JCRES
Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 5- Issue 1, Paper 38 January 2022

A Study On IoT - Botnet Detection Techniques

Rahul Sharma, Khyati, Harshit Singh, Chirag Joshi

School of Computing, DIT University, Dehradun

School of Computing, DIT University, Dehradun

Abstract **-**With a rapid growth in the area of communication, it is easy for people to connect with each other. The accumulation of data is also increased exponentially and it also facilitate many people around the world. Access to information is easier than ever before, which increases the amount of cyber attacks. Many attacks, such as viruses, worms, and trojan horses, have occurred in recent years. Attackers conceal themselves using many strategies, one of which is the use of a botnet. They are preparing and deploying Botnet-based attacks. Botnet assaults are not just aimed at computer systems, but also at IoT devices. One current attack that must be addressed is the IoT-Botnet. In this paper, we have studied various methods and techniques for the detection of IoT Botnet. We have also compared different methods on the basis of their methodology, dataset and performances. The study tells us the important of IoT Botnet and its detection methods.

I. INTRODUCTION

Computers are among the basic necessity in today's era. Abacus is believed to be the first computer, and hence the history of the computer begins [1]. Starting with the 1st gener- ation of computers, today is the time we have reached the 11th generation of them, which explains the development in terms of size and speed(citation). Computers, a device for storing and

A.History of Botnet

The term 'bot' emerges from 'ro-Bot' which refers to a script or set of scripts for pre-defined functions in an automated fashion [3]. In general terms, botnet refers to the set of connected devices affected by an attack, the devices affected are commonly known as bots and the whole controller of the attack is common as a bot-master or the bot-herder. The existence of these botnets has been traced several years ago, but in the modern era, it has taken the interest of the research community to a vast extent.

In August 1988, Jarkko Oikarinen University of Oulu, Finland, invented the first bot IRC [4] . 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 was Pretty Park in June 1999, against the first malicious bot named as GT-Bot [2]. 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" [5]. In this attack, the hacker hacked all the files and other information and for a handsome amount of $7,50,000 [5]. Another kind of crimes refers to the credit card frauds which are best described in [42]. For this purpose, a special fraudulent detection was used which was based upon Intelligent ML technique.

A variety of deep learning approaches are being developed for knowing the developmental factors of botnet and for the multiclass botnet identification [6]. In the early 1990's, the

manipulating data, also perform a wide range of tasks. It is media's attention to lesser-known worms, such as the Father an integral part of modern society, a programmable electronic Christmas Worm and the Worms Against Nuclear Killers, device used to process, store and retrieve data [1]. The two combined with the development of far-reaching tools and the different families of these devices, the digital computers, based proliferation of Internet-connected machines. Thus, the first on binary digits are the modern computers as compared to non-malicious IRC bot, Eggdrop, the first violent IRC bulletin the analog computers reflecting the properties of data being published in 1993, was designed to control and secure chat modeled. channels [7]. IRC channel. In 2003, Oregon hackers took

The tremendous increase in the use of these computers has control of 20,000 botnet hosts and launched DDoS [8]. Future resulted in a massive increase in cyber-attacks which results in and Current concerns include the ever-expanding Internet of exposing sensitive data which can be misused for any non- Things, vul- nerabilities in network botnets, and the potential social cause. According to research around 40% of the 80 for machines to become botnet hosts [9]. million computers are affected by some or other attacks [2].

The classification of these attacks into several types including *B.Defense Mechanism* malware, phishing, Denial of Service (DoS), etc and their The Botnet is one of the most tremendously growing attacks in symptoms including lack of speed, unusual shutdowns and the current scenario and hence tracking and detection are a many more confirms the existence of such threats. trending topic of research in current scenario. Applying tradi- tional techniques to IoT is quite difficult due to large volumes of

data, real-time scenarios, and many other characteristics, thus the modern solutions require large memory requirements and heavy weight computations. There exist 2 approaches for such detection an dtracking:- [10]

Setting up honeynets (generally used for understanding the technology and characteristics).

Passive network traffic monitoring and analyzing (numer- ous other classifications including signature-based, anomaly-based, mining-based, and many more).

Another technique is based on the Logistic Regression which is used at the propagation stage. It is based on the probability based on values of a collection of variables – predictors. It works by performing brute-force attacks on TEL- NET or SSH services [11]. The problem of the concept of erosion of the patterns of error changes over time, is among the most common problems in IoT Botnet [10]. The work done on finding IoT Botnets is usually unable to detect bots using advanced escape strategies such as dynamic IP addresses and URLs in a manner similar to a Domain-name generation algorithm (DGA). The cShield used as a distributed and expandable solution provides recommendations to the network that holds its own responses to threats [12].

C. IOT Botnet

"Internet of Things" (IoT), the name was given by Kevin Ashton in 1999, when he was working with Auto-ID labs [1]. 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 [2]. 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 [3]. IoT bot is generally a robot that scans for viruses and converts into a bot if any virus is found. For checking, it passes through a sequence of stages which includes:- [4]

- checking for initial and secondary infection
- establishment of Command and Control (C&C) server
- dealing with the attacked part
- working on up-gradation and maintenance

Working upon the centralized and P2P architecture of IoT Botnet, it helps the botmaster for monitoring and installs malware updates.

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

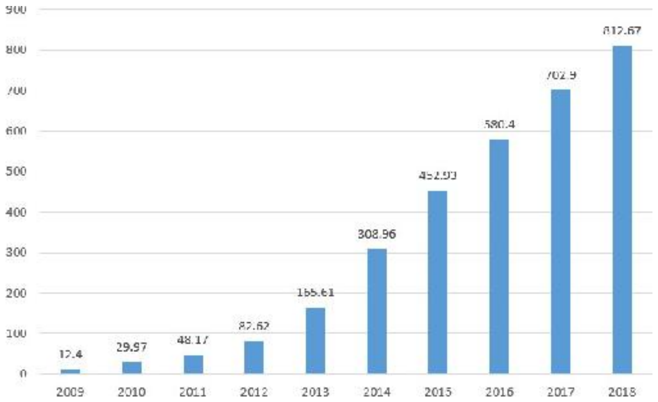

Fig 1. Statistics regarding growth of IOT Botnet

Botnet formation being one of the attacks that activates at a very fast rate and harms the server [16], the main focus of this paper is towards a brief study about the types of IOT attacks, their architecture, life-cycle, life-span and their defense mechanisms. The concepts and methods of emsemble learning techniques are widely used in this area [17] and the major focus is displayed over this concept. The techniques of ensemble learning involving machine learning techniques, mining and various others are displayed in a very effective way for better understanding of the concept. The types of attacks as defined in the paper gives a clear view of how it affects a system and the pre and post consequences of this attack.

The description of the paper is defined as follows: Section(2) defined the background work of botnets. Section(3) defines the architecture of the botnet including the various types of models used in these architectures along with the diagram- matic explanation. Section(4) involves the defense mechanism techniques of IOT Botnet which are used for the detection of IOT Botnet attacks. Section(5) includes a brief discussion about IOT Botnet, it's composition, development and the stages of life it passes from. The paper ends with Section(6) which includes the total work done in this paper as result and conclusion.

II. RELATED WORK

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 [3], a network of devices connected and attacked forms a background of botnet. August 1988, a date to remember [4], gives the description about the origin of botnet from the first bot IRC to it's development stages going on. According to Kevin Auston [13], 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. Various researches have been made for the detection of IOT Botnet which gives us various approaches for the same, the 2 base methods are given by Mr. Sandeep Sonaware[10] . The concept of drift was focused to a greater extent as it is widely helpful in detection [18]. Detection based on Artificial Immune System [19],the DNS based detection[20], and various others are broadly explained in this paper.

S No.	AUTHORS	Dataset used	Signaturebased detection method	Mining based detection method	DNS based	AIS based	Machine learning based detection method	Adaptive learning based	DGA based	DOTA based	DDoS based	C & C based	Autoencoder based	Cloud Computing based	Net Flow based

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

1	Amit Kumar Tyagi et al [4]	Real Life Dataset	✓	✓			✓					✓			
2	Mehdi Asadi et al [48]	ISOT,Bot-Iot					✓					✓			
3	Nickolaos Koroniotis et al [47]	BotIot,ISCX,TUIDS,KDD99					✓					✓			
4	Farhoud Hosseinpour et al [14]	Real Life Dataset				✓									
5	Smita Dange et al [10]	Real Life Dataset	✓	✓	✓		✓								
6	Somayeh Soltani et al [13]	Benign dataset								✓					
7	Kamal Alieyan et al [49]	Bot-Iot dataset													
8	Anton O. Prokofiev et al [6]	Mirai dataset					✓								
9	Zhou Shao et al [12]	Benign traffic and Botnet traffic						✓							

10	Carlos A. Rivera A. Et al [7]	Benign traffic and mixed traffic							✓						
11	Ruchi Vishwakarma et al [17]	Bot-Iot dataset					✓								
12	Michele De Donno et al [18]	Mirai								✓	✓				
13	Harry Owen et al [22]	UNBS-NB		✓								✓			
14	Tie Luo et al [24]	WSN											✓		
15	Shun Tobiyama et al [26]	Benign dataset					✓								
16	A. Geoge et al [27]	Real Life Datset												✓	
17	Saeed Abu-Nimeh et al [29]	Real Life Datset					✓	✓							

18	Lakshya Mathur et al [30]	CTU-13, ISOT					✓								
19	Hao Zhao et al [31]	MedB-IOT	✓	✓	✓		✓								
20	Pamela BeltránGarcía et al[5]	PsyBot, SSH					✓				✓	✓			
21	Pavel Celeda et al [6]	PsyBot										✓			✓

Table 1:- Comparison with Existing Literature

III. BOTNET ARCHITECTURE

A Botnet, being traditional or being an IoT Botnet, shares the same architecture, i.e centralized and P2P. With command line infrastructure and botnet, 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 botmaster and bots. An excellent example of this is the IRC-based model, in which one channel is used for communication. 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.

3.1 **Centralized model**

The offensive architecture is divided into different cate- gories, the models defined within this architecture are as follows [40]:-

I. Agent-handler model

- This model comprises clients, handlers/masters, and agents/bots. The structure is defined below in fig.2.
- The client refers to a hardware system that is used by an attacker for communication purposes.
- The handlers refer to a piece of software which are led into the Internet and helps in the communication process.
- The agents/bots refer to the malicious/infected software running on an infected machine.

Handlers are led at a place containing a large amount of traffic, such as routers or servers so that it is harder to detect, but here the handler and agent, both need each other’s identity for communication, so if a single bot is discovered it may lead to the detection of the whole botnet at a very easy rate [23].

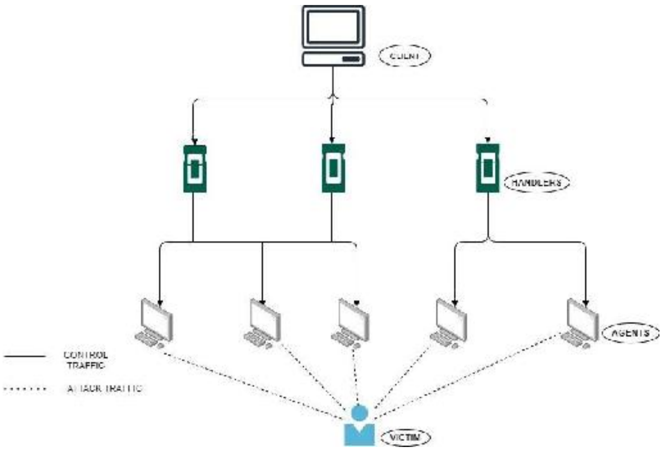

Fig. 2 :- Agent-Handler Model

II. Reflector model

Here is the addition of one more element as compared to that of the agent-handler model, i.e, the reflectors. The structure is defined below in fig.3. In this, the traffic is sent to other infected machines (reflectors) instead of directly sending them to victim. In this, the main role is of IP, which means that the source IP is replaced with a fake IP (victim's IP) to mislead and create a huge traffic towards the target.

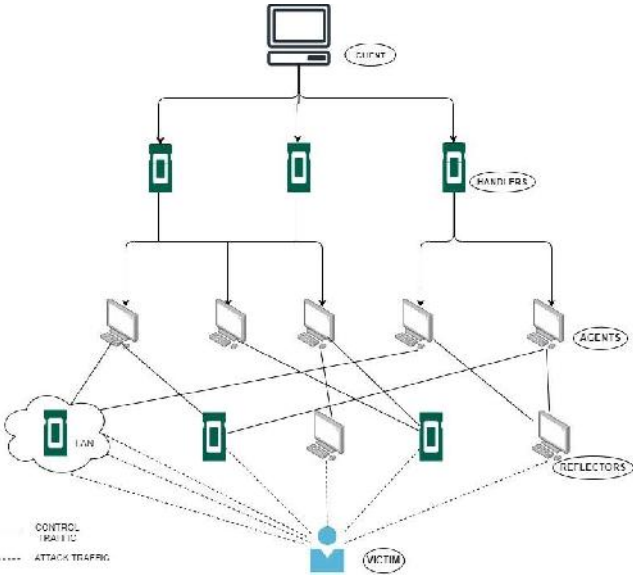

Fig 3:- Reflector Model

III. IRC Model

It stands for Internet Relay Chat-based model. Here, instead of handlers, a textual protocol called an Internet Relay Chat channel is used as command and control for communication between clients and bots, else is the same as the agent-handler model. The structure is defined below in fig.4. It is a multichannel or multi-user chatting system at the application layer. The several benefits of this channel are as follows [10]:-

- The use of legitimate ports for communication makes the traffic higher, thus marking the low traceability.
- Logging onto IRC server creates a list of all the servers available, thus there is no need to make any such list physically.
- An easy file sharing benefit is available which helps the attackers, the security of secondary victims to participate in the attacks.

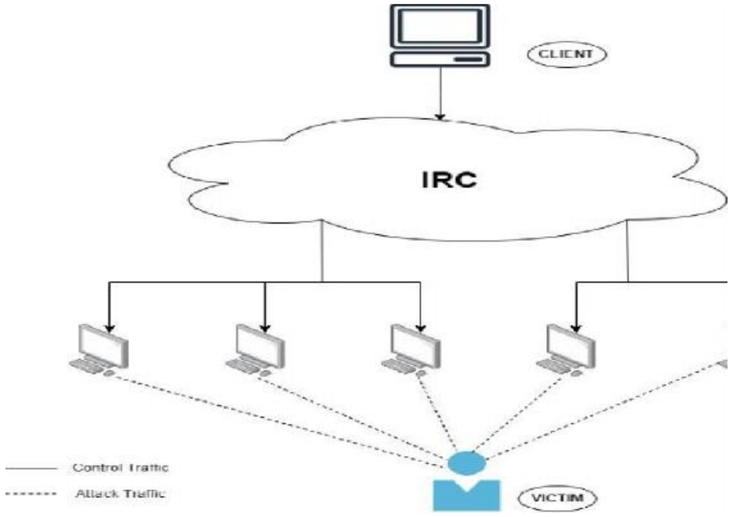

Fig 4 :- IRC/Web Based Model

IV.Web Based model

Here, the architecture is the same as that of the IRC model. The only difference in this architecture is that the Internet Relay Channel is replaced by a website, which is used as a C&C server for bots. The structure is the same as that of the IRCbased model, which is depicted in fig.4. The control traffic here can easily be said as legitimate as the ports used for web browsing are common, i.e 80 and 443, which are used for communication. It is highly comparable to calling someone from the website while the website is out of service [11]. It has advantages over the IRC model are as follows:- Ease of setup and website configuration.

- Less bandwidth and increased command and reporting functions.
- Botnet hijacking resistant via chat-room hijacking [11].
- Ease of acquisition and use.

3.2 **Decentralized / P2P model**

It stands for the peer-to-peer-based model. Here occurs only a network of agents between the client and the victim, with no single point command and control. It comes under decentralized architecture. The structure is defined in fig.5. It reduces the chances of tracing to a great extent as in centralized it is possible to uncover some part of infrastructure by detecting one agent due to intermediate nodes, but here commands are sent to a few bots, then further to a few bots, and so on, thus reducing the chances of tracking. It can be said that a

P2P is a kind of computer network where several computers or devices are connected for sharing of resources by direct exchange of data instead of exchanging via a server. The main factors affecting this model are as follows [12]:-

- Effectiveness – refers to how powerful a P2P model can be while launching an attack.
- Efficiency – is how much time the command can be issued to all the members.
- Robustness – how resilient a botnet is to failure.

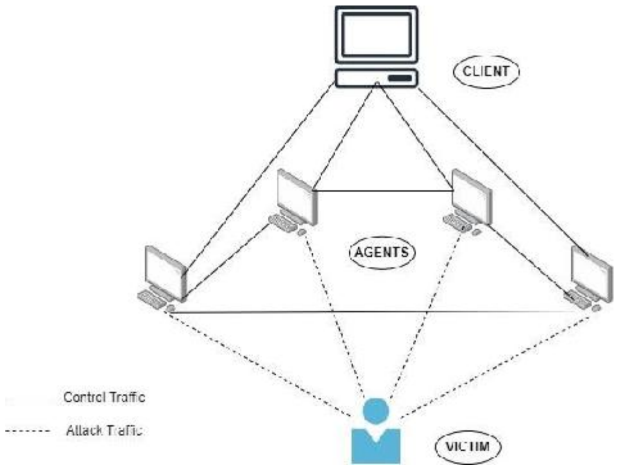

Fig 5 :- Decentralised / P2P Based Model

3.3 **Hybrid model**

It is a combination of both centralized and decentralized technologies, as the name suggests hybrid. The structure is defined in fig.5. Here, there are two types of bots, some of them referring to as both clients and servers while some only as clients. This maintains the high latency of the message. The best way for a hybrid IoT botnet is to use centralized topology as a part of communication and P2P as another part.

IV. METHODOLOGY

Botnet tracking and detection is a key concept of research in current scanario. Before moving towards the concept of defense mechanisms, it is necessary to have sufficient knowledge about the types of infection mechanisms, i.e, how a system can be affected by a particular bot. As discussed by Basudev et al, there are three main methods for bot propagation [13]:-

- **Web Download** – These are very common types of bot nowadays, as many people regularly download some or other things from the web, which may contain malicious data, which acts as a botnet-like structure and affects the functioning.
- **Mail attachments** – Emails are a major field of work in the current era, every day our mail is filled with many emails, which may contain bots in the form of mailing worms. Spam technique in the mail helps to reach out and identify them.
- **Automatically scan, exploit and compromise –** The bots automatically infect the hosts with vulnerabilities.

JCRES
Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

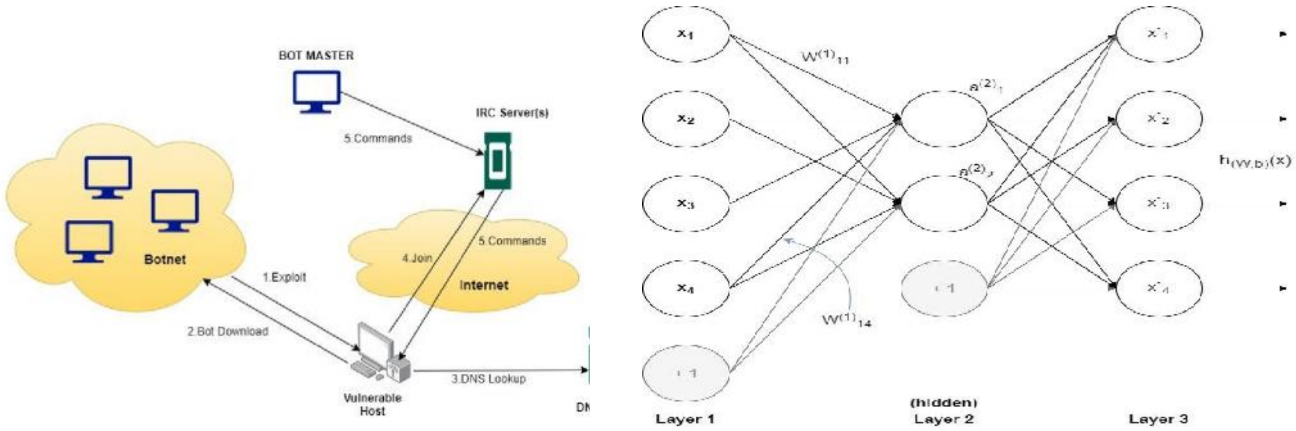

Fig 6. Botnet infection

There exist 2 approaches for the detection and tracking of such bots:-

- Setting up honeynets (generally used for understanding the technology and characteristics).
- Passive network traffic monitoring and analyzing (numerous other classifications including signature-based, anomaly-based, mining-based, and many more).

Examples of such detection mechanisms include Anomalybased, Botnet based, Fuzzy-logic based, phishing-based detection, and many more. Signature-based detection by Chen et al, for assisting the detection of anomalies states its drawbacks for implementation on large corporate networks and changes to network communication [14]. Various bot detection tools including SLINGbot, Rahim, and Bin Muhaya use a Botnet Scenario Driver (BSD) for the controlling of botnet and a Composable Botnet Framework (CBF) for ensuring the connection between the information of routing and botnet[14].

4.1 Machine Learning Based Detection

Keeping in mind the concept of Internet of Things (IoT), machine-based learning plays a very important role in botnet detection on the type basis of anomaly. The algorithm used here is the auto-encoders, where neural networks are used for anomaly detection[15]. This technique presented by Luo and Nagarajan helps us to know about encoders which are referred to as the deep learning models, traditionally used for recognition of image and other problems of data mining like spacecraft telemetry data analysis[16]. The structure being very simple performs at it is topmost level because it is an inherent power in reconstructing the data as input.

Fig 7 :- Auto-encoder Neural Network

The algorithm is designed to be a two-part kind of algorithm, residing on the IoT cloud and sensors [16]:-

- Detection at sensors without communication with any other sensors or cloud in a distributed manner.
- Cloud for the handling of computation-intensive learning tasks as the communication between cloud and sensors is at a much lower frequency than sense. The proposed distribution anomaly detection based on the concept of encoders has received the highest accuracy rate and the lowest rate of false alarms. Besides the minimum amount of computation and communication requirement, this algorithm also possesses an advantage from unsupervised learning features that the encoders have and dissolves the common challenge that a supervised machine learning model often lacks[16].

A. *Signature - Based Detection*

This technique makes use of the signature of current botnet for the detection. It is a type of intrusion detection system, which can be considered a well-known Botnet detection technique in each group and monitors trafficking of network and finds signs of intrusion [17]. It is only capable of the detection of well-known botnets and also botnets with slightly different signatures may get missed and a zero-day bot attack cannot be detected [18]. The most popular technique for this kind of detection mechanism is the 'Rishi', which matches the known nickname patterns of IRC bots [18]. They use a scoring system for the detection of bots having different channels for communication and an n-gram analysis, which are not detected easily by classical intrusion detection systems.

B. *Decision – tree Based classification*

A prediction on data of input involving network traffic is conducted and then selection of features is done for the creation of decision trees to find out whether a botnet or malware is present or not [14]. The concept of feature extraction has a various amount of hidden layers which contribute to the performance of the algorithm and the overall outcome of the prediction. It makes the dataset small by reducing the size to a limited amount of data features, which makes the accuracy much higher [19].

C. CLUSTERING

In general terms, clustering refers to the division of elements into different groups containing the same type of data. Here, it means the division of data points into different groups, so that similar data types are in the same group[43]. Providing both, simplicity as well as accessibility, there are various ups and downs in the mechanism such as involving data into only numerically valued data which becomes difficult to understand and thus is not much preferred by analysts.

D. LOGISTIC REGRESSION

A botnet detection algorithm that works on the concept of binary digits, i.e 1 and 0 which means true or false. If malware is found in the data it is marked as 1 or else as 0. Kumar et al have organized the data in the form of a
confusion matrix [44] **:-**

	Positive Prediction	**Negative Prediction**
Positive Class	True Positive(TP)	False Negative(FN)
Negative Class	False Positive(FP)	True Negative(TN)

Table 2:- Botnet classification confusion matrix

- True Positive (TP) is the case of actual malware.
- False Positive (FP) is the case of a non-botnet being mistakenly referred to as a botnet.
- True Negative (TN) is the case correctly referred to as non-botnets.
- False Negative (FN) is the case of an actual botnet being mistakenly treated as non-botnets.

The two main concepts which are in use here are [[14]:-

- **Precision** – Defines how precise a matrix is by dividing the true positive by the sum of both true and false positive, i.e, P = TP / (TP + FP).
- **Recall** – It is obtained by dividing the true positive by the sum of true positive and false negative, i.e R = TP / (TP + FN).

It is considered an effective solution for classification of botnets, as the confusion matrix gives a whole sum detail of whether the given data contains a botnet or not.

E. Fog Based Detection

As the name suggests, it creates a fog-like structure and covers the whole data storage within the network which here works like fog. It is highly compatible with IoT systems and has a very fast response time. It is a kind of framework that requires all the data to be accessed within the framework, which may lead to the revealing of confidential information to unauthorized users [20].

According to research, it is found that it is a very timeconsuming process due to the huge amount of network bandwidth. One way to move out of this problem is to delete programs or files using large quantities of bandwidth, which increases the risk of deleting the files which may be assets to the organization [14]. Hence, it becomes the responsibility of the developer to take care of which files or programs are to be deleted.

4.2 FUZZY LOGIC BASED DETECTION

According to Wesley Chai Fuzzy logic is a computer program based on "true degrees" rather than the common "true or false" (1 or 0) Boolean concept on which the modern computer is based. Artificial Intelligence (AI) has used the mysterious mind to mimic human thinking and understanding [45].

Harry et al [14] concluded that the newly developed features allow the system to detect any confusion within the network and make connections to its site and access botnet activity on its own. Chirag et al [46] suggested how they extract new features from existing databases, with a very limited number of features. To analyze the features they have proposed The Artificial Neural Network (ANN) and to use the add-on method. Identified features from the database can be converted into ambiguous features. And the width of the feature is low, medium, and high. When the release is made all new features are tested for their role in acquiring Botnet. Then enter them in the final database if the participation rate exceeds the specified limit value.

4.3 PHISHING DETECTING

Phishing is among the most common types of botnet attacks or malware that can occur in any system or a website and can cause harm to the system. In general terms, it can be said as a kind of electronic identity theft that comprises a combination of fake website creating methods and social engineering so that the user may disclose his/her confidential or valuable details, said Moghimi and Varjani[21]. Here, a pre-processing methodology is used for filtering the unwanted data and the data redundant in Machine Learning. Thereafter comes the concept of featuring the data according to the features described in the model [14]. The final determination of the result is based upon the true positive and true negative values which determine whether a website is affected by phishing or not.

Another method for this kind of detection is the detection with the help of a neural network. It is not a very common kind of technique because it is implementation is quite difficult as compared to other algorithms as suggested by Abu-Nihem et al [22]. The example of the feature extracted here can be referred to as the poorly spelled or grammatically incorrect messages starting with the phrase "Dear Customer" [14]. Another important feature of spam filtering is very important as it classifies the incoming emails as spam or actual emails. It uses the concept of logistic regression for this purpose. As logistic regression is based upon the concept of binary digits, here spam emails are marked as (1) and real emails are labeled as (0) values. The results are shown with the help of a confusion matrix as shown in table 2 [14][44].

Here again, precision and recall are used for incorrect classification.

4.4 MINING BASED DETECTION

Another mechanism for botnet detection is the detection via mining of network traffic flow. According to the model proposed by Mathur et al[23], qualities like the diversified data, and generalized detection of different kinds of botnet which results in a high level of accuracy in a timely and precise manner. The mechanism is divided into three
parts:-

A. *DATA COLLECTION*

To ensure a high level of accuracy, it is very important to test and train the data with a wide variety of sources containing different types of a botnet. In the experiment[23], nfdump and nfcapd were used while working in VMWare software upon different operating systems to obtain the network data. The other datasets used

V. IOT BOTNET

Internet of Things (IoT) is a network built by several devices connected physically to the Internet. Currently, around 20 billion IoT devices are present in the world and according to research [24], this number will reach around 50 billion by 2025. The security concern of these devices is to the least extent as it could be as the manufacturers have not taken keen interest keen interest in this field but in increasing the production of the devices. The network attacks in this context include Distributed Denial of Service (DDoS) attacks, leaking sensitive information, spam, and more. Botnets, considered a major source of cyberbullying, are a common source of malicious activity on the aforementioned networks.[24]. With the increase in the years and the development of the technology, botnet attacks have increased from the initial stage, i.e attacks over PCs and other devices to the current stage including attacking extensively based on various network equipment.

for network traffic data collection were:-

- CTU-13 dataset (labeled dataset with traffic consisting of 13 scenarios and 14 different captures of botnet samples).
- ISOT dataset (combination of malicious and nonmalicious data in pcap format).

B. MACHINE LEARNING ALGORITHMS

The most effective classification techniques which have been used in the experiment[23] and gave maximum results are described below:-

- Logistic Regression – As described in the above section, it is based upon the binary concept, i.e 1 or 0.
- Random Committee – an average of all the predictions made by the base classifier, keeping in mind the combination of seed values and base classifiers.
- Random Subspace – A method based on decision tree containing multiple trees constructed over randomly chosen spaces for good accuracy.
- Multi-class classifier – recording into 3 or more classes and applying output codes of errorcorrecting for increased accuracy.

C. FEATURE EXTRACTION

In experiment[23], filter and wrapper techniques are used for feature extraction purposes, where the wrapper technique helps in the optimal subset of features fitting in an algorithm while the filter technique helps in measuring the highest predictive power among the subsets. It is found that all the attributes do not equally participate in the result, some are very relevant while some cause redundancy.

Time	Botnet	Influence
2008	Linux/Hydra[5]	The earliest targeting IoT Botnet devices with opensource code and propagation mechanism for DDoS function.
2010	Chuck Norris[6]	Majorly for Linux devices along with Brute Force attacking. D-Link router authentications were added.
2012	Light Aidra/Aidra[24]	Found while the invasion of Botnet named Carba. Supports multi-system architecture and has open-source code on Github.
2014	Linux.Darlloz[24]	Use of PHP vulnerabilities for infection over more than 31,000 IoT devices and prevent Telnet login.
2016	KTN-RM/Remaiten[25]	A combination of malicious codes based on IRC agreements to control Linux devices to launch DDoS attacks. More than 1 million devices are infected here.
2018	Torii[24]	Modular structure, rich functions, get instructions for use with encrypted multi-layer communication.
2019	Mozi[7]	Combines at least 3 malicious codes (Gafgyt, Mirai, IoT Reaper) and builds a P2P botnet.

Table 3:- IoT Botnet Development

5.1 DEVELOPMENT OF IOT BOTNET

5.2 BOTNET STAGES

Botnet performs its function in 3 important categories: - [5]

- **Infection**: An unsuitable computer program for hiring new bots that spread the word of risk, downloading emails, installing software and hosting becoming part of the botnet. • **Command and Control**: Once the system is infected, the bot communicates with the botnet controller to receive any command.
- **Malicious Activities**: - Attack attacks such as DDoS, spam, etc.

5.3 BOTNET COMMUNICATION

The exchange of messages takes place in a sequence that is needed to achieve a specific task:-

- **Coordination:-** Instruction to the bot about the tasks to be performed.
- **Scan:-** It refers to the IMP/ TCP/ UDP scan for vulnerable services.
- **Data:-** It refers to arbitrary files, binary bots, or network data that communicates to the botnet.
- **Register:-** Required messages for the record.
- **Execute:-** Executed tasks by the bot.

54. COMPOSITION AND STRUCTURE

The IoT Botnet architecture is mainly divided into 4 parts:-
[24]

- **BOT: -** Infected device is used as a platform for continuous detection of endangered IoT devices and expansion of botnet resources. Detects and executes attack commands and other functions assigned to the C&C server**.**
- **C&C Server: -** It is completely under the control of the botnet administrator, who sends the command to the botnet and usually receives a transaction response.
-

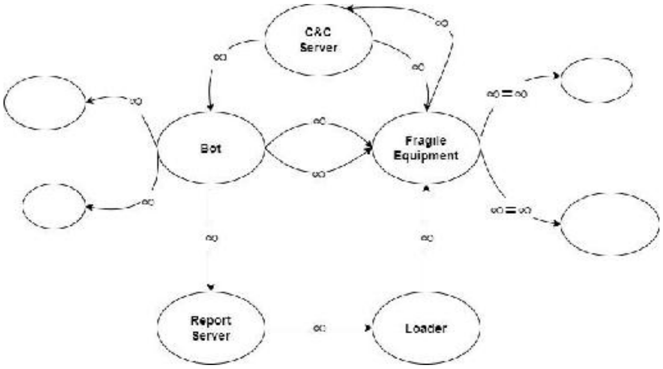

Fig. 8 :- Construction Process of IOT Botnet

5.5 LIFE CYCLE OF BOTNET

The Life cycle of IoT Botnet contains stages of detection, infection, program download, and command execution, as shown in fig. 7. The specific description of the stage is as follows: [24]

- The program uses a controlled terminal for open port detection and detects 22, 23, and various ports operated via remote control.
- After this, methods like username and password blasting login are invaded.
- After login, the device information is reported to IP and then provided to the loader.
- The Loader then obtains shell permission and executes and downloads bots of corresponding version of architecture.
- Once the program runs successfully, it connects to the C&C server to become a registered bot member.
- This C&C server issues the command attack to the bot.

CONCLUSION AND RESULT

The command usually contains the type of attack, the address of the attack, and the duration of the attack.

- **Uploader: -** With the new IoT machine, the bot will find the processing architecture and will direct you to download the corresponding binary file to the uploader.
- **Report Server: -** Usually used to store information and status of all infected devices, including Internet address, port number, processor configuration, and login details.

	Accuracy	Precision	Recall	F-Measure	Error
Anton O. Prokofiev et al [6]	97.30%	94%	98%	96%	NA
Mehdi Asadi et al [48]	99.63%	59.82%	NA	55.64%	0.37%
Lakshya Mathur et al [30]	98.4%	NA	NA	NA	0.4%
Mehdi Asadi et al [48]	99.39%	94.12%	NA	96.97%	0.61%
Shun Tobiyama et al [26]	89.3%	99.41%	NA	NA	NA
Saeed Abu-	93.58%	93.26%	83.5%	88.005%	NA
Chirag	99.94%	99.92%	99.96%	NA	0.005%

Table 4 :- Comparison of different algorithms and their accuracy

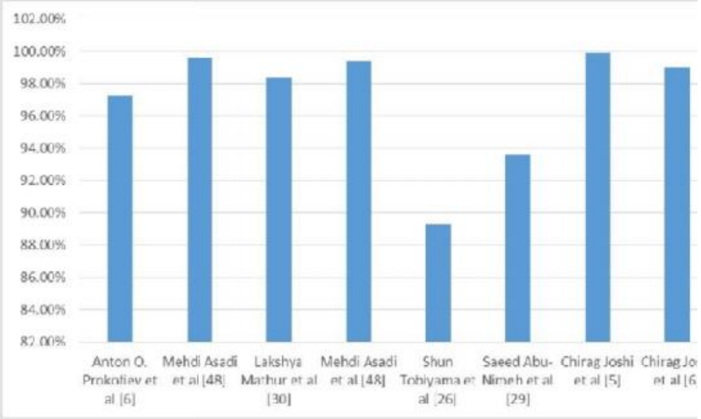

Fig. 9 :- Analysis of Accuracy for different models

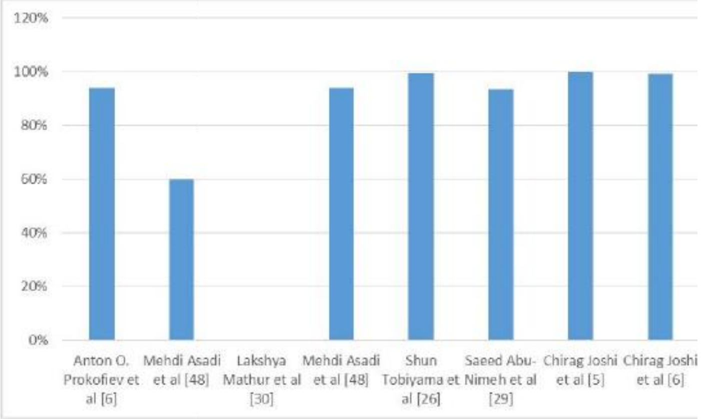

Fig. 10 :- Analysis of Precision for different models

The paper analyses the accuracy, precision, recall value, f-measure, and error for different methodologies used in different papers. By the various observations, we came to the conclusion that the best result among all the methodologies and dataset, we dealed with in this paper, is given by Chirag Joshi et al [5], while working with CTU-13 dataset. The accuracy of 99.94%, along with the precision value of 99.92%, recall rate of 99.96% and the error of 0.005%, is one of the best results, we found while working on this paper. Though this model gives best result. But it has some of it's limitations and also it is required to be tested among various other datasets.

In this work, we have summarized the IoTBotnet detection methods and the architecture. In this paper, we have address most of the issues of IoT-Botnet and also their detection methods. The detection methods ranging from Signature based to Deep Learning based are discussed in this study. Our study will help in finding the working, types and detection methods of IoT Botnet. We have also compared the performance of different detection methods with the dataset used in those methodologies. The study also concludes about the best performing methods in IoT Botnet detection.

REFERENCES

[1] G. O’regan, “A Brief History of Computing.”

[2] Institute of Electrical and Electronics Engineers., *Innovative Computing, Information and Control (ICICIC), 2009 Fourth International Conference on : date, 7-9 Dec. 2009.* IEEE, 2009.

[3] M. S. Thakare and K. K. Chhajed, “Botnet Attack and its Detection Techniques,” *IJARCCE International Journal of Advanced Research in Computer and Communication Engineering*, vol. 10, 2021, doi: 10.17148/IJARCCE.2021.107103.

[4] A. K. Tyagi and G. Aghila, “A Wide Scale Survey on Botnet,” 2011.

[5] C. Joshi, R. K. Ranjan, and V. Bharti, “A Fuzzy Logic based feature engineering approach for Botnet detection using ANN,” *Journal of King Saud University - Computer and Information Sciences*, 2021, doi: 10.1016/j.jksuci.2021.06.018.

[6] C. Joshi, R. Ranjan, and V. Bharti, “ANN based Multi-Class classification of P2P Botnet,” *International Journal of Computing and Digital Systems*, vol. 11, no. 1, pp. 1319–1325, Apr. 2022, doi: 10.12785/ijcds/1101107.

[7] P. Wainwright and H. Kettani, “An analysis of botnet models,” in *ACM International Conference Proceeding Series*, Mar. 2019, pp. 116–121. doi: 10.1145/3314545.3314562.

[8] J. Vania, A. Meniya, and H. B. Jethva, “A Review on Botnet and Detection Technique,” International Journal of Computer Trends and Technology, [Online]. Available: http://www.internationaljournalssrg.org

[9] “(45) ENISA threat landscape report 2018 - Publications Office of the EU”.

[10] M. Sandip Sonawane, “A Survey of Botnet and Botnet Detection Methods.” [Online]. Available: www.ijert.org

[11] S. Shaposhnikov, Sankt-Peterburgskiĭ gosudarstvennyĭ ė lektrotekhnicheskiĭ universitet “LĖ TI,” Natsional′nyĭ issledovatel′skiĭ universitet “MIĖ T” (Russia), Institute of Electrical and Electronics Engineers, Institute of Electrical and Electronics Engineers, and Institute of Electrical and Electronics Engineers., *Proceedings of the 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus) : January 29 - February 01, 2018, St. Petersburg and Moscow, Russia*.

[12] C. A. Rivera A, A. Shaghaghi, and S. S. Kanhere, “Towards a Distributed Defence Mechanism against IoT-based Bots,” in *Proceedings - Conference on Local Computer Networks, LCN*, Nov. 2020, vol. 2020-November, pp. 449–452. doi: 10.1109/LCN48667.2020.9314830.

[13] S. Sennan, P. Srinivasan, S. Sankar, and P. Srinivasan, “Internet of Things (IoT): A survey on empowering technologies, research opportunities and applications,” 2016. [Online]. Available: https://www.researchgate.net/publication/312522213

[14] E. Hopali, Ö. Vayvay, and E. Hopali, “Internet of Things (IoT) and its Challenges for Usability in Developing Countries Quality Management in Healthcare View project Technology and Innovation Management in Telecommunications INdustry View project Internet of Things (IoT) and its Challenges for Usability in Developing Countries,” 2018. [Online]. Available: https://www.researchgate.net/publication/322714582

[15] L. C. Jain, G. A. Tsihrintzis, V. E. Balas, and D. K. Sharma, Eds., *Data Communication and Networks*, vol. 1049. Singapore: Springer Singapore, 2020. doi: 10.1007/978-981-15-0132-6.

[16] “(6)IOT Botnet”.

[17] “(13)Adaptive Online Learning for IOT Botnet Detection”.

[18] S. Amin, H. Seno, M. Nezhadkamali, R. Budiarto, S. Soltani, and R. Budirato, “A survey on real world botnets and detection mechanisms,” *International Journal of Information & Network Security (IJINS)*, vol. 3, no. 2, pp. 116–127, 2014, doi: 10.11591/ijins.v3i2.6231.

[19] H. R. Zeidanloo, F. Hosseinpour, and P. N. Borazjani, “Botnet detection based on common network behaviors by utilizing Artificial Immune System(AIS),” in *ICSTE 2010 - 2010 2nd International Conference on Software Technology and Engineering, Proceedings*, 2010, vol. 1. doi: 10.1109/ICSTE.2010.5608967.

[20] K. Alieyan, M. Anbar, A. ALmomani, R. Abdullah, and M. Alauthman, “Botnets Detecting Attack Based on DNS Features,” Mar. 2019. doi: 10.1109/ACIT.2018.8672582.

[21] M. Feily, A. Shahrestani, and S. Ramadass, “A survey of botnet and botnet detection,” in *Proceedings - 2009 3rd International Conference on Emerging Security Information, Systems and Technologies, SECURWARE 2009*, 2009, pp. 268–273. doi: 10.1109/SECURWARE.2009.48.

[22] R. Vishwakarma and A. K. Jain, “A survey of DDoS attacking techniques and defence mechanisms in the IoT network,” *Telecommunication Systems*, vol. 73, no. 1. Springer, pp. 3–25, Jan. 01, 2020. doi: 10.1007/s11235-019-00599-z.

[23] M. de Donno, N. Dragoni, A. Giaretta, and A. Spognardi, “DDoS-Capable IoT Malwares: Comparative Analysis and Mirai Investigation,” *Security and Communication Networks*, vol. 2018. Hindawi Limited, 2018. doi: 10.1155/2018/7178164.

[24] S. Specht and R. Lee, “Taxonomies of Distributed Denial of Service Networks, Attacks, Tools, and Countermeasures,” 2003. [Online]. Available: www.princeton.edu

[25] E. Alomari, S. Manickam, B. B. Gupta, S. Karuppayah, and R. Alfaris, “Botnet-based Distributed Denial of Service (DDoS) Attacks on Web Servers: Classification and Art,” 2012.

[26] P. Wang, B. Aslam, and C. C. Zou, “Peer-toPeer Botnets.”

[27] H. Owen, J. Zarrin, and S. M. Pour, “A Survey on Botnets, Issues, Threats, Methods, Detection and Prevention,” *Journal of Cybersecurity and Privacy*, vol. 2, no. 1, pp. 74–88, Feb. 2022, doi: 10.3390/jcp2010006.

[28] R. Alhajri, R. Zagrouba, and F. Al-Haidari, “Survey for Anomaly Detection of IoT Botnets Using Machine Learning Auto-Encoders,” 2019. [Online]. Available: http://www.ripublication.com

[29] T. Luo and S. G. Nagarajan, “Distributed Anomaly Detection using Autoencoder Neural Networks in WSN for IoT,” Dec. 2018, doi: 10.1109/ICC.2018.8422402.

[30] H. R. Zeidanloo, M. J. Zadeh, Shooshtari, P. V. Amoli, M. Safari, and M. Zamani, “A taxonomy of Botnet detection techniques,” in *Proceedings - 2010 3rd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2010*, 2010, vol. 2, pp. 158–162. doi: 10.1109/ICCSIT.2010.5563555.

[31] S. Tobiyama, Y. Yamaguchi, H. Shimada, T. Ikuse, and T. Yagi, “Malware Detection with Deep Neural Network Using Process Behavior,” in *Proceedings - International Computer Software and Applications Conference*, Aug. 2016, vol. 2, pp. 577– 582. doi: 10.1109/COMPSAC.2016.151.

[32] A. George, H. Dhanasekaran, J. P. Chittiappa, L. A. Challagundla, S. S. Nikkam, and O. Abuzaghleh, “Internet of Things in Health care using Fog Computing.”

[33] M. Moghimi and A. Y. Varjani, “New rule-based phishing detection method,” *Expert Systems with Applications*, vol. 53, pp. 231–242, Jul. 2016, doi: 10.1016/j.eswa.2016.01.028.

[34] S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, “A Comparison of Machine Learning Techniques for Phishing Detection.”

[35] L. Mathur, M. Raheja, and P. Ahlawat, “Botnet Detection via mining of network traffic flow,” in *Procedia Computer Science*, 2018, vol. 132, pp. 1668–1677. doi: 10.1016/j.procs.2018.05.137.

[36] H. Zhao, H. Shu, and Y. Xing, “A Review on IoT Botnet,” in *ACM International Conference Proceeding Series*, Jan. 2021, vol. PartF168982. doi: 10.1145/3448734.3450911.

[37] P. Beltrán-García, E. Aguirre-Anaya, P. J. Escamilla-Ambrosio, and R. Acosta-Bermejo, “IoT Botnets,” in *Communications in Computer and Information Science*, 2019, vol. 1053 CCIS, pp. 247– 257. doi: 10.1007/978-3-030-33229-7_21.

[38] P. P. Pavelčeleda, R. Krejčí, J. Vykopal, and M. Drašar, “Embedded Malware-An Analysis of the Chuck Norris Botnet.” [Online]. Available: http://87.98.163.86/pwn/syslgd;

[39] “(39)DDoS-Capable IoT Malwares”.

[40] Institute of Electrical and Electronics Engineers, Zewail City of Science and Technology, IEEE Circuits and Systems Society, and Jāmiʻat alQāhirah, *The 31st International Conference on Microelectronics : (ICM 2019) : December 15-18.*

[41] J. Zhou, Z. Xu, A. M. Rush, and M. Yu, “Automating Botnet Detection with Graph Neural Networks,” Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.06344

[42] Amit Singh, Ranjeet Kumar Ranjan & Abhishek Tiwari (2021) Credit Card Fraud Detection under Extreme Imbalanced Data: A Comparative Study of Data-level Algorithms, Journal of Experimental & Theoretical Artificial Intelligence, DOI:10.1080/0952813X.2021.1907795

[43] Kaushik, S. An Introduction to Clustering and Different Methods of Clustering. 2016. Available online: https://www.analyticsvidhya.com/blog/2016/11/anintroduction-to-clustering-and-different-methods-ofclustering/(accessed on 30 December 2021).

[44] Brownlee, J. How to Calculate Precision, Recall, and F-Measure for Imbalanced Classification. 2020. Available online: How to Calculate Precision,Recall, and F-Measure for Imbalanced Classification(machinelearningmastery.com)(accessed on 30 December 2021)

[45] “What is Fuzzy Logic? - Definition from SearchEnterpriseAI,” *SearchEnterpriseAI*, Jun. 01, 2021. [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition /fuzzy-logic. [Accessed: Apr. 18, 2022]

[46] “A Fuzzy Logic based feature engineering approach for Botnet detection using ANN - ScienceDirect,” *A Fuzzy Logic based feature engineering approach for Botnet detection using ANN - ScienceDirect*, Jul. 01, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1319 157821001658. [Accessed: Apr. 18, 2022]

[47] N. Koroniotis, N. Moustafa, E. Sitnikova et al., Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: BotIoT dataset, Future Generation Computer Systems (2019), https://doi.org/10.1016/j.future.2019.05.041

[48] M. Asadi, M.A.J. Jamali, S. Parsa et al., Detecting botnet by using particle swarm optimization algorithm based on voting system, Future Generation Computer Systems(2020),doi: https://doi.org/10.1016/j.future.2020.01.055.

[49] Ali Muhammad, Muhammad Asad, Abdul Rehman Javed, "Robust Early Stage Botnet Detection using Machine Learning", *Cyber Warfare and Security (ICCWS) 2020 International Conference on*, pp. 1-6, 2020.

[50] Vania, Jignesh, Arvind Meniya, and H. B. Jethva. "A review on botnet and detection technique." *International Journal of Computer Trends and Technology* 4.1 (2013): 23-29.

[51] Wainwright, Polly, and Houssain Kettani. "An analysis of botnet models." *Proceedings of the 2019 3rd International Conference on Compute and Data Analysis*. 2019.