AN EFFICIENT DETECTION OF DGA BASED BOTNETS

M. Arun Kumar, G. Aravindh*

Abstract

The recent malicious attacks on the network i.e. Internet are of greater extent and their main intention mostly to gain financial benefits of any organizations, governments, countries defense plans etc., The Botnet is named as one of the most fearsome malware and by the help of DGA's (Domain Generation Algorithms) they were increased rapidly and that too in recent days their strength has been increased abundantly, which provide more stealth to the bots and makes more difficult to detect. They use C&C (Command & Control) server to stay undetected. This paper presents the comparative study on various DGA bots and discusses the challenges in detecting the bots; different types of algorithms used in the process and proposed new botnet detection technique. This paper concludes with the suggesting better way to find DGA botnets and to classify them efficiently.

Keywords : DGA (Domain Generation Algorithm), C&C (command &control server), IRC (Internet Relay Chat).

I. INTRODUCTION

The term "botnet" comprises of two components "bot/robot" and "network". The bots are otherwise called as zombies. A Botnet is made up of a network of devices which are connected to the Internet.

Simply, a bot is a software application which is capable to perform automated tasks over the global network.Typically, bots are designed to perform undertakings that are both basic and fundamentally tedious and they perform at a significantly higher rate to spread the attack widely at great speed.

Botnets can be utilized to perform malicious exercises like DoS/DDoS attack, steal data without user's permissions, send spam, and permit the assailant to access the device remotely. The Botmaster is the one who is responsible for the creation of bots by which he controls the botnets using C&C servers (command and control server).

The botnet which uses DGA algorithm is called as DGA bots, the DGA bots are capable to stay undetected on the Internet because, the bot developer design the bots in such a way to give stealth cover to the bots so that they would remain undetected on the network and make the detection of bots difficult. The famous bots which are being used worldwide.Confickerc, Kraken, Zeus, cybot, sanity, mobile, zeroaccess, marrowfat. People use C&C servers (Command & Control server) to control all the bots which are infected by the malicious software and they give commands to their bots which are created by the botmasters over the network. Sometimes this can also be detected by the various techniques, but by using the p2p technique the bots become more stealth over the Internet and give more hard time for the botnet detection techniques and other security measures.

There are two ways by which people could reach the destined website, either by using the IP address or DNS (Domain Name System). But the easiest way to remember the address is by using DNS because it uses human language e.g. English. The DNS (Domain Name System) is a system used to convert a hostname into an IP address to reach the

Department of Electronics and Communication Engineering,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
* Corresponding Author

destined website. The domains in the domain name are made up of a tree structure, where it is divided into sub domains. The top-level domains are the root domains where the sub domains are rooted down. The domain name is made up of labels or names, which are combined with the parent node and separated by a (.)dot. Either the DNS zone can be a single domain or it can be made up of many domains and sub domains, based on the organization's zone management which is based on their zone requirement. DNS can also be partitioned based on its classes; each separate class can be individually minded as an array of namespace trees. The network administrator is responsible for creating any additional zones. Permissions to the new zone are designated to the name server. The Parent zone will remain as the definitive zone for the recently created new zones.

Contrasted two methodologies to recognize the C&Cs botnet [2]. In the approach, firstidentifiedareas were transiently associated or correlated with Mahalanobis distance and Chebyshev's inequality to recognize the peculiar domains. In the second approach, they investigated the repeating DNS "dynamic" replies for theresponses of NXDomain. These examinations demonstrated approach was incapable, and a few genuine administrations utilize DNS. Yet, their second approach yielded identification and identified C&C spaces weresuspicious.Domain names were anomalous transiently focused on rates query or unusually repeating (Dynamic DNS) answers, individually. The merits were the Bayesian bot detection technique successfully recognized the C&C servers with various domain names, and in the meantime, it created few or no False Positives. The affectability of the examination suggested that this technique was powerful.

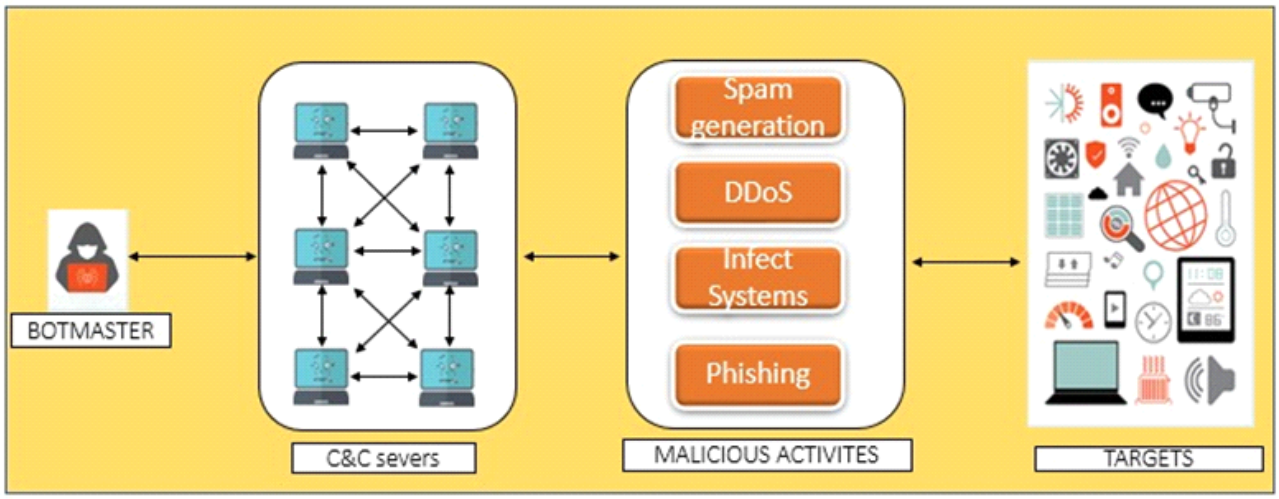

Figure 1. General architecture of Botnet

II. LITERATURE SURVEY

Detectionbotnet presents strategycombines both botnet activity traffic and communication traffic [1]. Clustering is connected to play out the cross-plane relationship to recognize botnets. Merits are that theBotMiner indicates great recognition precision on different sorts of ordinary movement. Demerits over here are some Botstry to evade the detection by using evasion technique especially P2P botnets would use this method.

The outcome demonstrated the absence of false positives and false negatives for a genuinely wide range of parameters. So that, if the parameters were not all around tuned, the strategy may create false positives if the domain name was questioned just by an affected host or by one or more number of uninfected hosts. This outcome in mistakenly ordering an uninfected host as tainted in light of the fact that it has questioned the domain name which was not likewise questioned by an adequate number of other uninfected hosts which is the demerit of this technique.

Worthy of the real-time system, the benefits are accomplished by utilizing Naive Bayesian classifier it gives an exact method for identifying both quick transition spaces and DGA area names [3]. The outcome demonstrates the benefits of this paper is the Naive Bayesian classifier gives the level of best exactness negligible positives false, trailed by classifiers Bayesian lastly and Probability classifier giving comparative outcomes. Proposed the arrangement gives us exact for means enhancing. Also, it gives a successful extra network of layer defense, barrier frameworks. Furthermore, the demerit is that the exploratory outcomes demonstrate that the framework can distinguish which is the demerit of this paper.

On that point, it creates a binary matrix for every domain group, who represents the host and columns which speaksabout the period [4]. The merits develop metric model to detect the botnets in the networks in a large-scale environment. DNS Botnets use to rally the malicious, assaults it uses them to revise the codes and later updates them. So, they gave a case study to train the Botgad to respond the situation where the demerits start when the evasion is done by restricting attack target. Evasion by minimizing the synchronicity, Evasion by inducing IP-churn, Evasion by threshold attacks, Evasion by the botnet subgrouping are the methods by which the Botgad can be evaded.

Exhibited a technique on the progress of training. Added to it, the detection act is dependent on the accessibility of the training data with a sufficientmeasure [5]. The legitimacy is that the Pleiades proposed the framework, which precisely identifies machines that are inside the observation zone which is been affected with DGA-based bots. Pleiades screens the movement beneath the local recursive DNS server and evaluates the surges of unsuccessful DNS resolutions, rather than depending on the manual reverse engineering of bot malware and their DGA algorithms. The Pleiades can accomplish a high detection precision and the disadvantage of this technique is that once another DGA, Pleiades can assemble a statistical model of how those DGA domains "resemble", yet it isn't possible for it to remake the exact domain generation algorithm. So, it creates a specific number of false positives and false negatives.

The key commitments are the relative execution portrayal of each and every metrics in various situations[6]. Moreover, when the technique is connected to the Level 1 ISP's it can trace and detect the renowned bots, DGA bots like Conficker and even other obscure and unclassified bots, which we called a Mjuyh which are their merits and the demerit is the attacker may gather great database bring down the entropy which makes it harder to identify such anomalies.

Botnet detection method in a view for gathering the in a hosts specific timeframe is factually probably won't be questioned again in the accompanying timeframe[7]. In the same way, the proposed plot clusters the queried domains. Detection module identifies areas whose names are produced automatically, and are subsequently mixed up or unimportant.

Acquainted this technique to recognize DGA domains from non-DGA domains by utilizing both linguistic and IP features[8]. Phoenix comprises of three phases, a discovery phase, a detection phase, and an intelligence phase. The detection phase identifies subsequently mixed up and it ensures a false high rate alarm and if there is no adequate history of suspicious domain activities then it won't be considered.

Distinguish DGA-based bots without having any earlier information about the DGA [9]. It utilizes Cumulative Sum (CUSUM) as the CPD calculation since it has been turned out to be powerful and it has been utilized as a part of numerous different works. A host acts ordinarily, yet when we break down the DNS movement on the traffic. We can readily

identify an individual bot by which it is gathered from a single network by utilizing a chain of proof, including the quantity proof, temporal proof, and linguistic proof. The last outcome demonstrated that the BotDigger identifies over 99.8% of the bots with under 0.5% false positives. Be that as it may, the bot can sidestep them by querying C&C gradually, such as querying a domain like every five minutes. On the attempt to detect the bots by query domains we may gradually expand the time window and at that point more false positives will be presented.

DGA-based botnets in reasonable system situations [10]. It can distinguish new DGA botnet like Mjuyh botnet behavior which has to detect the power,multiple sub domains usethe botnets while creating domain series. It also has the capacity to provide an additional safeguard consequently equipped for dealing with vast graph diagrams in a sensible time. Having played out the grouping procedures, the grouping results become the result for the group identification module to be handled in future. The NXDomains creates the consequence of grammatical mistakes or transitory site terminations with most clients [11]. They were grouped as per the correlation of their query behavior while we are working on the recognition time window. Accordingly, DBod can't be recognized and can stay torpid at considerable length of time or even a day if it gets connected to the server.

III. COMPARATIVE STUDY BETWEEN DIFFERENT TECHNIQUES

S.No	Paper's Title & Author	Algorithms used	Merits	Demerits
1	BotMiner (Gu,2008)	Clustering algorithm and cross-plane correlation.	Excellent detection, accuracy, very low false positive rate.	Some botnets try to evade the detection by using evasion technique especially the P2P botnets.
2	Bot gad (Choi,2009)	X-means clustering algorithm, Binary matrix algorithm.	Metric model to detect the botnets in the largescale networks in real -time. It updates their codes, creates case-study to train themselves.	Evasion by restricting attack target, Evasion by minimizing the synchronicity, Evasion by inducing IP -churn, Evasion by threshold attacks, Evasion by the botnet subgrouping.
3	DF Botkiller (sharifnya,2015)	The Spearman's rank correlation Coefficient Jensen-Shannon divergenceand The distance Levenshtein algorithm.	It is as reputation system in online negative.Identifies suspicious domain failures.	Has high false alarm rate. Prejudgment of hosts.

4	Botdigger (Zhang,2016)	CumulativeSum (CUSUM) as the CPD algorithm.	They detect DGA -bots without having any prior knowledge. It can detect every single bot by analyzing at their DNS traffic. Detects > 99.8% of botswith false positives < 0.5%.	The botnets can outflank the detection mechanism. Increases the time window, increases more false positives.
5	DBod (Wang,2017)	Chinese Whispers (CW) algorithm.	It can detect DGA -based botnets in Realtime environments. Has the ability to identify the bots using multiple subdomains. It can defend against existing DGA -botnetsand also against emerging botnet patterns.	If a typo or temporary website closures occurred then most users try to reconnect the website which clustered in similar to the query behavior which results in botnet evasion where the compromised hosts stay stealth for hours, even days.
6	Phoenix (Schiavoni,2014)	DBSCAN clustering algorithm and linguistic features.	It produces the knowledge - based behavior of each tracked botnet.	It needs registered domains to function. So, the data is being given with longer collection periods. It uses pronounceable domains to evade unseen DGAs and future DGAs.
7	Bayesian bot detection (Villamarin,2009)	Bayesian method.	It successfully recognizes the C&C servers with multiple domain names, and with minimum or no False Positives. This approach is robust.	The result shows the absentees of false positives and false negatives for a given wide range of parameters.
8	Domain flux (Yadav,2012)	Jaccard index, KullbackLeibler divergence and edit distance Levenshtein	It's been enforced Tier-1 ISP's which could trace and detect the well -known bots, DGA bots and even other unknown and unclassified bots.	An experiment exhibitto about the entropylower, it makes harder to detect the anomalies.

Table 1: Table for comparison between different botnet papers.

IV. CONCLUSION

It is important to give fitting speak for the real risk on network security and to its significant supporters of undesirable network traffic. In this paper, has done a detailed survey on the various botnet and DGA-based botnet detection technique and compared them according to the algorithms used, merits and demerits. From the above study, it is clear that the previously proposed detection techniques were unable to distinguish DGA-based
botnets in the real-world environment. In future, DGA based botnet detection technique has to be improved, by proposing an architecture with highly powerful algorithms by which the rate of detecting the DGA based bots will be increased to a greater extent. The proposed architecture will not only improve the efficiency on detecting the bots but also helps in increasing the accuracy, detection rate and rate of false positives.

REFERENCES

[1]. Gu G, Perdisci R, Zhang J, Lee W BotMiner "Clustering analysis of network traffic for protocol-and structure-independent botnet detection", in USENIX Security Conference, 2008, pp. 139–154.

[2]. Choi H, Lee H, Kim H "BotGAD detecting botnets by capturing group activities in network traffic". In Proceedings of the fourth international ICST conference on Communication system software and middleware, ACM, 2009, pp. 2-6

[3]. Sharifnya R, Abadi M. DFBotKiller "Domain flux botnet detectionbased on the history of group activities and failures in DNStraffic",Digit Invest 2015, pp.15–26.

[4]. Wang T.S, Lin H.T, Cheng W.T, Chen C, "DBod Clustering and detecting DGA based botnets using DNS traffic analysis", Comput. Secur, vol. 64, pp. 1-15, Jan. 2017.

[5]. Schiavoni S, Maggi F, Cavallaro L, Zanero S."Phoenix: Dga based botnet tracking and intelligence. In Detection of Intrusions and Malware, and Vulnerability Assessment", Springer (2014), pp. 192–211.

[6]. Villamarin Salmon R and Brustoloni J.C"Bayesian bot detection based on DNS traffic similarity". In SAC'09 ACM conference on Applied Computing, 2009.

[7]. Yadav.S, Reddy.A, Reddy.A, and Ranja.S"Detecting algorithmically generated malicious domain names". ACM, November 2010.

[8]. Bilge L, Kirda E, Kruegel C, Balduzzi M "Exposure finding malicious domains using passive DNS analysis". NDSS, 2011.

[9]. Antonakakis.M, Perdisci.R, Nadji.Y, Vasiloglou. N, Abu-Nimeh.S, Lee.W, and Dagon.D"From throw away traffic to bots Detecting the rise of dga-based malware". In Proc. of the 21st USENIX Security Conference (Security'12), Bellevue, Washington, USA, USENIX Association, August 2012, pp. 48–61.

[10].Stalmans E, Irwin B. "A framework for DNS based detection and mitigation of malware infections on a network"Protocols and encryption of the storm botnet, Black Hat, USA, 2008.

[11].Zhou Y.L, Li Q-S, Miao Q, Yim K. "DGA based botnet detectionusing DNS traffic". J Internet Serv Inf Secur 2013, p.p.116–23.[1].