Using SVMs use optimization techniques to find the hyperplane that best separates the data into different classes. The SVM training process involves solving a quadratic optimization problem, and this process is iterative but not divided into epochs. The proposed system is implemented using google colab. The experiments are performed to detect Multi Layered attacks in the dataset.

6.1.1 Accuracy

The ratio of accurate predictions to total predictions is known as accuracy. This is one of the most straightforward metrics, and models frequently use it as a key performance indicator.

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN} \qquad (6.1)$$

6.1.2 Recall

Recall (also known as sensitivity or true positive rate) is the ratio of true positive predictions to the total number of actual positives. It measures the model’s ability to captures all the positives.

$$Recall=\frac{TP}{TP+FN} \qquad (6.2)$$

6.1.3Precision

The ratio of true positive predictions to all predicted positives is known as precision, or positive predictive value. It gauges how well the model predicts the positive outcomes.

$$Precision=\frac{TP}{TP+FP} \qquad (6.3)$$

6.1.4 F1 Measure

The harmonic mean of recall and precision is known as the F1 Measure. It offers a single measure that, particularly in cases of an uneven class distribution, strikes a balance between precision and recall concerns.

$$\text{F1}=\frac{2\times Precision\times Recall}{Precision+Recall} \qquad (6.4)$$

6.2 Experiment Results

6.2.1 F1Score and Accuracy Recall in Attacks

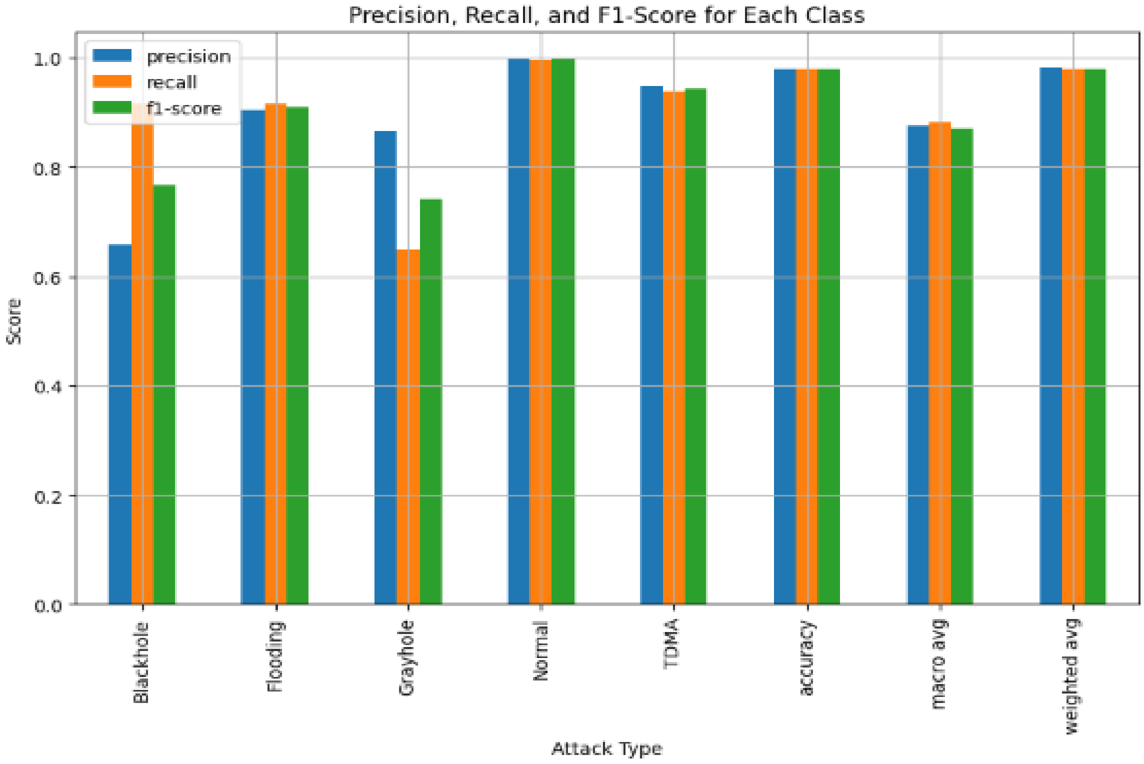

Figure 6.1: F1-Score and Accurate Recall in Attacks

The blackhole, flooding, grayhole, and TDMA attacks are depicted in Figure 6.1, along with the percentage of accuracy of the SVM training model using our suggested model. We only used significant parameters in the SVM model. The SVM model was able to classify the attacks with 98% accuracy, and the highest accuracy was 97.85%

6.2.2 Accuracy Results

Attacks	Precision	Recall	F1 Score
Blackhole	0.66	0.92	0.77
Flooding	0.90	0.92	0.91
Grayhole	0.87	0.65	0.74
TDMA	0.95	0.94	0.94
Accuracy	-	-	0.98

Table 6.1: The prior model’s F1 score, recall, and precision

The percentage of our SVM training model's accuracy when we used all of the parameters is displayed by the accuracy of the model per malicious scenario. Examine the significance of utilizing important parameters at the scaling step to enhance malicious CC 94% precision and accuracy of 80% with a total accuracy of 98%

6.3 Comparison of Approach in Proposed Framework

Accuracy percentage of SVM model per malicious scenario shows the accuracy percentage of SVM trained model when all parameters are used. The scaling step confirms the importance of using meaningful parameters to increase the malicious C& C 94% accuracy and 80% accuracy, resulting in an overall accuracy of 98% as we learned that there are so many studies done in IoT Security with using KDD CUP-99 dataset, KDD99 dataset but on IOT23 it is a wireless sensors dataset that are recently published for that dataset improving accuracy upto 98% is achieved by this model.

7. CONCLUSIONS AND FUTURE WORK

We suggested training and evaluating an SVM machine learning model with both benign and malicious input. Additionally, the derived SVM model was evaluated using training data.

The detection accuracy levels reach upto the 97.85% to 99% that is achieved for new dataset like IOT 23 as in WSD wireless sensor data that are attacked by various attacks like blackhole, grayhole, flooding, TDMA, malicious C& C. which may cause vulnerabilities for network that we have detected and improved accuracy for IoT security to the internet.

References

[1] Rasheed Ahmad and Izzat Alsmadi. Machine learning approaches to iot security: A systematic literature review. Internet of Things, 14:100365, 2021.

[2] Zeeshan Ahmad, Adnan Shahid Khan, Cheah Shiang, and Farhan Ahmad. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Transactions on Emerging Telecommunications Technologies, 32, 01 2021.

[3] Mohammed Ali Al-Garadi, Amr Mohamed, Abdulla Khalid Al-Ali, Xiaojiang Du, Ihsan Ali, and Mohsen Guizani. A survey of machine and deep learning methods for internet of things (iot) security. IEEE Communications Surveys & Tutorials, 22(3):1646–1685, 2020.

[4] Christiana Ioannou and Vasos Vassiliou. The impact of network layer attacks in wireless sensor networks. In 2016 International Workshop on Secure Internet of Things (SIoT), pages 20–28, 2016.

[5] Christiana Ioannou and Vasos Vassiliou. Classifying security attacks in iot networks using supervised learning. In 2019 15th International Conference on Distributed Com- puting in Sensor Systems (DCOSS), pages 652–658, 2019.

[6] Christiana Ioannou and Vasos Vassiliou. Network attack classification in iot using support vector machines. Journal of sensor and actuator networks, 10(3):58, 2021.