Machine Learning Approach for Anomaly Detection of IoT Cyberattacks in smart city

Sushmitha R[1], Deepa N P[2]

[1]*MTech DEC, Department of ECE, DSCE, Bangalore, India*
[2]*Assistant Professor, Department of ECE, DSCE, Bangalore, India*

Abstract - *The deployment of IoT devices has led to create smart cities to use most advanced communication technologies. With the increase in the amount of data across the network in a smart city over IoT devices cause IoT cyberattacks. In order to overcome these cyberattacks, an Anomaly Detection of IoT (AD-IoT) system results based on Random Forest machine learning algorithm using the modern UNSW-NB15 dataset. Thus, this system can effectively meet highest classification accuracy with false positive rate.*

Key Words: **Anomaly Detection of IoT (AD-IoT), cyberattacks, Network-based Intrusion Detection System (NIDS), Machine Learning, Random Forest, smart city, UNSW-NB15 dataset**

1. INTRODUCTION

IoT covers a variety of applications by implementing IoT technology. Nowadays many attacks are found in an environment. The number of attacks increase rapidly due to interconnected devices that can communicate over the internet. Moreover IoT generates a huge amount of data that enables attackers to intercept the data while transmitting over the network. Hence, a new technique can be obtained to overcome these cyberattacks.

A smart city uses IoT technology to perform different services within a city. The number of connected IoT devices over heterogeneous types leads to the complexity of IoT networks. These IoT devices with the internet cause serious cyber threats and vulnerabilities for attacking the information in day-to-day activities in a smart city. There are two main security challenge in a smart city. The first challenge is to detect zero-day attacks from a variety of protocols of IoT devices. The second challenge is to detect cyberattacks from the IoT networks before damaging a smart city. These cyber threats can get unauthorized access to the IoT devices without the awareness of the administrator. Thus, to detect these threats, an anomaly detection system technique can be implemented based on machine learning algorithms.

1.1 Intrusion Detection Systems (IDSs)

Previously, traditional Intrusion Detection Systems (IDSs) was used to monitor the amount of data in the network and to detect only known attack. There are mainly two types of IDSs: first is Host-based IDS (HIDS) and second is Network-based IDS (NIDS). Host-based IDS scan the software that is installed on the computer. Thus, HIDS is not significant with some IoT devices. Network-based IDS monitor the amount of data in the network and can detect both malicious and non-malicious attacks based on signature-based and anomaly-based techniques. However, in this paper an anomaly detection system can be applied. This system is an intelligent anomaly detection method used to detect attack and normal which identifies attack based on the amount of data across the network in an environment.

1.2 Machine Learning

Machine learning is a branch of artificial intelligence that focuses on the development of computer programs that can access the data. Machine learning explores the study of algorithms that can learn and make predictions on the dataset based on training and testing phase. Machine learning algorithms can be classified into four categories: supervised learning, unsupervised learning, semi-supervised learning and reinforcement leaning.

1. Supervised learning – the input data contains the label of the dataset.

2. Unsupervised learning – the data is divided into two categories and the model will identify based on the training data on which the testing data should be.
3. Semi-supervised learning – there are no labels for all in the dataset or there are labels for all in the dataset.
4. Reinforcement learning – it performs some functions and make decisions on the basis of the reward obtained.

1.3 Random Forest algorithm

Random forest is a supervised machine learning approach based on decision tree. Random forest can be obtained by considering the features from the dataset. A forest can be created with a large number of different decision tree structures. After the forest is formed, a dataset that needs to be classified for each tree in the forest for classification. Each tree gives a majority count that indicates the tree decision about the feature of the dataset. The output variable can be obtained in both numerical and categorical.

There are two parameters considered:

1. The number of features to be chosen in a dataset.
2. The number of trees that build the forest.

The features of random forest are:

1. It runs effectively on large datasets.
2. It can handle unbalanced datasets.

1.4 UNSW-NB15 dataset

The dataset is used to evaluate the network anomaly detection system. The UNSW-NB15 dataset is used to train and test all the machine learning implementations. Therefore, the UNSW-NB15 dataset includes nine type of attack classifications to update the malicious behavior. The names of these attacks are Normal, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms.

2. LITERATURE SURVEY

Ibrahim Alrashdi, Ali Alqazzaz, Esam Aloufi [1] have proposed traditional IDS for smart city for IoT applications. An approach based on NIDS called AD-IoT system was introduced.

Rashmi H Roplekar and N V Buradkar [2] have proposed the use of machine learning for network anomaly detection. This improved how the accuracy of anomaly detection and also to reduce the false positive rate in the system.

Nour Moustafa and Jill Slay [3] have discussed the UNSW-NB15 dataset and concluded as a modern NIDS benchmark dataset.

Jadel Alsamiri and Khalid Alsubhi [4] aimed to detect the amount of data in the IoT network using machine learning methods.

Fatima Hussain, Rasheed Hussain, Syed Ali Hassan and Ekram Hossain [5] have discussed about the issues related to IoT networks.

3. CONCLUSION

In this paper, we propose an approach to Anomaly Detection of IoT system can effectively provide protection to all types of attacks and to reduce false positive rate in the system. The evaluation of UNSW-NB15 dataset is used to illustrate the model accuracy.

REFERENCES

[1] Ibrahim Alrashdi, Ali Alqazzaz, Esam Aloufi, “Anomaly Detection of IoT Cyberattacks in Smart City Using Machine Learning”, 2019, IEEE.

[2] Rashmi H Roplekar and N V Buradkar, “Survey of Random Forest Based Network Anomaly Detection Systems”, Vol. 6, Issue 12, December 2017.

[3] N Moustafa and J Slay, “Unsw-nb15: a comprehensive data set for network intrusion detection systems”, in Military Communications and Information Systems Conference, 2015, IEEE 2015.

[4] Jadel Alsamiri and Khalid Alsubhi, "Internet of Things Cyber Attacks Detection using Machine Learning", Vol. 10, No. 12, 2019.

[5] Fatima Hussain, Rasheed Hussain, Syed Ali Hassan and Ekram Hossain, "Machine Learning in IoT Security: Current Solutions and Future Challenges.