International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Security of IOT Devices using Machine learning

P. Sai Gaurav[1], B. S. Nithhiin[2], K. Raghu Vamshi[3], K. Madhu Babu [4]

[1]*UG student, Dept. of* Electronics *and Computers Engineering, Sreenidhi Institute of Science and Technologies, Telangana, India*
[2]*UG student, Dept. of Electronics and Computers Engineering, Sreenidhi Institute of Science and Technologies, Telangana, India*
[3]*UG student, Dept. of Electronics and Computers Engineering, Sreenidhi Institute of Science and Technologies, Telangana, India*
[4]*Asst. Professor, Dept. of Electronics and Computers Engineering, Sreenidhi Institute of Science and Technologies, Telangana, India*

Abstract –*As the Internet of Things (IoT) devices continue to expand and evolve rapidly, cyber attackers are finding new ways to exploit these devices for malicious purposes. The increasing volume of network traffic in the IoT makes it challenging to detect attacks and identify malicious traffic in its early stages. Moreover, traditional security measures are inadequate in securing IoT devices as they have limited storage and processing power. To overcome these difficulties, this study suggests a machine learning-based architecture that makes use of the classification techniques Support Vector Machine (SVM) and Random Forest (RF) to identify and categorise possible risks in IoT network data.*

***Key Words*: Support Vector Machines (SVM), Random Forest (RF), Internet of Things (IOT), Network Traffic.**

1.INTRODUCTION

The term "Internet of Things" (IoT) refers to a network of actual physical objects or "things" that are outfitted with sensors, software, and other innovations to enable communication and data exchange with other equipment and systems over the internet. These things can be anything from straightforward domestic goods to intricate industrial tools. There are currently over 7 billion linked IoT devices, and it is predicted that this number will rise to 10 billion by 2020 and 22 billion by 2025, suggesting a considerable development tendency. The unique feature of IoT is its ability to enable the creation of intelligent systems that can operate independently by making decisions and taking actions based on the data they receive and analyze. This autonomous operation allows for the optimization of various processes, improvement of efficiency, and enhancement of the user experience. By leveraging the power of IoT, businesses can achieve significant improvements in their operations, as well as providing more personalized and responsive services to their customers.

1.1 IOT ARCHITECTURE

Four important levels make up the IoT architecture: the Perception Layer, Network Layer, Middleware Layer, and Application Layer. The Perception Layer, or sensing layer, is the top layer of this design. This layer includes various devices such as sensors, RFID tags, and actuators that collect data from the physical world. The primary role of the Perception Layer is to gather data about the physical environment, objects, and people. Sensors are the fundamental components of this layer, as they can detect and measure physical properties like temperature, humidity, pressure, acceleration, and light intensity. They then convert these measurements into electrical signals that can be processed by the IoT system. Actuators are also part of the Perception Layer, and they are devices that can control physical systems in the environment.

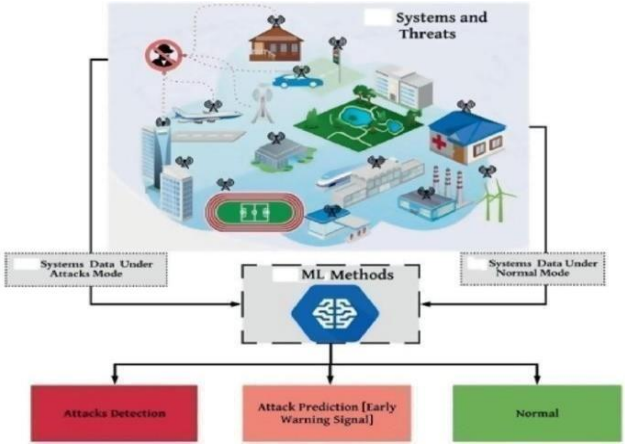

Fig -1 : IOT Architecture

2. LITERATURE SURVEY

In their work titled "Security and Privacy in IoT Using Machine Learning and Blockchain," Nazar Waheed and Xiangjian He proposed the utilization of both Machine Learning and Blockchain technologies to enhance network analysis in the Internet of Things (IoT) domain. The authors conducted a comprehensive examination of various types of features, including their performance metrics, techniques for extracting features, criteria for selecting features, accuracy rates, and detection methods. While multiple detection techniques were considered for each feature set, Random Forest (RF), Support Vector Machines (SVM), and K-Nearest Neighbours (KNN) were the most commonly used ones. [1]

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Khadijeh Wehbi, Liang Hong (2019 IEEE) "A Survey on Machine Learning Based Detection on DDoS Attacks for IoT Systems" discusses the most prevalent malicious assaults This model has been used to study and identify distributed denial of service (DDoS) and denial of service (DoS), which have become important security risks to all networks and in particular to IoT devices. In order to differentiate between regular and unusual traffic in their investigation, the authors investigated two categories of features: stateless and stateful features. They examined how these features could aid in the process of feature extraction. The researchers evaluated five distinct Machine Learning classifiers, namely K-Nearest Neighbors (KNN), Linear Support Vector Machines (LSVM), Neural Network (NN), Decision Tree (DT), and Random Forest (RF). Their findings indicated that K-Nearest Neighbors, Random Forest, and Neural Network classifiers exhibited the highest effectiveness in the analysis.. [2]

In their study titled "Internet of Things Cyber Attacks Detection using Machine Learning,"[4] Seven different machine learning methods were used by the authors to examine the Bot-IoT dataset. The selection of the Bot-IoT dataset was made possible by its regular updates, wide variety of attack methods, production of IoT traffic, and ability to create new features directly from the raw information. The researchers used CICFlow Metre, a network traffic flow generator that produces a CSV file and a visual representation of the created features, to extract features from the dataset. To extract features from the dataset, the Random Forest regressor approach was used. [3]

3.EXISTING METHOD

SVM is an effective supervised machine learning algorithm renowned for its ability to handle both regression and classification tasks.

In SVM, the primary goal is to identify a hyperplane that can successfully separate the two classes within a dataset. Although multiple hyperplanes may exist that perfectly separate the classes, SVM selects the most suitable hyperplane by maximizing the distance or margin between them. In other words, SVM aims to find the hyperplane that achieves the maximum separation between the classes. This approach ensures that the classification is accurate and the separation between classes is optimal.

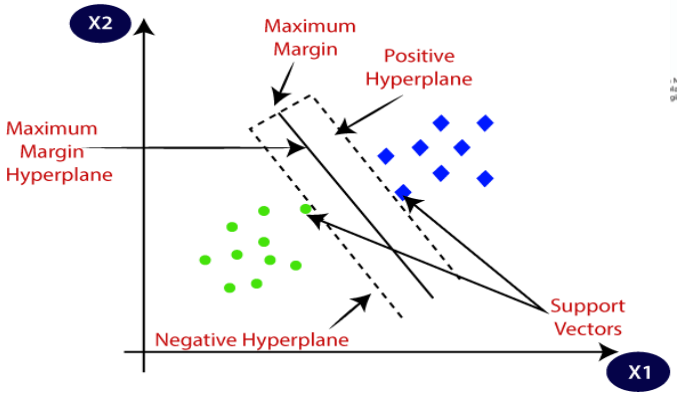

Fig- 3 :SVM

Both linearly separable and non-linearly separable data can be handled by Support Vector Machines (SVM). SVM discovers the best hyperplane to divide data into two groups when dealing with linearly separable data. SVM can map data into a higher-dimensional space where it can be linearly separated, but, in the case of non-linearly separable data.

4.PROPOSED METHOD

Classification and regression issues can be resolved using the machine learning technique known as Random Forest. It is an ensemble learning technique that combines various decision trees to produce an effective predictive model. Each decision tree created by the method is trained using a distinct segment of the training data that is randomly chosen. The algorithm then combines the predictions of these trees to produce a final prediction. This approach reduces the risk of overfitting and makes the model more robust to noisy data. Furthermore, the method effectively uses categorical and numerical data, and it can manage missing data.

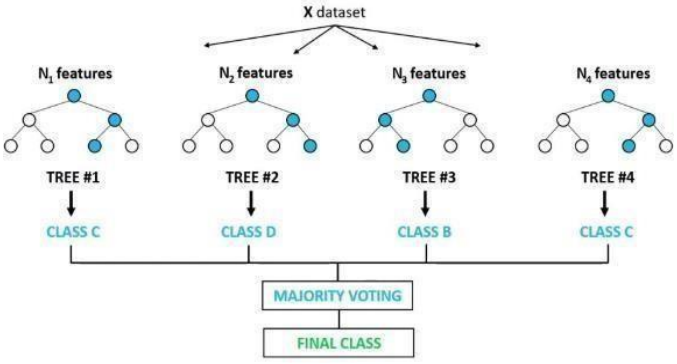

Fig -4: Random Forest

A versatile and potent machine learning algorithm called Random Forest can be used to solve a variety of problems in a variety of industries, including marketing, finance, and healthcare.

5.IMPLEMENTATION

To execute this project one needs to have python software downloaded in the system and the other software requirements are to have VS code editor with the downloaded Jupyter extensions and necessary libraries installed.

1. We downloaded the dataset UNSW_NB15 from the internet, which has a total of 82,332 values
2. We import the dataset into the variables used in the code

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

3. The dataset was then divided into a training dataset and a testing dataset at a ratio of 75%:25%, respectively.
4. Now we fit Random Forest algorithm to the training dataset.
5. We predict the test result after that.
6. Then we test the accuracy of the result.
7. Finally, we do the performance analysis of the results obtained.

6.RESULTS

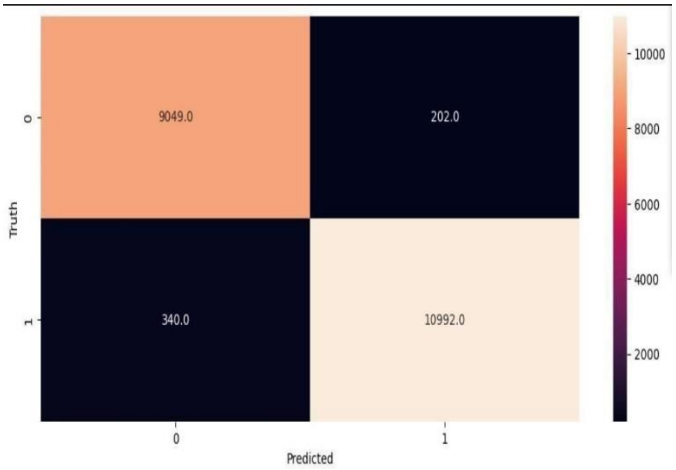

Fig- 6.1: Confusion matrix

By implementing the Random Forest algorithm this is the confusion matrix we obtained through our Machine Learning model on the choosen dataset(UNSW_NB15).

```
Accuracy: 97.37%
Recall: 97.37%
Precision: 97.38%
F1-Score: 97.37%
time to train: 15.72 s
time to predict: 0.24 s
total: 15.97 s
```

Fig -6.2: Performance Scores

These are the performance metrics we obtained for the model using Random Forest algorithm.

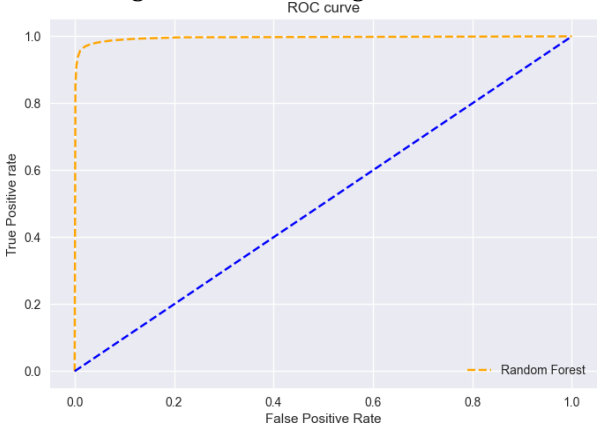

Fig -6.3: ROC curve of Random Forest Algorithm

This ROC curve is the graphical representation of the accuracy achieved by our Machine Learning model using Random Forest algorithm. The accuracy score of the Random Forest algorithm is 97.37%

7.CONCLUSION AND FUTURE SCOPE

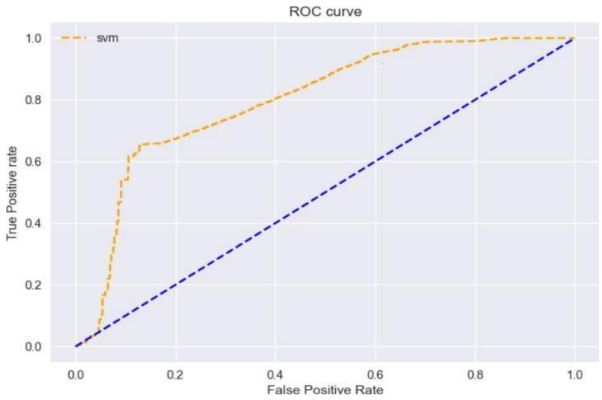

Fig -7.1: ROC curve of SVM Algorithm

The figure 7 is the ROC CURVE obtained for the Machine Learning model using SVM algorithm.The accuracy score of the SVM model is 74.66%.If we compare figure 6.3 and figure 7 we can clearly observe that the accuracy scores of the Random Forest model are way higher when compared to that of the SVM model.

In the future, it is possible that Random Forest can be combined with other machine learning techniques to improve the accuracy of intrusion detection in IoT networks. Furthermore, it can also be used to identify vulnerabilities in IoT devices and develop more secure systems. As IoT devices continue to become more ubiquitous, the use of Random Forest in securing these devices will become increasingly important.

REFERENCES

[1]https://iopscience.iop.org/article/10.1088/17551315/248/1/012002/pdf#:~:text=The%20Random%20Forest

[2]https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0268-2

[3] Abebe Abeshu and Naveen Chilamkurti. 2018. Deep learning: The frontier for distributed attack detection in fogto-things computing. IEEE Communications Magazine 56, 2 (2018), 169–175.

[4] www.ijacsa.thesai.org (IJACSA) International Journal of Advanced Computer Science and Applications,Vol.10,No.12,2019