IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125208

IoT Botnet Cyber Attack Detection

Kajal Sawant[1], Prajakta Jadhav[2], Shraddha Shirsath[3], Shubhangi Dhumal[4]

Department of Computer Engineering (SPPU), G. H. Raisoni College of Engineering And Managment,

Ahmednagar, Maharashtra, India[1-4]

Abstract—Internet Of Things (IoT) is the term which has been popular nowdays as an increasing number of users. Statistically results shows that in the future it will increasing more and more. But because of this large number of uses of users, to maintain their high degree of security is something that is critical. In this research we has been liked to improve the security of IOT devices through applying various Machine Learning Algorithms and through some efficient engineering techniques. In this Paper, we have set up an approach to detect botnet of IOT devices using three ML Algorithms that are: Support vector machine second one is Naïve Bayes and Third one is Decision Tree which all are Supervised Learning Algorithms. And for the purpose to detect Bot we have been using different Bot Datasets. After a number of pre-processing steps, we feed the pre-processed data to our supervised algorithms that can achieve a good precision score that is approximately 77–99%. The SVM achieves the best accuracy score, approximately 99% in every dataset, and Naïve Bayes accuracy score varies from 91% to 98%; however, the Decision Tree achieves lowest accuracy score that is from 77% to 99%. Our algorithms are cost-effective and provide good accuracy in short execution time.

Keywords— IoT, IoT botnet, Botmster, Bot attack, IoT devices, P2P, Datasets, DDoS, Feature Extraction, ML algorithms, Cyber Security, malicious, data pre-processing.

I. INTRODUCTION

In current Digital era, Cyber security is critical to maintain. A high degree of safety due to increasing use of digital communication. This form of assault is difficult to detect because the device continues to function normally, and the user or owner of the device will not realize whether his device is a victim of attack. To design an appropriate security model for detecting cyber threats, the representative dataset must be Well-structured for training the model and subsequently testing the proposed system

A Botnet is a collection of connected devices often within an IOT network that becomes infected and controlled by malware to benefit cyber criminals. Each individual machine or IOT device under the control of the bot-header is known as bot. from one central point, the attacker can command every device on it's botnet to simultaneously carry out a coordinal criminal action. The term 'bot' emerges from 'ro-Bot' which refers to a script or set of scripts for pre-defined functions in an automated fashion. The existence of these botnets has been traced several years ago.

In August 1988, Jarkko Oikarinen University of Oulu, Finland, invented the first bot IRC. These IRCs use low bandwidth and simple communication methods as well as simple construction with a moderate success ratio.The first Worm for remote control using IRC. Pretty Park in June 1999, Against the first malicious bot named as GT-BOT. Recently, within a few years of time period, in June 2020, a malware attack was made to a renowned food major of India named as "Ransomware". In this attack, the hacker hacked all the files and other information and for a handsome amount of $7,50,000. Another kind of crimes refers to the credit card frauds which are best described in . For this purpose, a special fraudulent detection was used which was based upon Intelligent ML technique.

In our research to detect the botnet we have been only focuses on the study of ML algorithms to train the model. And various steps like data pre-processing, feature extraction and classification for the purpose of to detect IoT bot.

II. RESEARCH MOTIVATION

In these day to day worldwide increasing use of Internet and their IoT devices to maintain the high level of cyber security is something that was always challenging task for researchers. As a result, Cybercriminals constantly research new approaches to carry out their illegal task. Hence, malware technique is now growing with new and innovation manner. So to avoid and secure our IoT devices from the Bot attack which is one of the great example of malicious activity. So, as IT student we would find challenging to do it.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125208

III.RESEARCH OBJECTIVES

The objectives of this study are as follows:

i. To transform the raw data into machine learning format using data transformation and pre-processing techniques.
ii. To develop the machine learning model which will be used to classify the IoT botnet attacks.
iii. By applying algorithms that are only type of supervised learning method

IV.RELATED RESEARCH

"Internet of Things" (IoT), the name was given by Kevin Ashton in 1999, when he was working with Auto-ID labs. It refers to the physical devices connected to share data among them, with no or less human interaction. The work of humans here is done by various kinds of sensors that can request or provide a service. This kind of self-learning of the physical systems is done with the help of machine learning, machine-to-machine communication, human-machine interaction, visualization, and data analysis. As these types of devices are more in demand, it has resulted in increased production, which has reduced the security of these systems. The default username and password of the devices, the concept of fixed key value which cannot be changed, are some of the major security issues, which are the doorsteps to different attacks, IoT Botnet being one of them.

As the Botnet has Command-And-Control Architecture. The Bot-master gives command to every bot which is each infected connected IoT device in peer to peer network (P2P). From, this the bot-master carry out malicious activities like Spam generation, DDoS, phishing, infect system etc.

Botnet formation being one of the attacks that activates at a very fast rate and harms the server.the structure of botnet objects and how they interact defines the structure of the attack.

As discussed, bots and botnets are hot topics of discussion in the current scenario due to the increasing growth in the botnet attacks. According to Thakare et al, a network of devices connected and attacked forms a background of botnet. August 1988, a date to remember, gives the description about the origin of botnet from the first bot IRC to it's development stages going on. According to Kevin Ashton, the name was given as "Internet Of Things" (IOT), which brings the concept of IOT Botnet into light. According to various researches, it is a kind of robot that detects and scans for viruses and the viruses are converted to a bot if found, while following various steps of the cycle.

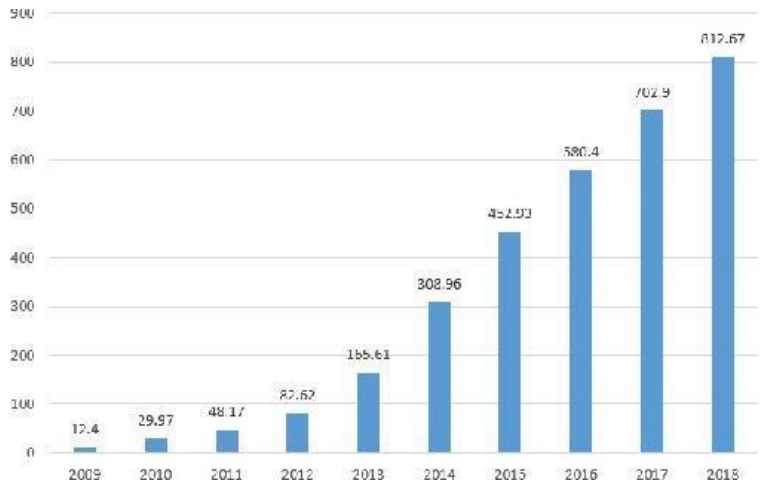

Fig 1. Statistics regarding growth of IOT Botnet

V. METHODOLOGY

The main focus of this paper is towards the [1] Architecture of bot and it's nature [2] the types of bot attack which has been done by attacker (botmaster) [3] types of IOT devices [4] Dataset types used to detect bot [5] training and testing of dataset [6] pre-processing [7]Feature extraction [8] classification [7] result

DOI: 10.17148/IJARCCE.2023.125208

4.1 Architecture of bot

As Kevin Ashton gives the concept of IOT bot into the light where it has Control-And-Command Architecture and each infected IOT Device is known as Bot which is in the control of Bot-master and connected in Peer to Peer network. the structure of botnet objects and how they interact defines the structure of the attack.

The centralized structure refers to the type in which a single point (C&C server) is used for communication between bot-master and bots. A decentralized architecture is a type of architecture where there is no central point of contact, that is, each bot maintains a certain or other connection with other bots. It usually refers to the type of P2P model.

4.2 Datasets

Here are some datasets that we are using for our research are as follows. By using it's parameter we can detect Whether IOT Device contain any bot or not

i. ***IOT_Fridge_dataset:*** used to detect the bot of IOT fridge device.
ii. ***Train_Test_IoT_Motion_Light:*** used to detect the bot of IOT based Motion light devices.

4.3 Types Of Bot Attacks-

Botmaster has a control over the bot of networks that is botnet where the inter communication between object defines the structure of bot. The Datasets contain following types of attacks through it's parameter

i. **DDos:** Distributed Denial of Service attacks in IoT botnet scenarios are cases where many intruders try to intrude into the extensive network to attack a single IoT device. This type of attack compromises the available network resources from intended users and makes the IoT network devices perform isolated
ii. **Backdoor:** in this type of bot attack the malicious activities carried out by a malicious software through which hackers can get prohibited access to the website. In this attack hackers can be get undetected as Malware is installed via weak network entry points such as old plug-ins fields.
iii. **Injection:** these attacks are known as SQL injection attacks where Attackers manage to inject malicious code by alternating SQL queries to manipulate the backend databases, By through which a attackers take advantage of user section such as login, contact etc. to inject a malicious code which will alternate actual SQL command

4.4 IoT Devices

IOT devices are any machines, appliances or gadgets that are programmed for certain applications and can transmit data over internet. Among broader range of IOT devices we will detect only three types of IOT smart devices that are

i. **Motionlight IoT devices:** A Motionlight triggers a response when motion is detected. They can be installed indoors, on walls, celling and indoors smart appliances as well as outdoors too.
ii. **Fridge IoT device:** IOT smart Appliances include smart refrigerator is a kitchen appliances that connects into your smart home system. Smart refrigerators feature a touchscreen interface and ability to connect to the internet through Wi-Fi to provide number of additional features. Some of them also has an internal cameras, more flexible control cooling option. And some Smart refrigerator can also connect with other smart devices in your home; such as speakers, smart TVs etc. and model.

4.5 Training and testing of dataset

Machine learning is concept which is now used worldwide that enables computers or machines to turn a huge amount of data into a predictions. Therefore, the best quality of data only results into a highly accuracy model. In this research we are using a machine learning technology to for detection of IoT device botnet. Train and test datasets are two main key concepts of machine learning.

Training data is fed to the ML algorithms and which helps to make predictions while once the training of the model done with training dataset, then testing of the model will done through testing dataset.

Test dataset is another subset of original data which is independent of training dataset, as test dataset is used to evaluate model.

To increase the model performance, predictability and efficiency we have to split the dataset into two parts that is train dataset and test dataset. For this we can use train_test_split function Scikit-learn.

As In our research to detect IoT botnet, we are using supervised learning algorithm to train the ML model and to detect the IoT botnet. Therefore, we will only using the dataset that has labeled. With the labeled datasets, the model makes predictions and provides accurate results.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125208

4.6 Data pre-processing

Splitting the dataset into train and test sets is one of the important parts of data pre-processing, as by doing so it can improve the performance and hence get result into a better predictability.

It is the first crucial step while creating a machine learning model As, it is the process of preparing or converting the raw data like missing values, noises into a data which is suitable for a machine learning model**.**

4.7 Feature Extraction

Feature Extraction is the process of dimensionality reduction by which an initial raw data is reduced to more manageable groups for processing. Feature extraction makes large number of features extracted into a small another new features where the previous that is original once discarded.

This process is useful when you need to reduce the number of resources needed for processing without losing important or relevant information. It is done through technique such as PCA, ICA, LLE etc.

4.8 Classification

Classification is a problem which is comes under the supervised learning technique. Supervised learning ML model will only learns from labeled data.

In our research for the classification purpose we have been using three types of supervised learning algorithm that are

i. **Support vector machine(SVM):** SVM is most popular Machine learning algorithm comes under the classification problem in ML. to create a best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in future. SVM can give 98% accuracy on the dataset.

ii. **Naïve Bayes:** In our research we are using Naïve Bayes based on bayes theorem for classification. As, it is one of the simple, most effective classification algorithm and also make quick prediction. It predicts on the probability of object.

iii. **Decision Tree:** we have using this algorithm as it usually mimic human thinking ability while making decision, it has tree like structured so internal nodes represent the feature of dataset, and branches represents the decision rules and each leaf node represents the outcome hence it is easy to understand.

4.9 Result

The result of this research will show the output in terms of prediction by filling the box with appropriate parameter and by clicking on submit system will show the output in predictable manner only. As it only detect whether that IoT device present with bot or not.

VI. PROPOSED SYSTEM

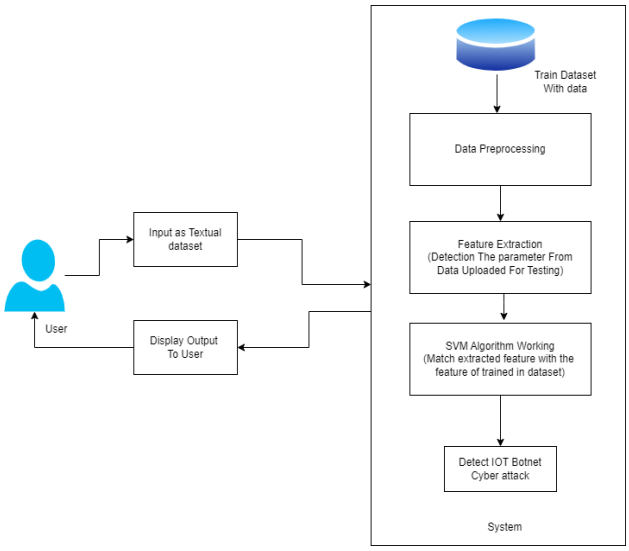

Fig 2: Proposed System

Module 1: User

User Registration If an user already has an account then user can directly login into the system but if doesn't have then user first needs to do the registration.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125208

Module 2: Data processing
It contains of three stages like [1] Data Pre-processing [2] Feature Extraction [3] Classification

VII. FUTURE SCOPE

Future research could include a broader range of IoT devices, and the applications of unsupervised learning classifier. Although many more smart-home devices exist and are being created, our study only looked at network activity from two devices. Anomaly detection on specific brands of the same IoT devices is being considered; this would provide information into the network and application layers of the specific IoT devices.

Finally, our study only looked at training from a labelled dataset; future research will include using unsupervised learning techniques like k-means clustering on unlabeled data to detect malicious activity.

VIII. CONCLUSION

Various malwares are exploiting the vulnerabilities in IoT devices, resulting in large-scale Cyber-attacks. In this propose system, We propose a novel approach for detecting IoT-botnet cyber -attacks. Based on botnet behavior analysis. We use supervised machine learning techniques with the observed attributes as inputs to detect the presence of these botnet cyber-attacks. We also used a variety of different machine-learning techniques and find it's suitability for efficient detection of IoT-botnet cyber-attacks.

ACKNOWLEDMENT

This work is supported in IoT Botnet Detection System for IoT device using ML and Feature engineering. Authors are thankful to Faculty of Engineering and Technology (FET), Savitribai Phule Pune University, Pune for providing the facility to carry out the research work.

REFERENCES

[1] a. yousefpour, g. ishigaki, and j. p. jue, Fog computing: Towards minimizing delay in the internet of things, in edge Computing (EDGE), 2017 IEEE International Conference on. IEEE, 2017, pp. 17-24.
[2] A. Abeshu and N. Chilarnkurti, "Deep learning: the frontier for distributed attack detection in fog-to-things computing," IEEE Communications Magazine, vol. 56, no. 2, pp. 169-175, 2018.
[3] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things," Computer Networks, vol. 57, no. 10, pp. 2266-2279, 2013.
[4] E. Borgia, "The internet of things vision: Key features, applications and open issues," Computer Communications, vol. 54, pp. 1-31, 2014.
[5] F. Restuccia, S. D'Oro, and T. Melodia, "Securing the internet of things: New perspectives and research challenges," IEEE Internet of Things Journal, vol. 1, no. 1, pp. 1-14, 2018
[6] J. Howell. Number of connected iot devices will surge to 125 billion by 2030, IHS Markit says - IHS technology.