First International Conference on Computer Engineering
International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
Volume 5 Issue 8, November-December-2020

Survey on Botnet and Its Detection Techniques

Shubham Gour[1], Yogesh Bhosle[1], Onkar Jagtap[1], Pratik Nirmale[1], Prof. Monika Dangore[1]
[1]Department of Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Maharashtra India

ABSTRACT

Botnet term was coined when multiple networks of bots came into existence. It is a number of Internet-connected devices, which runs single or multiple bots. Botnets can be used to perform Distributed Denial-of-Service (DDoS) attacks, steal data, Ransomware, send spams, and allow attackers to gain unauthorised access on devices and its connections. Command and control(C&C) software are used by the Owner (BotMaster) to control the botnet. This paper explores the survey conducted on botnet and its detection techniques.

Keywords - Botnet, Botmaster,Intrusion Detection System(IDS), Neural Network, P2P, Network Traffic.

I. INTRODUCTION

A bot is an automated program which runs over the internet, some run automatically, while some run when they are triggered by specific input. Internet connected devices are infected with a piece of software that is bot. These internet connected devices are nothing but the botnet. After infection, these internet connected devices steer the instruction commanded by the owner of Botnet known as Bot Master/Bot Herder in 4 phases.

Following are the phases of the botnet infection:

Phase 1 Infection Initialization

A- "Social media" posts targeted by cybercriminal, In the first instance cybercriminal will post a malicious link on social media websites like hoax advertisements, shammed icons etc. When users perform any action on these websites, their action proved to be erroneous, as the current page is redirected to a malicious website, where the software gets installed which was already planted by the BotMaster.

B- "Infection method" approach is followed by the cybercriminals. In this "Email Phishing" tactics are being used to lure users on malicious websites as the user gets redirected when a link is being clicked, and their system gets compromised.

C- "Email Attachments" cybercriminals embody malicious pieces of software with an email, which gets downloaded once clicked and infects the whole system.

Phase 2 Connection to C2C Server

System manifests a connection with a command-and-control (C & C) server which establishes unauthorised connection periodically or may consummate upon infecting the system with malicious activity. Any infected machine liaising with C&C server will comply to launch a coordinated attack. e.g P2P, TELNET, IRC

Phase 3 Control

Cybercriminal (BotMaster) superintends the command and control of botnets for remote process execution by installing botnets on compromised machines. BotMasters uses Tor/shells to hide their tracks by hiding their identities via proxies to disguise their IP addresses.

Phase 4 Multiplication

Attacks in the first 3 phases are incessanted by Botmasters to infect copious internet devices by malicious use of botnet by promulgating fraud, spam emails, DDOS, keylogger, Miria botnet etc. Most recent attack was the " Kashmir Black", an active botnet comprehending thousands of compromised systems across 30 countries and exploiting dozens of vulnerabilities by targeting their CMS. It is believed that the campaign of the "Kashmir Black" started around the end of November 2019 and was trained to aim CMS platforms like Vbulletin, Opencart, Yeager, Joomla!, WordPress. Thus after knowing these vicious internet attacks which happen on a daily basis. We decided to counter this issue by implementing an ML model. In this paper we are going to fill up the canvass of loopholes and vulnerabilities with our ML model. To grasp the enormous nature of Machine Learning models let us first know about the basic model of Botnet. The Figure 1 stages a basic model of botnet in which botmaster is directly or indirectly connected to every other entity such as server, bots, benign hosts through two way communication.

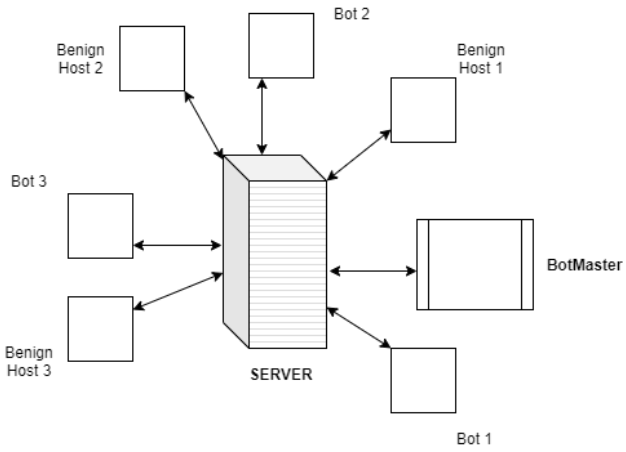

Figure 1.Model of Botnet

II. II.RELATED WORK

In the paper[1] author proposes botnet detection through graph-based feature clustering. In graph based feature clustering a novel botnet detection methodology has topological features like in degree, out degree, weights, clustering coefficient, node betweenness and centrality. A self methodize clustering is applied in networks to establish clusters of nodes based on these features. This technique is competent enough to isolate bots in clusters of small chunks while the majority of them are abided in the same big cluster, making them easily cast around a limited number of nodes. A procedure is developed to verify the algorithm efficiency by using[1] CTU-13 datasets against a detection method which is based on classification. Despite their assorted behaviour, results showed efficiency of the model is promising. In normal condition undesirable results may behave differently. Some models are trained to avoid the unfiltered data but the amount of humungus data challenges the computational expense.

The paper[2] proposes the detection of P2P botnets through network flow level behaviour analysis.In Network flow analysis an enhanced peer hunter

and Network flow based analysis system is used to detect peer to peer botnets. Once P2P network flow is detected then it moves to "correlative contacts" to cluster chunks of bots into communities. It uses community behaviour analysis to mark potential botnets. There are two evasion attacks, one where invaders know the attempts to evade the system by making peer to peer bots to act like legitimate P2P applications which showed high detection with some false positives can be achieved by using enhanced peer hunter technique. Botmaster uses three approaches adding randomized junk packets to depict bytes per characteristics , reducing the number of destination diversity rate from "mutual contacts", randomized spatial communication can be dealt if we adopt time communication features.

In the paper[3] the author proposes a system to detect botnets using a graph based behavioural model that extracts time dependent relations among the similar ones and detect the malicious connections. This process steers clear of the drawbacks due to statistical or signature-based techniques.

A graph is constructed based on the NetFlow records, after which an unsupervised mining method is used to trace root problem. Outlier detection and clustering techniques are used. From the flow records events are extracted from which sequence database is created using which graphs are plotted and outlier detection is used to detect anomalies.

The graphs generated are host-to-host and such graphs are generated for every individual host which entails a high overhead.

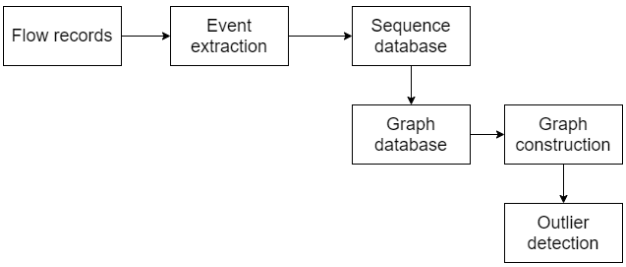

Figure 2.Steps in netflow records for BotGM

Genetic programming is used on streaming data by labels in this paper[4].Sometimes detection of a botnet behaviour is not as helpful, since bots can mix the malicious behaviour with normal behaviours or dynamically switch between different applications or using different versions of the same application. This model is being trained in a non-stationary stream of data. So, the prediction of a botnet cannot be apriori. The data can be labelled by humans to ensure the classification isn't manipulated by the attacker. The GP framework predicts the records that may be malicious, these records are then used to further train the classifier. In this method if the human operator who requests true labels, if he requests records from minor classes, then they are promoted aggressively which might result in reduced performance of major classes.

In the paper[5] Neural Networks are used for botnet detection.Botnets are one of the most increasing and significant threats to a large number of internet users in today's world. Botnet is a serious threat to the internet for committing cyber crimes such as DDoS attacks, stealing sensitive information etc.In this paper, the authors have proposed a Machine Learning technique for detecting P2P botnets. Convolutional Neural Network (CNN) has been used in the paper to train the model to detect the

botnets and Decision Tree for enhancing the decision rate to show the effectiveness of the proposed method for botnet detection. The proposed method transforms collected network traffic into flows, then extracts useful features and then utilizes algorithms to recognize botnets. Peer Rush and CTU datasets are used to validate robustness of the method proposed in the paper.

In the paper[6], indexing multiple for network traffic is used. In conventional firewalls, Intrusion Detection System (IDS) catalogues and manages network activities using the connection concept ‘‘flow', which encompasses five 'tuples' of source and destination IP, ports and protocol type. What is being fetched in a single connection can be obtained from TCP/IP field and packet content by inspecting the flows. In the era of the 21st century network has subsumed more connection and substantial bandwidth, the opprobrium of allowing permissions, granting access to IoT devices makes connection vulnerable. Triggering more malicious network threats, whose communications methods have been compromised. Factors like Miscalculations of duration of time and demand of instant data is responsible to acknowledge the network traffic behaviours to resolve this issue, additional two tuple and single tuple flow types being used to associate multiple 5 tuple communications, whereas discrete connection behaviour can be depicted by generating the metadata. These techniques ensure what network activities have been taking place over a course of time. To exemplify the proficiency of this approach a system rule set detects a Multi-peered Zeus botnet which communicates by establishing multiple connections with multiple hosts, thus imperceptible to standard IDS systems by monitoring 5 tuple flow types in isolation. As techniques are rule based, every parsing transpires in real time and does not require post-processing for further acknowledgment. This paper discusses the application for next generation firewalls and analyses the network traffic behaviour using those multiple tuple indexing.

In the paper[7] Machine learning has various applications and methods to solve real world problems in discrete domains. This is possible due to abundant data spread across over the network, significant furtherance of ML techniques, and advancement in computing capabilities. In the figure we discussed the components which are used to build a robust ML model for a given networking model.

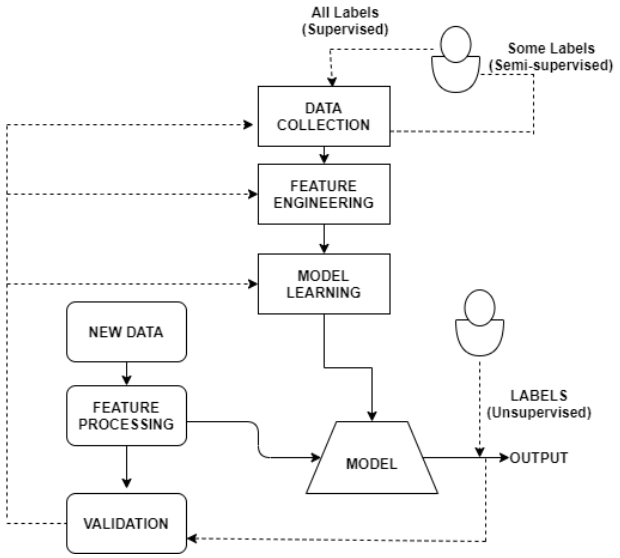

Figure 3. ML based solution.

ML has been applied to dispense its flexible nature to solve real world complex problems in network operations and other sectors. In survey we found out that perplexed problems across different network technologies can be unraveled by using diverse ML techniques which is an injunction with diverse application of Machine Learning.There are fragments like QOE , QOS

management, traffic prediction, congestion control, routing and classification management of networking which we have discussed in our paper to get the insights, scientific challenges and extent of ML in networking.Every effort is accountable and holds the responsibility to push the barriers of automatic network operations and their activities by using the features of ML in networking.

III. LITERATURE REVIEW SUMMARY TABLE:

Paper no.	Paper title	Methodology	Advantages	Drawbacks
1	Botnet detection using graph-based feature clustering	Topological features of nodes within a graph using clustering.	The flow-based detection systems have advantage over the packet-based systems.	If the dataset is large, the computational expense is often high for the detection approach, which is a huge disadvantage if faster detection is required
2	PeerHunter:Detecting peer-to-peer botnets through community behavior analysis	Flow based detection	Robust against changing behavior of botnets	The experiment results showed that the system is robust to PMMKL and will make the botnets less stealthy, less efficient and more exposed while conducting AMMKL But still there are chances that botnets are not detected.
3	BotGM: Unsupervised graph mining to detect botnets in traffic flows	Data mining	1.Uses netflow 2.Doesn't analyze individual behavior.	The graph generated for individual hosts entails a high overhead
4	On botnet detection with genetic programming under streaming data label budgets and class	Genetic programming	Works on a non-stationary stream of data.	If the minor classes are requested the efficiency decreases

	imbalance			
5	Botnet Detection Using Recurrent Variational Autoencoder	Convolutional Neural Network	Botnet detection using CNN has high accuracy rate and low false positive rate in detecting botnets as compared to other detection methods like anomaly based and signature based detection method	To conduct the classification of network traffic with the proposed technique is difficult as this technique is appropriate only for image and pattern recognition.
6	Peer Based Tracking using Multi-Tuple Indexing for Network Traffic	Multi-tuple indexing for network traffic analysis	Detection of multi-peered ZeuS botnet	Cannot locate features, generated by metadata to log traffic characteristics across multiple connections.
7	Survey on machine learning for networking: evolution, applications and research opportunities	Using ML in traffic prediction using MLP-NN	i.Traffic sampling and interpolation ii. Leveraging features other than traffic volume for traffic prediction.	It works on mostly on single-tenant and single-layer networks, re-architected is required to take into account the notion of multi tenancy in multi-layer networks.

IV. CONCLUSION

This paper construes various techniques and methods to deal with botnets under different situations over different networks. The main threat in bot detection is to avoid any loopholes or vulnerabilities in our own system while tracking them to terminate bot's network before their vicious goal is achieved by their botmaster.We realised that traditional techniques work fine but to make them further good and efficient we have to introduce Machine Learning . ML is the luminosity for the cataclysmic bots which exploit the vulnerabilities in networks and compromises the systems.

V. REFERENCES

[1]. Sudipta Chowdhury1*, Mojtaba Khanzadeh1, Ravi Akula1, Fangyan Zhang2, Song Zhang2, Hugh Medal1, Mohammad Marufuzzaman1, Linkan Bian1" Botnet detection using graph-based feature clustering".

[2]. Zhuang and J. M. Chang, "PeerHunter: Detecting peer-to-peer botnets through community behavior analysis,".

[3]. S. Lagraa, J. François, A. Lahmadi, M. Miner, C. Hammerschmidt and R. State, "BotGM: Unsupervised graph mining to detect botnets in traffic flows," 2017 1st Cyber Security in

Networking Conference (CSNet), Rio de Janeiro, 2017, pp. 1-8, doi: 10.1109/CSNET.2017.8241990.

[4]. Sara Khanchi, Ali Vahdat, Malcolm I. Heywood, A. Nur Zincir-Heywood,"On botnet detection with genetic programming under streaming data label budgets and class imbalance",Swarm and Evolutionary Computation,Volume 39,2018,,ISSN 2210-6502

[5]. Jeeyung Kim, Alex Sim, Jinoh Kim, Kesheng Wu," Botnet Detection Using Recurrent Variational Autoencoder".

[6]. Hagan, M., Kang, B., McLaughlin, K., & Sezer, S, "Peer Based Tracking using Multi-Tuple Indexing for Network Traffic".

[7]. Raouf Boutaba 1, Mohammad A. Salahuddin 1, Noura Limam 1, Sara Ayoubi 1, Nashid Shahriar 1, Felipe Estrada-Solano1,2 and Oscar M. Caicedo 2 "Survey on machine learning for networking: evolution, applications and research opportunities".

[8]. 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," Inf. Warfare Security Res., vol. 1, no. 1, p. 80,2011.

[9]. S. Chen, Y. Chen and W. Tzeng, "Effective Botnet Detection Through Neural Networks on Convolutional Features," 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, 2018, pp. 372-378, doi: 10.1109/TrustCom/BigDataSE.2018.00062.

[10]. B. Alothman and P. Rattadilok, "Towards using transfer learning for Botnet Detection," 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), Cambridge, 2017, pp. 281-282, doi: 10.23919/ICITST.2017.8356400.

[11]. G. Vormayr, T. Zseby and J. Fabini, "Botnet Communication Patterns," in IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2768-2796, Fourthquarter 2017, doi: 10.1109/COMST.2017.2749442.

[12]. H. Dhayal and J. Kumar, "Botnet and P2P Botnet Detection Strategies: A Review," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, 2018, pp. 1077-1082, doi: 10.1109/ICCSP.2018.8524529.

[13]. C. Czosseck, G. Klein and F. Leder, "On the arms race around botnets - Setting up and taking down botnets," 2011 3rd International Conference on Cyber Conflict, Tallinn, 2011, pp. 1-14.

[14]. K. Alieyan, M. Anbar, A. Almomani, R. Abdullah and M. Alauthman, "Botnets Detecting Attack Based on DNS Features," 2018 International Arab Conference on Information Technology (ACIT), Werdanye, Lebanon, 2018, pp. 1-4, doi: 10.1109/ACIT.2018.8672582.

[15]. W. Zhang, Y. -J. Wang and X. -L. Wang, "A Survey of Defense against P2P Botnets," 2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing, Dalian, 2014, pp. 97-102, doi: 10.1109/DASC.2014.26.

[16]. W. Sun and H. Gou, "The Botnet Defense and Control," 2011 International Conference of Information Technology, Computer Engineering and Management Sciences, Nanjing, Jiangsu, 2011, pp. 339-342, doi: 10.1109/ICM.2011.218.

[17]. M. Khosroshahy, M. K. Mehmet Ali and D. Qiu, "Scomf and SComI botnet models: The cases of initial unhindered botnet expansion," 2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Montreal, QC, 2012, pp. 1-5, doi: 10.1109/CCECE.2012.6334871.