ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

A Study on Detection of Botnet by Machine Learning

Swathi Rathi

EAS SAP Analytics, Cognizant Technologies, Bangalore, India

ABSTRACT: At past many of the advancements came into existence regarding the concept of botnet in Machine Learning. As the technology is growing day by day as a result the privacy terms is increasing simultaneously. Even though how secure the server may be the attacker perform different techniques and decrypts the data. The action took by the attacker depends on the security leakage and privacy leakage. Many technologies and in terms of privacy issues were raised regarding this concept in social media. Many mechanisms were proposed regarding the concept of maintaining and controlling the traffic by using the Machine Learning. In this paper we are focusing the regarding the botnet in the domain of traffic monitoring and controlling in Machine Learning and we also added advantages, limitations, performance etc., by using this concept. We have made a special focus and concentration while generating results. Finally we have added the limitations and challenges by using this concept.

KEYWORDS: Machine Learning, Botnet, Traffic, Monitoring, Controlling, Privacy, Prevention.

1. INTRODUCTION

As the usage of internet became cheap the internet came into usage for all the domains such as Banking, Marketing, Industry and Retail etc, even though the services are good there are a lot number of Privacy issues and all few among them are denyl of service attack, Man in the Middle attack, networking attacks etc., as the cyber crimes were increased the number of attacks were raised in the network. [1] Many authors studied and concluded that the most of attacks was happening in Local area network (LAN) and Wide area Network (WAN). Present the concept of Internet of Things (IoT) came into existence many of the devices were handheld and all the process were automated based on user instruction. Regarding this issue the cyber security stands as a challenging task. Generally when several devices were connected using internet security acts as a major role. In corresponding with the advancement and extension of online services, cyber crimes have been executed by cybercriminals to disturb and bargain PC frameworks. This malware imperils the security of client's information. With the development of online users and an expansion in administrations the web renders, consistent development has additionally been seen in the spread of malware. Malware has been believed to have experienced a quick advancement expanding the instrument of spread,[2] vindictive exercises and flexibility to bring down endeavors. The most advantageous and dangerous attacks is the malware attack over the bootnet for the several components. Botnet can be defined as a component of multiple networks which are connected by a single channel called as "Boot master". They attack in the unexpected manner and attack over the device very rapidly and suddenly.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

As per the previous research on the concept of bootnet these are generally categorized into two types namely:

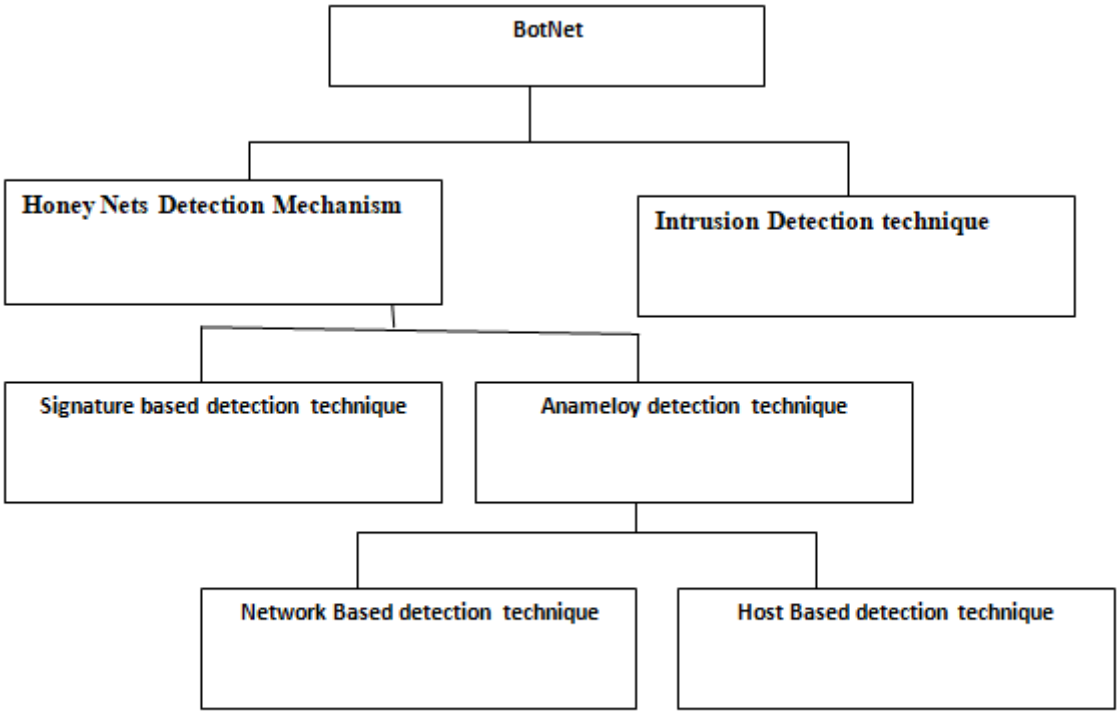

Figure 1: General Architecture of Botnet

Anomaly detection technique is a technique which generally we use in Botnet. One of the most famous detection techniques for malicious is Botnet this mechanism we use in Botnet. The main thing in the concept of Machine Learning is in identifying the correct nodes and identifying the nodes [3] while at the time of conjunction control, flow control and Traffic control over the network. The main advantages of Botnet in machine learning are because of its Effectiveness and it has a greater comp actability in detecting unknown nodes over the network. Regarding the concept of digital signature researchers worked a lot in identifying the nodes which are authorized [4] and which are unauthorized by the help of Botnet Mechanism. the another thing identified is by using the botnet technology it's easy to verify and detect the ineffective nodes in the network.

This paper mainly focuses on the concept of Botnet technique and other mechanisms which are used in the Machine Learning. It concludes the verifying regarding the identifying of intrusion over the networks.[5]

II. LITERATURE REVIEW

Some of the authors namely Yan et al, Fe Alejandra etc., have used the features of the present Botnet available one and picks 21 features in the Botnet out of 16 features were extracted from Machine Learning. The result generated was on average of 75%. [6] The authors proposed a mechanism and mechanism depends on the Peer 2 Peer networks. The complete network depends on the 7 layered OS I reference model and the network traffic mechanism. The authors have proposed the mechanisms in which the detection and selection procedure can takes place in the bootness.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

III. CONCEPT OF BOTNET

This domain was treated as one of the privacy thread from many decades. There ere many decades and attacks took place while sharing data and lot of sensitive information such as bank account details, health information etc., was leaked. So by observing all these challenges many of the researchers were motivated to do this task. This domain was constantly evolved and made the complete concept in a proper structured manner. The Botnet architecture was evolved continuously because the change in the internet. The Botnet attack was performed by different networks such as File Transfer Protocol (FTP), [7] Hyper Text Transfer Protocol(HTTP), peer-peer network, by these different nodes the attacker attacks and connects to the server and thefts the data. The attacker retrieves all the information such as user entered commands, the victim location, IP address of the victim etc., the botnet makes sure regarding the updation of the server from time to time.

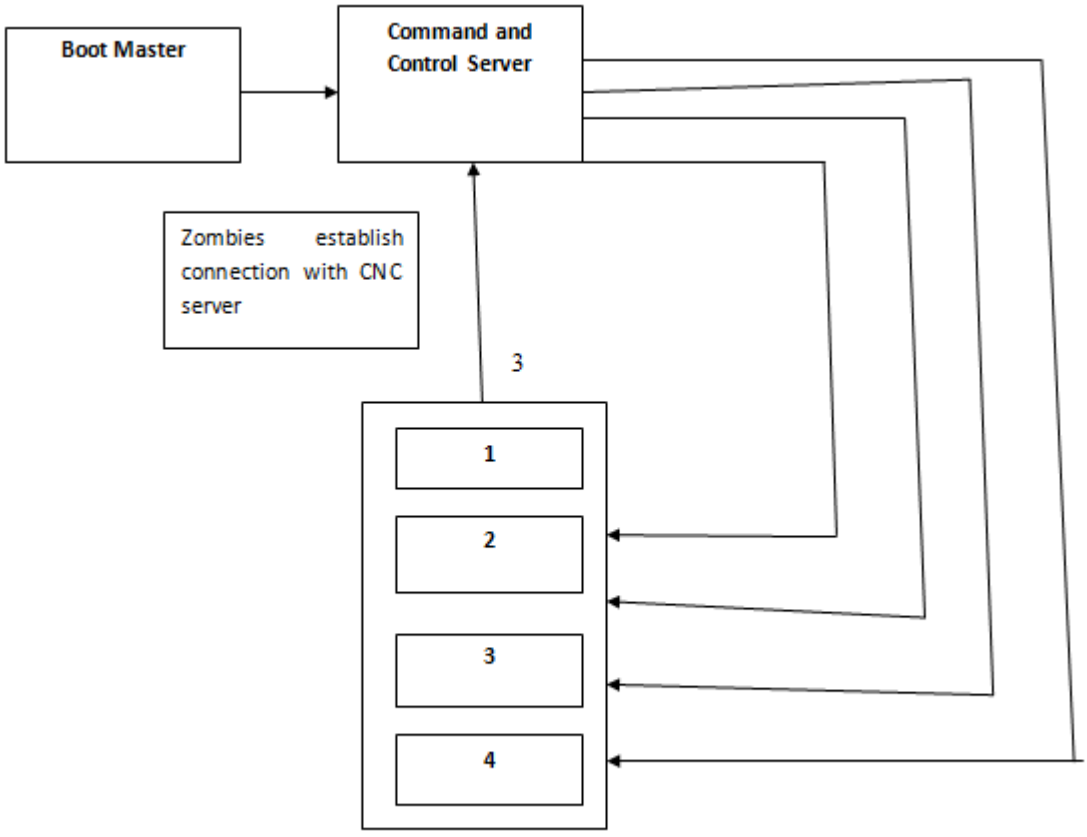

Figure 2: General Life Cycle of Bonet

3.1 Botnet Architecture

The bonet architecture was classified into three categories namely:

- Centralized architecture
- Decentralized Architecture and
- Hybrid Architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

The detailed architecture classification was explained below:

Centralized architecture

Here the request and the messages come from the centralized system or Hub. The Botnet controls the task from the centralized devices from the network communicating channel for communication from centralized server to the client.

Decentralized Architecture

Here only one machine acts as a main role in communicating channel for communicating of messages. There are several centralized devices and mechanisms in detecting and verifying the server. By the help of Zombie and Botnet controlling mechanism and Zombie we can detect the suspicious nodes in the network.

Hybrid Architecture

The hybrid mechanism is a combination of centralized mechanism and decentralized mechanism. Here we have 2 bots namely Client Bot as well as Server Bot. monitoring and detecting finds more typical in finding the best among them.

3.2 Machine Learning Concept

It is a concept derived from the Artificial Intelligence. The main theme of this concept is to establish and identify from the previous experiences. Here many models were defined to model the accurate and defined data. Here the complete concept of Machine Learning depends on the Botnet and it also depends on the past things. [8] Here based on the past things and past expectations many of the things were came into the existence such as identifying the data and differentiating the data from the correct one to the past one. The concept of the machine learning came into the existence by the help of different things trends such as History , Politics, Sociology, Arts, Science etc., there are several machine learning algorithms the theme of all these algorithms is to identify the type of the data. Generally the data is classified into 2 types of classifications namely:[12,13,14]

- Supervised Data and
- Unsupervised Data

Supervised data is a type which generates the values to generate the desired result from the set of data. Here the results will be generated in the accurate manner and in the correct manner. Here all the values in the stored variables were labeled and have a label id. [9,15,16]

Unsupervised data is a type of dataset which have the data with the different dataset and values. The defined things in the unsupervised data are in the improper order. For searching and getting the results in the accurate manner takes lot of time and in some cases it's not possible. [10] Here first we have to identify the type of data because here we cannot have any input such as data type and other reverent values.

In the domain of Machine Learning many of the mechanisms were used for Botnet Detection.

IV. PROPOSED APPROACH

For the development of this mechanism we have took the clustering process. The complete set up was defined as below:[17,18]

Hardware

For the developing of this project we took the configuration of minimum system specifications such as Intel processor, 2 GB ram, dataset with some records, some nodes were deployed in the physical connected nodes.

Software

Here we have designed complete project by the help of Java Programming. Here the particular structured library [11] was designed for the development of the project.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

Dataset

We have taken 13 different datasets which having 13 different sub-category concepts. These datasets having the different items related information. [12,19]

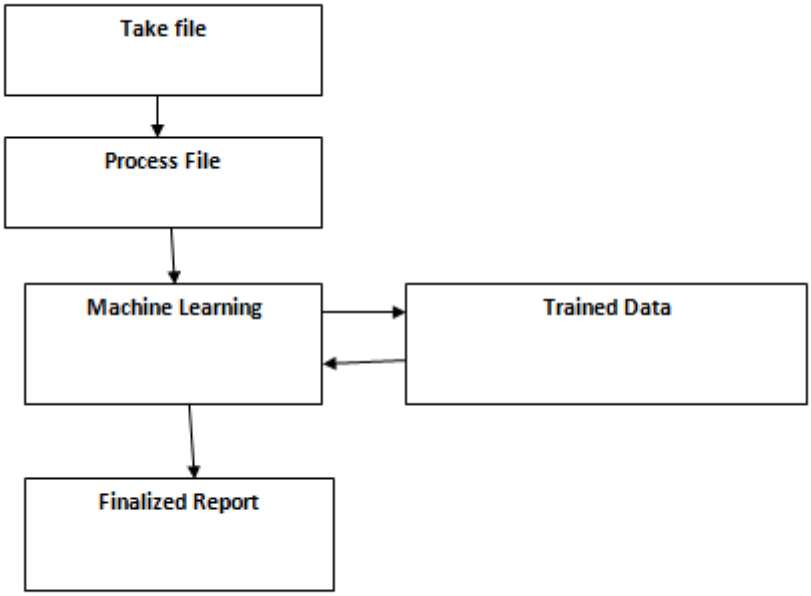

Figure 3: Project Execution Work Flow

V. ALGORITHM

To overcome the proportion we have used the concept of Noise ratio. It gives the robustness of the mechanism. Here in the algorithm we have used the value of X. The term "X" represents the noise ratio proportion and the values needed to add. The noise ratio proportion ranges in between -1 to 1. The formula is represented by

$$X' = X + VAR * NOISE * X$$

Here

X represents Feature need to add ratio.
Var represents the random number ranges from -1 to 1.
Noise represents random values ranges from decibel values.

VI. EXPERIMENTAL RESULTS

To perform this project successfully we have developed 3 Synthetic datasets namely Zeus, Virut and Waledac. The size of all the datasets are 1024 MB. The features and Properties of all the datasets were same. At last the dataset is tested for performing for the comparison of one dataset to other. Here we have used the things namely

- Pre-Processor
- Construction flow and
- Calculating features of machine learning.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

Here for datasets we have few things such as

- Sender port
- Receiver Port
- Protocol
- Total packets transferred
- Number of NULL packets transferred
- Number of small packets transferred
- Proportion among the incoming and outgoing packets
- Flow time
- Re-establishment of connection
- Size of packet 1
- Total packet size
- Average payload
- Total payload send
- Total payload received
- Standard deviation of packet
- Average bits size per second
- Medium bits of packet
- Variable packet size
- Variable packet sent
- Variable packet size received

This procedure intends to discover typical information love for the model we assembled. Initially, lessen the occurrences of typical information to about half to make the attributes of contaminated information progressively self-evident. Besides, increment the level of ordinary information over contaminated information to make it closer to genuine circumstances. At long last, we include distinctive level commotion into testing informational collections in payload and between appearance time and some other related highlights to assess the model.

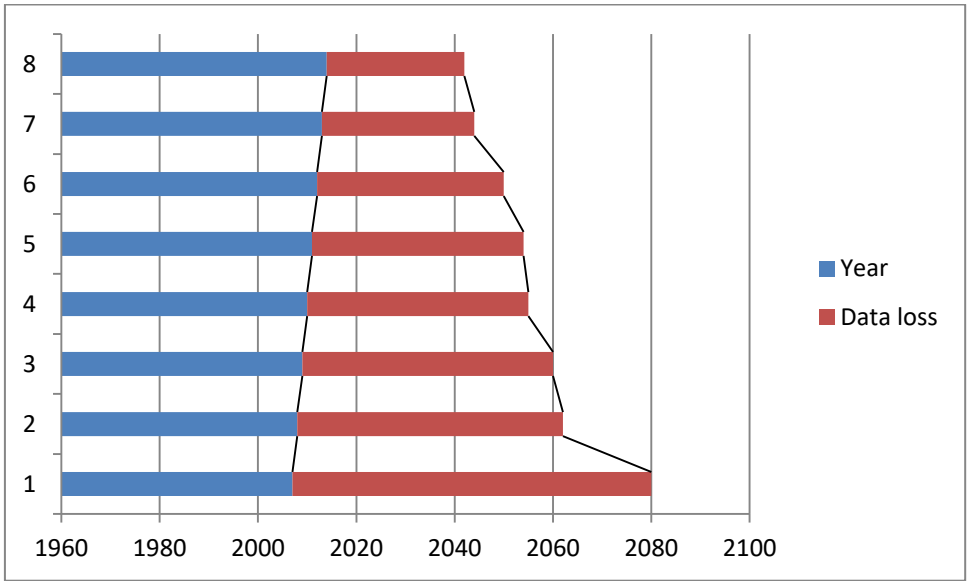

Figure 4: Graph for showing Data loss when Botnet came into Existence

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

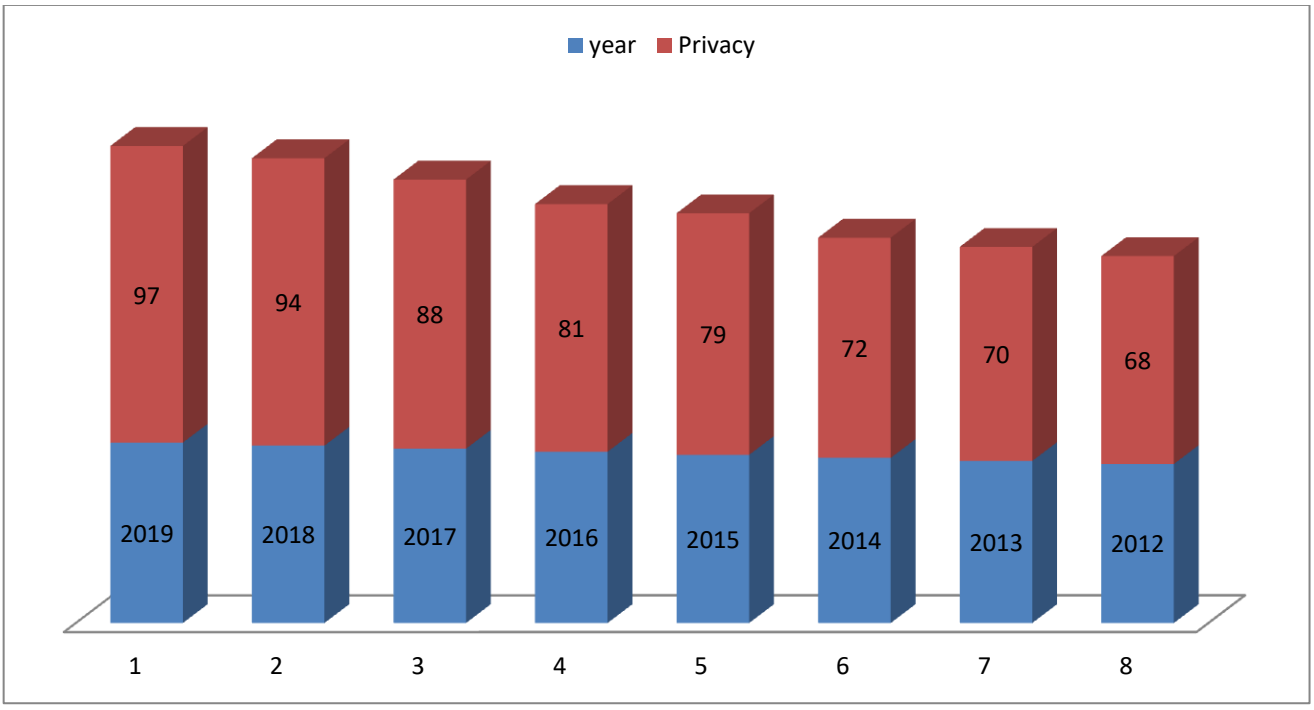

Figure 5: Graph for showing improving privacy in recent decades

Our model can oppose more clamors contrasted with the framework in. Under this unique situation, we add highlights to a higher layer like HTTP highlights and it for sure aides in HTTPS all the more frequently these days.

VII. CONCLUSION

Botnet location dependent on AI has been the subject of enthusiasm of the examination network bringing about the various identification strategies that depend on various botnet heuristics, that target various sorts of botnet utilizing differing AI calculations and that thus give the fluctuating exhibition of discovery. This paper exhibits an audit of the absolute most unmistakable contemporary botnet location strategies that utilization AI as a device of recognizing botnet-related traffic. The displayed investigation tends to scarcely any location strategies, proposed in the course of the most recent decade. The strategies have been broke down by researching bot-related heuristic expected by the recognition frameworks and AI methods utilized to catch botnet-related information. Moreover, the strategies have been inspected by breaking down their qualities, exhibitions, and constraints. The investigation of these identification approaches shows a solid capacity of this class of ways to deal with be utilized for recognizing botnet organize traffic. Notwithstanding, the investigation additionally demonstrates a few parts of AI-based methodologies that could be additionally improved.

Initially, the advanced AI-based identification frameworks should target accomplishing information versatile, on-line and productive recognition. The methodology ought to have the option to adjust to the changing examples of botnet traffic, and it ought to work in the on-line design to give convenient recognition and satisfy the necessities of promptbotnet balance. At long last, the identification strategies ought to be time and computationally proficient so they could be effectively sent at center systems, covering bigger system scopes and giving progressively careful knowledge of traffic created by botnets. Second, the extensive testing and assessment of the proposed identification frameworks are required, where progressively exhaustive informational indexes would be utilized, covering pernicious traffic

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

beginning from the higher number of botnet and non-vindictive traffic that catches the "genuine" idea of the Internet traffic. Furthermore, the assessment ought not just to focus on the appraisal of the precision of discovery yet also on the evaluation of the exhibitions of information pre-preparing and the AI calculations, so qualified judgment on identification exhibitions and the adaptability of strategies could be made.

VIII. FUTURE WORK

Here we are planning to work out for a larger datasets as well as identify the effects which are caused in the work. Here we have defined the deduction, perfectness of the work. We also proposed the latest works regarding the work latest features and capabilities to improve the quality and correctness in the work.

REFERENCES

[1] J. Francois, S. Wang, W. Bronzi, R. State, and T. Engel, "Botcloud: Detecting botnets using mapreduce," in 2011 IEEE International Workshop on Information Forensics and Security. IEEE, 2011, pp. 1–6.
[2] D. Zhuang and J. M. Chang, "Peerhunter: Detecting peer-to-peer botnets through community behavior analysis," in Dependable and Secure Computing, 2017 IEEE Conference on. IEEE, 2017, pp. 493–500.
[3] S. Saad, I. Traore, A. Ghorbani, B. Sayed, D. Zhao, W. Lu, J. Felix, and P. Hakimian, "Detecting P2P botnets through network behavior analysis and machine learning," in Privacy, Security and Trust (PST), 2011 Ninth Annual International Conference on. IEEE, 2011, pp. 174–180.
[4] H. R. Zeidanloo, A. B. Manaf, P. Vahdani, F. Tabatabaei, and M. Zamani, "Botnet detection based on traffic monitoring," in Networking and Information Technology (ICNIT), 2010 IEEE, 2010, pp. 97–101.
[5] W. T. Strayer, D. Lapsely, R. Walsh, and C. Livadas, "Botnet detection based on network behavior," in Botnet detection. Springer, 2008, pp. 1–24.
[6] A. Karasaridis, B. Rexroad, D. A. Hoeflin et al., "Wide-scale botnet detection and characterization." HotBots, vol. 7, pp. 7–7, 2007.
[7] S. Khattak, N. R. Ramay, K. R. Khan, A. A. Syed, and S. A. Khayam, "A taxonomy of botnet behavior, detection, and defense," IEEE Communications Surveys & Tutorials, vol. 16, no. 2, pp. 898–924, 2014.
[8] H. Debar, M. Dacier, and A. Wespi, "Towards a taxonomy of intrusion detection systems," Computer Networks, vol. 31, no. 8, pp. 805–822, 1999.
[9] J. Caballero, C. Grier, C. Kreibich, and V. Paxson, "Measuring pay-perinstall: the commoditization of malware distribution," in Proceedings of USENIX Security Symposium, 2011, pp. 13–13.
[10] J. M. Ehrenfeld, "WannaCry, Cybersecurity and Health Information Technology: A Time to Act," Journal of Medical Systems, vol. 41, no. 7, p. 104, 2017.
[11] R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano, and O. M. Caicedo, "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," Journal of Internet Services and Applications, vol. 9, no. 1, pp. 1–99, 2018.
[12] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," Leading Issues in Information Warfare & Security Research, vol. 1, no. 1, p. 80, 2011.
[13]Kumar. Attangudi P. Perichappan, S. Sasubilli and A. Z. Khurshudyan, "Approximate analytical solution to non-linear Young-Laplace equation with an infinite boundary condition," 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, 2018, pp. 1-5. doi: 10.1109/ICOMET.2018.8346349
[14] S. Chandrasekaran, "Contemplated Method for Predicting Disease by Deep Learning Approach Over Big Data," 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), San Salvador, 2018, pp. 1-5.
doi: 10.1109/RICE.2018.8509090
[15]Gopinadh Sasubilli,Uday Shankar Sekhar, Ms.Surbhi Sharma, Ms.Swati Sharma, "A Contemplating approach for Hive and Map reduce for efficient Big Data Implementation" 2018 Proceedings of the First International Conference on Information Technology and Knowledge Management pp. 131–135 DOI: 10.15439/2018KM20
[16]Sreenivas Sasubilli,Kumar Attangudi Perichiappan Perichappan,P. Srinivas Kumar,Abhishek Kumar,"An Approach towards economical hierarchic Search over Encrypted Cloud", 2018 Proceedings of the First International Conference on Information Technology and Knowledge Management pp. 125–129, DOI: 10.15439/2018KM38
[17]R. T. Mylavarapu and B. K. Mylavarapu, "Huge information extraction techniques of Data Security," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 179-183.
doi: 10.1109/ICICCT.2018.8473017
[18]B. K. Mylavarapu, "Implementing Machine Learning in Intensive Care Units: For Avoiding Over Medication," (2018) International Journal of Pure and Applied Mathematics, Volume 118 No. 20 2018, 4799-4811URL: https://acadpubl.eu/hub/2018-118-21/articles/21f/33.pdf
[19] R. T. Mylavarapu, "A Method for Approximated Deep Learning Towards Dynamic Sharing from Big-Data Analysis," 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), San Salvador, 2018, pp. 1-6.doi: 10.1109/RICE.2018.8509060