International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

BOTNET DETECTION USING MACHINE LEARNING

Mr. A. Sankaran[1], A. Krithika Bavani Murat[2], M. Tharrshinee[3], G. Yuvasree[4]

[1]Assistant Professor, Department of Computer Science Engineering, Manakula Vinayagar Institute of Technology, Pondicherry- 605 107.
[2,3,4]UG Scholar, Department of Computer Science Engineering, Manakula Vinayagar Institute of Technology, Pondicherry- 605 107.

ABSTRACT: *The growth of internet of things leads to rise of botnet attacks. Botnet are the group of computers which connected to each other to perform n number of respective tasks to process the website to keep on working. One of the most powerful ways to pursue any computationally challenging task is to leverage the untapped processing power of a very large number of everyday end points. The idea behind the botnet is a collection of workstations and servers are distributed over the public internet, this leads to the agenda of malicious or criminal entity. The foremost target of the botnet to attack as possible as many devices along with spreading most optimistic through malicious code. The botnet attacks together with infect all kind of technology, rudimentary of internet security suites, firewall including antivirus dispense some protection. In advance we proposed dynamic analysis, looking up for sign of infection in behavioral analysis along with network and picking up unusual network traffic. The attack on botnet symptoms on individual with network levels. In this paper, performance of network dataset has been compared to predict the accuracy and anomalies on the network. The machine learning algorithms which have been used here is Logistic Regression (LR). Our experiments shows, that our approach can compare benign traffic and the junk traffic effectively and reaches the accuracy of 99.98%.*

Keywords – Botnet, Mirai, Bashlite, Logistic Regression.

1. INTRODUCTION

Now-a-days, there is a countless internet of things (IoT) devices has promoted effectually and reached throughout the world. A different types of internet connected devices that are not personal computers are taken as a part of work to get the traffic traces. The rapidly increasing number of IoT devices which can be more leads to enlarge in occurrences of IoT botnet attacks. In order to obtain new thread, there required to developed new method for detecting attack. We put forward a new methodology to detect IoT botnet using machine learning algorithm. Our proposed method has the ability to accurately and instantly detect the attacks as they were being a part of the botnet. Massive exchange of sensitive information in cloud and other wireless transfer. While IoT gives huge benefits of individual and business, it also gives a hoard of security concerns which one cannot turn a blind eye to. IoT, unlike common desktop systems, foundation of the embedded system is build upon IoT, the protocols can vary from device to device and application to application. A unified central system wherein security measures can be established is absent presently. Hence, as the volume of data interchanged increases, the risks involved in security also reaches new heights.

Large number of difficulties in the area of interconnected network. In which the main ideal of the paper is to make a thread free network so we are chosen the botnet detection in the means of thread free connection. The compatibility of the network services was taken as the data. Contradiction in network services was included to evaluate the variance of the network through the detection methodology.

- A Denial of Service (DoS) attack happens when attackers attempt to prevent legitimate users from accessing the services (1 Computer). Distributed Denial of Service (DDoS), which result from a large number of systems maliciously attacking same target from different sources. This is often done through a botnet, where many devices are programmed to request a service at exactly the same time (Multiple computers). DoS wouldn't steal information or lead to a security breach, but the loss of reputation for the affected company can still cost a large amount of time and money. It is a cyber attack in which the network is stopped and often collapsed by flooding it with useless traffic and thus preventing the legitimate network traffic. DoS attack first occurred in 1974 courtesy of David Dennis—a 13-year-old student. It is the first large-scale DdoS attacks occurred in August 1999, when a hacker used a tool called "Trinoo".

- Linux.Aidra – Also known as Linux.Lightaidra, botnet which was discovered in 2012 by security researchers at ATMA.ES. It was first noticed when researchers found a large number of Telnet-based attacks on IoT devices.

- Bashlite – Also known as Gayfgt, Qbot, Lizkebab and Torlus, IoT botnet which was determined in 2014 with the Bashlite the source code published in 2015. Few variants of this botnet reached over 100,000 infected devices, serving as the precursor to Mirai.

- Mirai – Gaining worldwide attention in September 19, 2016, the Mirai botnet consisted of record-

breaking DdoS attacks on Krebs, OVH and Dyn. The botnet targeted closed-circuit television cameras, routers and DVRs, generated traffic volumes above the value 1Tbps. Ten pre-defined attack vectors, the botnet confused the infrastructure of service providers and cloud scrubbers. Some of the points include GRE floods and Water Torture attacks. Mirai took advantage of these insecure IoT devices in a simple but clever way. Rather than attempting to use complex wizardry to track down IoT gadgets, it scanned big blocks of the internet for open Telnet ports, then attempted to log in using 61 username/password combos that are frequently used as the default for these devices and never changed. In this way, it was able to amass an army of compromised closed-circuit TV cameras and routers, ready to do its bidding.

- Linux/IRCTelnet – Discovered in 2016 by Malware Must Die, this IoT botnet targets routers, DVRs and IP cameras. It can send UDP and TCP floods along with other methods in both Ipv4 and Ipv6 protocols.

- Once the desired number of devices is infected, attackers can control the bots using two different approaches. The traditional client-server approach involves setting up a command and control (C&C) server and sending automated commands to infected botnet clients through a communications protocol, such as Internet Relay Chat (IRC). The bots are often programmed to remain dormant and await commands from the C&C server before initiating any malicious activities.

- Peer-to-peer (P2P) botnet – The other approach to controlling infected bots involves a peer-to-peer network. Instead of using C&C servers, a peer-to-peer (P2P) botnet relies on a decentralized approach. Infected devices may be programmed to scan for malicious websites or even for other devices in the same botnet. The bots can then share updated commands or the latest versions of the botnet malware.

The P2P approach is more common today, as cybercriminals and hacker groups try to avoid detection by cybersecurity vendors and law enforcement agencies, which have often used C&C communications to monitor for, locate and disrupt botnet operations. Besides machine learning, IoT services are also applied to these domains. The growing complexity in IoT infrastructures is raising unwanted vulnerability to their systems. In IoT devices security breach and anomaly has become common phenomena nowadays.

2. PROJECT THEME

All IoT devices are connected to the server/peer-to-peer network. If the server have this facilities installed, it continuously monitor all devices. So, if they find any abnormalities from any connection then it immediately remove that device from the connected device.

So, the malware should not spread across multiple device.

3. LITERATURE REIVEW

There are several works done in IoT fields. Still, researchers are working in this area.

Pahl et al. He proposed a firewall and detector for anomaly of IoT services. Clustering methods like BIRCH and K-Means implemented in different microservices. Different clusters were in the same standard deviation distance. Online learning technique are updated using clustering model. In this model he immitate and obtain overall accuracy is 96.3%.

Liu et al. he developed a On and Off attack detector by using malicious network node. Would IoT network might be attacked by a malicious node when it was in initial state or on state. The network became normal when the malicious node in off state or inactive. This method is developed by using light probe outing mechanism and estimation of neighbour node for the detection of anomaly.

Anthi et al. he represented the intrusion detection system be entitled in the IoT world. He has been successfully identified many numbers of machine learning classifiers using simple forms of denial of service (DoS) attacks and scanning probing. Weka software were used for classifier machine learning.

Diro et al. he disputes the comparison study of deep and shallow neural network by using dataset. By using fog-to-things architecture he detected the four class of anomaly and attack. The system has been accomplished the 98.27% accuracy for deep neural network model and 96.77% accuracy for shallow neural network model.

Kozik et al. his classification based on the attack detection utilizing cloud architecture. In this paper he employed on the artificial netflow data by extreme learning machines (ELM) and scalable in Apache Spark cloud framework. The accuracy values of his classification are 0.99, 0.76 and 0.95 respectively. Scanning, command and control and infected host are the three significant used in this IoT systems.

Zang and Green represented IoT defense algorithm to prevent DDoS attacks by making IoT devices as intelligent as bots, while preserving a lightweight and inexpensive solution. To understand the difference between a b engine and a malicious request, a node analyses the consistency of the packet content. Although results showed that this approach helps to prevent attacks, it depends on the limited resources of every bot. A storage is missing when monitoring node to deal with the extra demand.

Ukil et.al he analyse the detection of anomaly in healthcare at the platform of internet of things. He also introduced in healthcare to use IoT sensors, medical image analysis, bio-medical sensor analysis, big data mining, predictive analyse, model of cardiac anomaly detection takes place through smartphones.

Wang et al. he proposed a peer-to-peer detection approach based on the botnets. He studied the control flows of initial intervals (10 minute). Then he measures the stability of flow and exploits property of the bots which is reveal the comparable behavior in command search and performance of their tasks.

4. PROPOSED DETECTION METHOD

In this method we propose for detecting IoT botnet attacks using machine learning. In real world, there is difficulties in the obtaining of botnet malicious datasets. In order to evaluate our system, we try to generate a network traffic traces which contain both malicious and non-malicious traffic, these traffic traces are collected from the popular network applications. Both malicious and non-malicious traffic are mixed with each other because these are occurred in same period of time. After this malicious and non-malicious traffic traces are separated into two different datasets.

In this we train and test the dataset which is extracted from the traffic data. By this we can detect the anomalies in the server.

This method consists of the following main stages: (i) Pre-processing of data, (ii) feature extraction, (iii) training and testing data.

Pre-processing of data: We used to collect data from the network traffic data using port mirroring on the switch through the organizational traffic flows. Data may have incomplete records, noise values and inconsistent data, data cleaning is used to find the missing values to recognize the correct data. The dirty data may lead to poor output. In Data reduction transform the digital information into experimentally into corrected, ordered and simplified form, the large number of data taken place to the meaningful parts. Cleaning the duplicate value, redundant data and many missing values in the data.

Feature extraction: The feature extraction are belongs to dimensionality reduction. The process of feature extraction is one of the very essential part when you mush need to reduce the number of resources that are needed for processing without any lackage of important or relevant information. In given analysis, it also reduce the amount of redundant data. Test of correlation among the feature is performed. Test of feature significance detection is done using logistic regression technique.

Testing and training data: First of all, we must separate the data into two types one is testing data another one is training data. In this dataset most important part is data mining model, the data separated for training set has to be trained with a given functionality and the small portion of the data (20%) is tested for testing. The testing data is unseen data. Trained data to fit in the model testing the data to test it to gain accuracy. Tested data make sure that how well machine can predict and give the accuracy. The relation among its input are compression ensures that the network learns the meaning full concepts. An anomaly has been observed by the given classification when apply unseen traffic model are maximise the true positive rate and minimize the false positive rate.

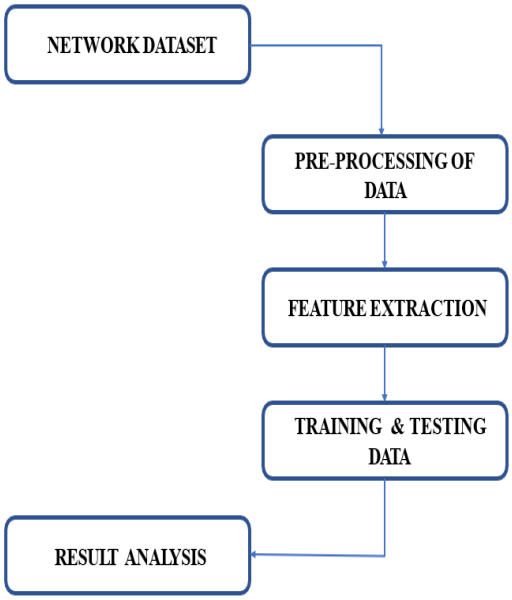


Fig.1

A.LOGISTIC REGRESSION

To detect botnets, we used logistic regression model in this work focused. In machine learning, logistic regression is basically a supervised machine learning algorithm. Logistic regression take path in two possible ways. It is also called as supervised classification algorithm. Logistic regression is almost resemblance like linear regression. The assumptions made by logistic regression about the distribution and relationships in your data are much the same as the assumptions made in linear regression. This regression must contain minimum of two target variables is popularly used to solve classification problems. It's quite related to linear regression. Three types of logistic regressions are (1) binominal (2) multinomial (3) ordinal. This algorithm works with binary data where either the attacks happened or doesn't attacks happen.

Logistic regression is based on the logistic function f (y), provided below:

$$f(y) = \frac{1}{1 + e^{-y}}$$

In logistic regression y is expressed as a linear function of n input variables:

$y = \beta 0 + \beta 1\ x1 + \ldots + \beta n\ xn$

Then, based on the input variables x1,x2,…,xn , the probability of an event is shown below

P(x1,x2 ,…,xn-1)= f (y)

In this, the logistic regression model has been used to estimate the probability that the device that initiated connection was a part of IoT botnet. In order to build the model, data of 100 botnets oriented to IoT devices and performing brute-force attacks to increase their scale was collected and employed.

To create the logistic model the following parameters were selected as predictors:

Mean Interval between requests. Mean interval between requests is not very large with a small deviation.

Mean size of packets. A packet is not large enough, because it contains a default username or a password.

The packets always contain default user credentials and it is usually alphanumeric.

Variance of the packet will change when the botnet is affected.

Even number requests. Most of all the number of requests is even.

The malicious requests send from host has at least one open port.

B.Z-SCORE NORMALIZATION

Z-Score normalization is a strategy of normalizing data which avoids the outlier issue. The formula for Z-Score normalization is below,

$$Z = \frac{Value - \mu}{\sigma}$$

where, μ - Mean value of combined data

σ – Standard deviation of the combined data

If combined data is exactly equal to the mean value of the combined data, then it normalized to 0. If it is below the mean, it will be negative and above the mean it will be positive. The size of the negative and positive numbers was determined by the standard deviation of the combined data.

C. PSEUDOCODE

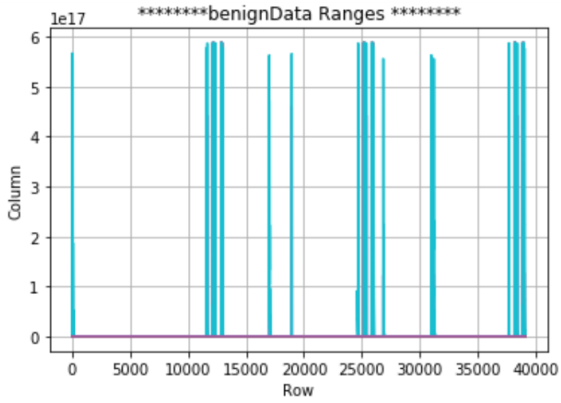
1) Import libraries (Numpy, Pandas)
2) Import data/dataset
3) Check dim (Rows/Col)
4) Assign the attributes as 0/1
5) Combine two datasets
6) Final output is “out”
7) Convert single dimensional array from n-dimensional array
8) Normalize the data using z-score normalization
9) Import sklearn model for logistic regression
10) Train and test data using test-train.split
11) xtr -> x-axis training | ytr -> y-axis training

    Xte -> x-axis testing | yte -> y-axis testing
12) fiting model
13) Find accuracy (using testing model) -> (xte,yte)

5. EXPRIMENTS AND RESULT ANALYSIS

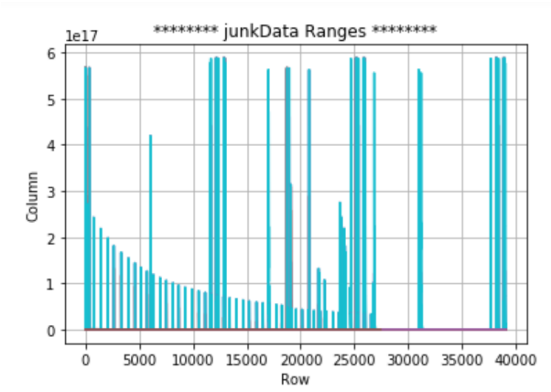
The set of data was collected in the lab which is corresponding to the IoT devices. It was divided into two data ranges one is benign data and another one is junk data. In this benign data is full of non-attacked or non-malicious data set which is collected from the network.

The data ranges are shown in the figure (a). The pure data are recorded in the dataset from the data network.



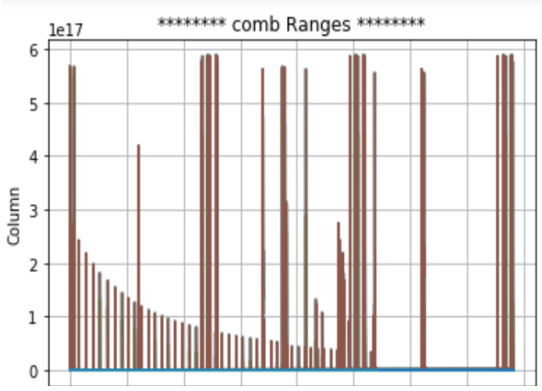
(a)

In junk data malicious data set is present which is collected from the network. The data is full of impure and attacked data. These data has malicious records taken from the network. The data ranges are shown in the figure (b).

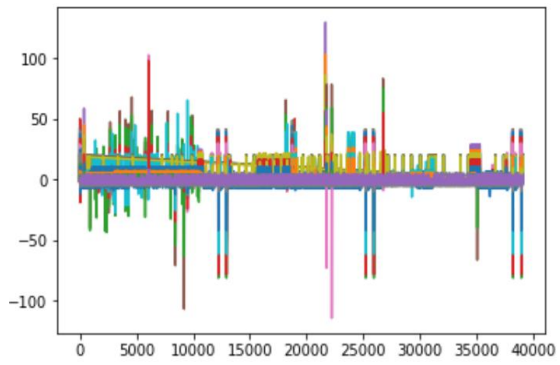


(b)

It is the combination of the both benign and junk data. Which means malicious and non-malicious dataset is represented as the graph. The attacked data and non-attacked data are shown in the figure (c) and (d).



(c)



(d)

6. CONCLUSION

Finally, to sum up the botnet attack detected by using machine learning algorithm. By most probably this year attacks and illegitimate action has been full tilt proliferate rapidly. The large amount of internet connected device in sense IoT device is envisage to grow exponentially to billion. Although, the information security of these device still remains lacking in security set up, there is an unauthorized access of data and loss of data. Thus, the security system aren't strong Enough still security issue are here it perhaps to increase in IoT botnet attack such as massive DDoS attack employed by cybercriminals. To beneficially conserve against IoT botnet, essential techniques are used to detect them. One of them we are applied in this paper machine learning algorithm to analyse whether the device infected nor not. The propagation stage was discussed in this paper. The accuracy of this detection is comparatively high when compare with other techniques. The dataset is trained by virtual environment data. An empirical study we mainly focused on the real -time problem face in our day to day life. We postulate that the predictability of infected device from server side with an accuracy and precision. The provided detection of IoT botnet are applicable model. Therefore, we approach pre-processing techniques and machine learning proposed in this paper have a substantial accuracy in IoT botnet.

REFERENCES

[1] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and Other Botnets," Computer, vol. 50, no. 7, pp. 80–84, 2017.

[2] Junyi Liu, Shiyue Liu, Sihua Zhang "Detection of IoT Botnet Based on Deep Learning" July 27-30, 2019.

[3] Christopher D. McDermott, Farzan Majdani, Andrei V. Petrovski "Botnet Detection in the Internet of Things using Deep Learning Approaches" 2018 International Joint Conference on Neural Networks (IJCNN)

[4] M. Ozcelik, N. Chalabianloo, and G. Gur, "Software-Defined Edge Defense Against IoT-Based DDoS," in 2017 IEEE International Conference on Computer and Information Technology (CIT). IEEE, 8 2017, pp. 308–313.

[5] S. Hilton. (2016) The DDoS that didn't break the camel's VAC. [Online]. Available: https://dyn.com/blog/dynanalysis-summary-of-friday-october-21-attack/

[6] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "Ddos in the iot: Mirai and other botnets," Computer, vol. 50, no. 7, pp. 80–84, 2017.

[7] Akami. (2017) Threat Advisory Internet of Things and the Rise of 300 Gbps DDoS Attacks. [Online]. Available: https://www.akamai.com

[8] B. Krebs. (2016) KrebsOnSecurity Hit With Record DDoS. [Online].Available:https://krebsonsecurity.com/2016/09/krebsonsecurityhit-with-record-ddos/

[9] Anton O. Prokofiev #1, Yulia S. Smirnova#2, Vasiliy A. Surov "A Method to Detect Internet of Things Botnets" 2018 IEEE.

[10] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. Robinson, "Deep learning for unsupervised insider threat detection in structured cybersecurity data streams," in Artificial Intelligence for Cybersecurity Workshop at AAAI, 2017.

[11] Prokofiev, A. O., Smirnova, Y. S., & Silnov, D. S. Examination of cybercriminal behaviour while interacting with the RTSP-Server. 2017 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), St. Petersburg, 2017, pp. 1-4. DOI: 10.1109/ICIEAM.2017.8076437

[12] Dobbins, R. Mirai iot botnet description and ddos attack mitigation. Arbor Threat Intelligence, 2016, 28.

[13] A. Tuor et al., "Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams," Proc. AAAI 2017 Workshop on Artificial Intelligence for Cybersecurity, 2017; https://arxiv.org/pdf/1710.00811.pdf.