International Journal of Research in Advent Technology, Vol.4, No.7, July 2016
E-ISSN: 2321-9637
Available online at www.ijrat.org

Botnet Malicious Activity Detection Based on DNS Traffic Analysis

Pooja Devi[1, 3], Sanjeev Kumar[2, 4], Neeraj Sharma[1, 5]
[1] *Computer Science and Engineering Department, Chandigarh University, Mohali, Punjab*
[2] *Cyber Security Technology Division, CDAC, Mohali, Punjab*
{[3] sharmap625@gmail.com, [4] ror.sanjeev@gmail.com, [5]Neerajkirti@gmail.com}

Abstract- In the field of internet security botnet is becoming the significant threat as more number of users are connected to internet. Botnet which is a collection of infected computers so called (bots) are becoming the major threat to internet community. The difference between a malware and botnet is that bot is remotely controlled by a C&C server which are under the control of a botmaster. Here in this research, a DNS traffic based approach is presented for detection of the malicious activities performed by botnets with inclusion of IP-Domain features. At first, a detailed literature is presented for botnet detection, and then a DNS traffic analysis technique is proposed with inclusion of a) IP to Domain pairing, b) deep packet inspection (DPI), c) anomalous behavior of traffic exchanged between bot infected PC to C&C server. A shell script is developed to automatically fetch the network traces from a victim Honeypot machine for further analysis for botnet infections in network traces. Further the global data feeds related to botnet attributes are integrated with the system but the problem with reputation engine is that they only determine the suspicious domain. To determine the actual botnet infections; there is a need to apply another technique in the form of DNS traffic analysis. A prototype system is developed which profile the DNS traffic for botnet determinations, in the end experimental results are presented to validate our research.

Index Terms- Bot, Botnets, Malware, Computer security, Reputation

1. INTRODUCTION

Nowadays web attacks are rapidly increasing, so to stop these attacks is becoming the important and necessary task for internet security researchers. Today botnet is becoming the biggest and serious threat for internet users and slowly botnets are spreading in all over the world. DNS is the core component of Internet but the strange thing is that it is also becoming the part of botnet. Most of the internet attacks are nowadays based on DNS, so it is important to inspect the DNS traffic for the detection of Botnets. Figure 1 depicts the structure of botnet.

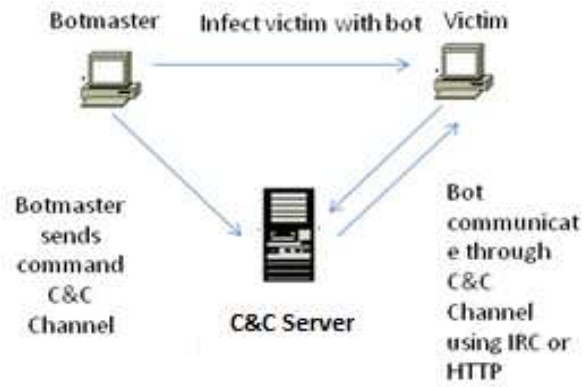

Figure 1: Botnet structure

A botnet is a group of infected computers which can be presented anywhere, anytime (ubiquitous) from home to offices. This group of computers communicates with each other through C&C channels. They work under the control of single user known as botmaster. In botnets botmaster and bots exchange information through C&C channels, which can be implemented using different protocols, such as IRC, HTTP, and DNS. DNS is clearly a critical component of internet which is used for translation of domain names to IP addresses, but there are times when this protocol is used for malfeasance. The most common DNS based attacks are zero day attack, Cache Poisoning, DOS, DDOS, DNS amplification and Fast Flux DNS. For example domain names increasingly playing a role for the management of botnet C&C server download websites where malicious software is hosted and phishing pages that aims to steal the sensitive information from victim machines. Using domains names gives attacker the flexibility for migrating their malicious servers with ease. That is the malicious "services" that the attacker offers become more "fault tolerant" with respect to IP addresses where they are hosted. In this paper we introduce a passive DNS traffic analysis module and detection system to effectively and efficiently detect domain names/IPs that is involved in malicious activity. Many

researchers have use DNS before as a way to analyze, measure and estimate the size of existing botnets in the past years. Some solutions have then attempted to use DNS traffic to detect malicious domains of a certain type. However all these approaches have only focused on specific types of malware. Our approach in comparison is much more generic and not only limited to certain types of attacks (e.g. only botnets). Our system is basically analyze the passive DNS traffic and detect the malicious domains and compares these malicious domains with the blacklisted domains/IPs. If any malicious domain/IP is found then our system generates an alert. Further we check the payload of that particular DNS packet which belongs to malicious traffic. The contributions of this paper are as follows:

- A novel approach for the detection of malicious activity domains/IPs is presented that is based upon DNS traffic Analysis. It does not rely on prior knowledge about the kind of service the malicious domains provide (e.g. botnet that use domain generation algorithm, fast flux services). This is significantly different from existing techniques that only target fast flux or DGA based domains used in botnet operations.
- The implementation of analysis and detection method that detect the malicious domains is presented which incorporate the reputation feeds from online scanner.
- The evaluation of DNS Traffic Analysis and Detection engine with two datasets from the pcap captured on honeynet, as well as Blacklisted dataset is presented.

1.1TECHNICAL BACKGROUND BOTNET

A botnet is a network of compromised machines called bots or zombies under the remote control of a human operator called botmaster. The bot is piece of software that infects and compromises the machines. The infection is carries out through a variety of so called- distribution channels, which vary from compromised websites that serve malware via driven-by mechanisms to phishing. Once infected, the machine will continue to work as nothing changed to the eyes of the legitimate user, while it is capable of executing malicious activities on the behalf of the botmaster, who will employ a command and control server (C&C) to dispatch orders to and gather information from the bots.

Botnet Topologies

Different botnet topologies imply different benefits and weaknesses. Our study will focus on the DNS based botnets, as it is more spread, even though recent reports indicate a rise in DNS based botnets.

a. Centralized

This topology reflects the classic and well established client server pattern. The bots communicate directly with the botmaster, which forwards the messages between clients. This technique guarantees the low latency and control over the packet delivery. Usually these servers are hosted in some bullet proof hosting service providers or in hostile websites, thus they are difficult to take down.

b. P2P

In contrast to the centralized topology, in P2P topology, there is no centralized C&C server, the botmaster can contact and send instruction to any bot. the communication in this topology does not depend on the one or a few selected C&C servers, thus the botnet becomes more resilient to detection and mitigation. As a result defenders cannot defeat this kind of botnet by taking down a small number of bots. Nevertheless, implementing a P2P C&C model requires more work by the bot's authors because of the constraints posed by the P2P topology The main advantage in employing a P2P technology consists of a much more robust and resilient infrastructure, as we do not have anymore a single point of failure but each bot is responsible for broadcasting the message received to the other bots.

c.Unstructured

This is another way to design a botnet, featuring zombies that are completely agnostic with respect to the same botnet they belong to. The bot in an unstructured C&C topology do not contact and report to the C&C server. Whenever they need to send a message to the infected network, they encrypt it, randomly scan the internet and pass along the message when they detect another bot. even though the design is quite simple, it would not be able to guarantee the actual deliver and it would also be prone to extremely high latencies.

BOTNET INFECTION LIFECYCLE

Predominantly, botnet infection lifecycle is categorized into five major phases: initial infection, secondary infection, connection to C&C, malicious activities, and maintenance phases. Figure 2 represents overall diagram of botnet infection lifecycle. In the initial infection phase attacker scan the target machine for known vulnerabilities and infect victim machine through different exploitation methods. This phase is crucial for the progress of the whole infection; if failed further infection is impossible. After successful initial infection, the compromised host contacts the malware server to download the real bot executable. This phase is named as secondary injection since the binary having all bot logic gets executed on this stage. Also

the infected host executes a script known as shell code. Binary download may be realized by the means of FTP, HTTP or P2P protocols.
Further infected host contacts with C&C server in order to registered as online bot and retrieve the commands to be executed. Bot behavior from this point depends on the retrieve commands.

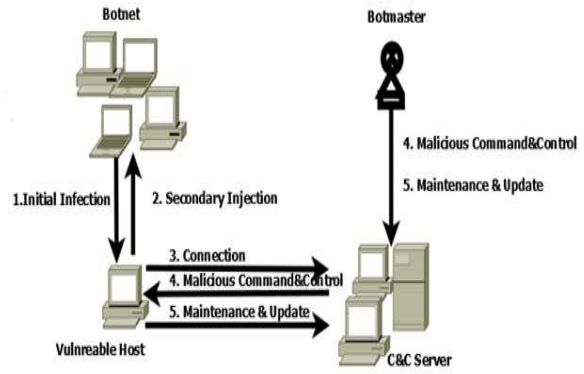

Figure 2: Botnet Infection Lifecycle

BOTNET DETECTION APPROACHES

A simple taxonomy for classifying Botnet Detection Approaches as shown in figure 3 is to categorize the technique as static or behavioral. However, the following sections classify botnet detection approaches into signature-based, anomaly-based, DNS-based and mining-based detection.

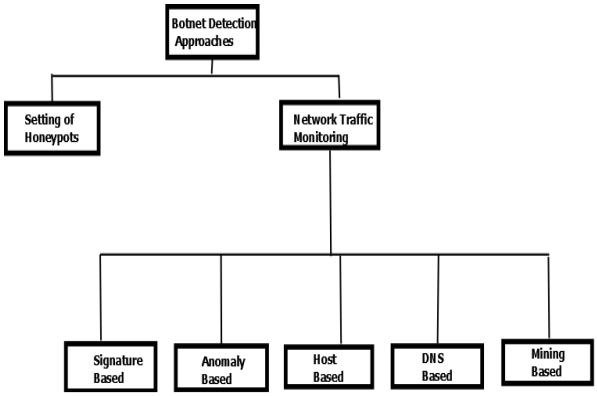

Figure 3: Botnet Detection Approaches

II.RELATED WORK

Recent research on DNS traffic based botnet detection has proposed number of approaches that classify the different features between malicious and benign DNS usage. In [7], they tried to prevent the end users from visiting malicious websites which are target of domain flux technique. In particular they explored the use of statistical methods from machine learning for classifying site reputation based on the relationship between URLs.
In [5] authors introduced a passive DNS analysis approach and a detection system, EXPOSURE, to effectively and efficiently detect domain names that are involved in malicious activity. They used 15 features that allow them to characterize different properties of domain names and the ways that they are used (i.e., queried). After that, any given domain can be classified as malicious or legitimate using the tree.
In [3], Antonakakis et al. introduce Notos to build models of known legitimate and malicious domains using 18 network-based features, 17 zone based features, and 6 evidence-based features. These models are then used to compute a reputation domains from non-DGA domains by using the both linguistic and IP features.
In [8], Yadav et al. measure K-L with unigrams, K-L with bigrams, jacquard index, and edit distance from training data set, and then use them to detect DGA-based botnets. Later on they introduced another method that utilizes entropy of NXDomain and C&C domains for detection. Techniques and heuristics for detecting DNS blacklist (DNSBL) reconnaissance activity, where botmasters perform lookup against the DNSBL to determine whether their spamming bots have been blacklisted, is suggested in [13].
The complex work by Bilge et al. [5] uses passive DNS analysis, examines a wide set of DNS traffic features and in corporate machine learning techniques. In [19] authors introduced BotDigger, a system that detects an individual bot by only using DNS traffic collected from single network. It utilizes a chain of evidence, including quantity, temporal and linguistic evidence.
In [12], authors introduced a framework to analyze domain name system traffic, such as NXDomain queries, at several premier Top Level Domain (TLD) authoritative name server to identify malware related domains. Ibrahim Ghafir et al. have described a methodology for detecting any connection to malicious domain. This detection method based on blacklist of malicious domains. The blacklist is updated automatically and the detection is in the real time. This methodology is applied on pcap files which contain traffic to malicious domains [15].
Pengkui luo et al. described a systematic study of DNS failure using a large ISP datasets. It demonstrates that attackers are employing a disparate domain name patterns for their malicious activities. A comprehensive evolutionary learning framework is used to detect diverse clusters of suspicious DNS Failures [14].
Guodong Zhao et al. introduced a novel system IDS placed at the network egress points to detect malware infection inside the network combined with DNS traffic analysis. Some new features are extracted and a

reputation engine is built on big data which include approximately 400 million DNS queries [16].

III. METHODOLOGY

In this section a methodology is proposed for detecting any connection to malicious domain/IP. Our detection method is based on a blacklist of malicious domains/IPs. As it is shown in figure 1, we pass the network traffic to the analysis engine. The analysis engine extracts all the DNS traffic and extract the DNS parameters. We analyze all the DNS requests and match the query with the reputation database the reputation database of malicious domains/IPs is updated automatically each day.

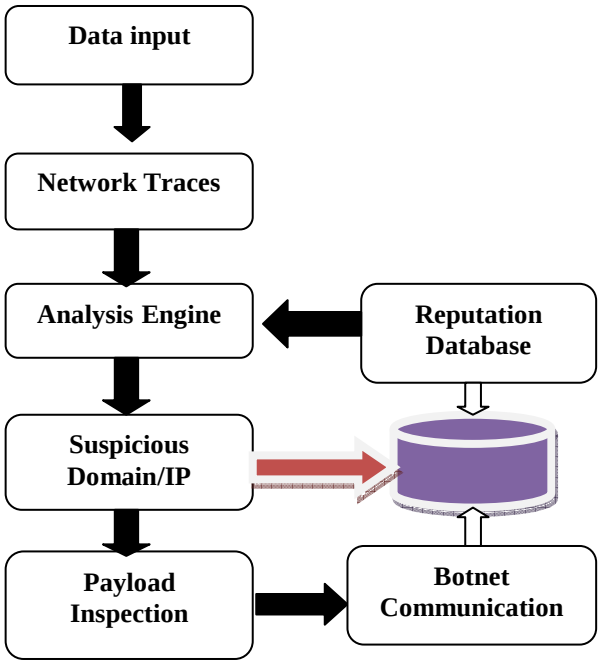

Figure 3: Methodology for malicious Domain/IP detection

If the DNS query matched with Blacklist then an alert is generated automatically by analysis engine.

Figure 3 shows the implementation of developed Analysis and Detection Engine. Firstly pcap traces are extracted from a Honeypot machine corresponding to a malware captured and process them to a analysis engine. The purpose of analysis engine is to decode the DNS packets through DPKT library and store the extracted information in the form of "Query_type, Query_ name, query_ID, Query_Response, detected URL, Resolve IP, websense, Reputation status IP and Domain/IP Reputation status" into a MySQL database. Any researcher may use this database as a feed for further analysis and applying data mining techniques on generated data sets. The motivation behind this module is to develop a database of malicious domains/IP extracted from network traces of a Honeypot machines. The domains/IP extracted are further checked with certain reputation scanners for their reputation status. If a domain/IP is found malicious, then it is indentified two indications a) Honeypot activities b) Reputation status which leads for further payload inspection of suspicious domain. The archive of domain/IP may further be useful for blacklist databases.

The major blocks of the developed system are:

A) Data Extractor

A python based shell script is developed which fetch the network traces corresponding to a dropped malware on a Honeypot machine. The complete session of the network traces are extracted during which a malware is detected by a Honeypot machine. Further these traces are submitted to analysis engine for processing and extracting the DNS related intelligence.

B) DNS Analysis Engine

An analysis engine is developed which take those network traces, decode the DNS traffic with respect to source and destination and its metadata. The extracted intelligent data are first checked with repudiation databases for determination of suspicious domain.

C) Reputation Scanner

The reputation status of extracted domains/IPs are checked with integrated database of global feeds who provide the list of suspicious domains/IPs such as MalwareDomains, Zeus Tracker, Malware Bytes, Palevo Tracker etc. Presently 12 such feeds are integrated into main system and more than 60,000 command and control IPs are recorded into database and which are currently active domains. The assumption here is that the technique used by the reputation scanners to determine a domain/IP as suspicious, is the indicator for analysis module to further inspections.

D) Botnet attributes

The extracted information from payload inspections are botnet attributes in the form of command and control domains/IP and its communications with foreign IP address. The set of bot commands are extracted and recorded into databases such as PING, PONG, PRIMSG, NICK, JOIN etc.

IV.EXPERIMENTS AND RESULTS

Bin ID	MD5 hash	Virus Total Label	Suspicious domain/IP detected

1847	560498d6c2133ebeaa40e480c097dbf4	[Spammer*Win32/Tedroo.A]	**198.144.156.187 hub.werbeayre.com**
1920	Bcc24e902f7f4cee608adb9608282d57	PWS-Zbot	**198.144.156.187 hub.werbeayre.com**
1924	13fc8c1ce61e823d277e12ef92fbb205	WormWin32/Pushbot.gen	**198.144.156.187 hub.werbeayre.com**

Table 1: Suspicious Domain/IPs

As shown in the above table, the domain hub.werbeayre.com is detected as suspicious domain by threat seeker after DNS packet decoding by Analysis engine. It is an indication for further analysis of this malicious domain for further extraction of botnet infections.

Deep Packet Inspection (DPI):

1) There is lots of packets exchanged between the source and suspicious domain as shown in the below screenshot. The suspicious domain is making the full TCP 3-way handshake to the infected machine.

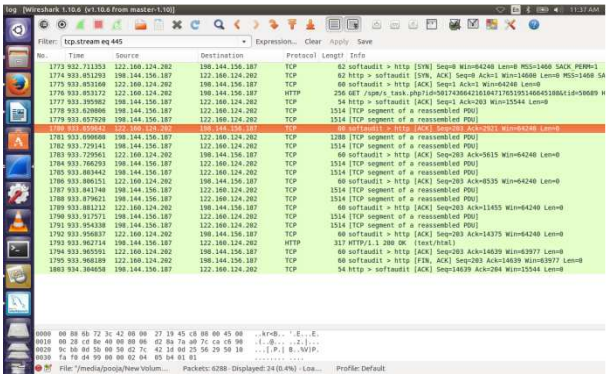

Figure 4: TCP 3-way handshake with suspicious domain

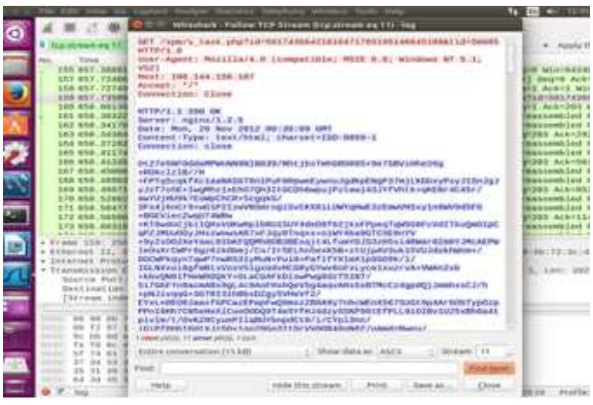

Figure 5: Payload inspection of suspicious domain

2) After full 3-way TCP handshake with suspicious domain, it is further noticed that it was exchanging some encrypted data with the infected system. It is clearly shown in the header of the payload that it is trying to host some PHP pages on a host machine through GET command. Further certain Bot commands are observed in communication in the form of HTTP Bot infections with its version number. In content/payload inspection, the mechanism is able to populate the database with botnet communications in the form of IRC commands exchanged with bot infected machines and remote command & control server.

Command and control server identifications- as shown in the above analysis and table (table no), the suspicious domain is trying to control many malwares (3 malwares in above analysis), which is the clear feature of botnet infections from command and control server. There are other certain communications recorded to determine the botnet command & control domains/IP.

V. CONCLUSION AND FUTURE WORK

In this research, the technique of DNS traffic is applied to botnet infection determinations. There are certain methods are available for botnet detection but as per the research conducted, a few data sets are available for determination of botnet suspicious domains. The Honeypots based research is globally available but it only give the data sets in the form of malwares and network traces on which there is a need to apply another technique to determine the botnet infections. The Suspicious domain/IP detected by developed DNS traffic analysis module which inspect and decode the DNS packets from PCAP dump is also determined as malicious Domain/IP by third party reputation scanners as well the same is validated through SNORT IDS engine incorporated with emerging threat [ET] Signatures & alert produced by SNORT engine is determining the suspiciousness of detected Domain/IP. Through research investigation, it is formed that the dataset of malicious domain is not publically available. Instead few services are there which provide the list of different datasets, but the problem with these available datasets is that they are non uniform, non reliable and not accurate. Therefore the research motivation was to prepare the blacklisted Domains/IP through pcap dataset processing of honeynet network dump as well the malware samples captured on Honeynet. The development of GUI console for further visualization of data set is remained in these research implementations. The integration of reputation scanner specific to botnets is a continuous process to improve the system.

REFRENCES

[1] Wang, Haining, Danlu Zhang, and Kang G. Shin. "Detecting SYN flooding attacks." INFOCOM

2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE. Vol. 3. IEEE, 2002.

[2] Ramachandran, Anirudh, Nick Feamster, and David Dagon. "Revealing Botnet Membership Using DNSBL Counter-Intelligence." SRUTI 6 (2006): 49-54.

[3] Antonakakis, Manos, et al. "Building a Dynamic Reputation System for DNS." USENIX security symposium. 2010.

[4] Yadav, Sandeep, and AL Narasimha Reddy. "Winning with DNS failures: Strategies for faster botnet detection." International Conference on Security and Privacy in Communication Systems. Springer Berlin Heidelberg, 2011.

[5] Bilge, Leyla, et al. "EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis." NDSS. 2011.

[6] Krmicek, V. "Inspecting DNS Flow Traffic for Purposes of Botnet Detection."GEANT3 JRA2 T4 Internal Deliverable (2011): 1-9.

[7] Edwards, Benjamin, et al. "Beyond the blacklist: modeling malware spread and the effect of interventions." Proceedings of the 2012 workshop on New security paradigms. ACM, 2012.

[8] Yadav, S.; Reddy, A.K.K.; Reddy, A.L.N.; Ranjan, S., "Detecting Algorithmically Generated Domain-Flux Attacks With DNS Traffic Analysis," in Networking, IEEE/ACM Transactions on , vol.20, no.5, pp.1663-1677, Oct. 2012.

[9] S, Khattak; N.R, Ramay; K.R, Khan; A.A. Syed, and S.A. Khayam; "A taxonomy of botnet behavior, detection, and defense," IEEE Commun. Surveys & Tutorials, vol. 12, no. 2, pp. 898–924, Oct. 2013.

[10] Kara, A. Mert, et al. "Detection of malicious payload distribution channels in DNS." 2014 IEEE International Conference on Communications (ICC). IEEE, 2014.

[11] Binsalleeh, H.; Kara, A.M.; Youssef, A.; Debbabi, M., "Characterization of Covert Channels in DNS," in New Technologies, Mobility and Security (NTMS), 2014 6th International Conference on , vol., no., pp.1-5, March 30 2014-April 2 2014.

[12] Thomas, Matthew, and Aziz Mohaisen. "Kindred domains: detecting and clustering botnet domains using DNS traffic." Proceedings of the 23rd International Conference on World Wide Web. ACM, 2014.

[13] Kührer, Marc, Christian Rossow, and Thorsten Holz. "Paint it black: Evaluating the effectiveness of malware blacklists." International Workshop on Recent Advances in Intrusion Detection. Springer International Publishing, 2014.

[14] A. M. Kara, H. Binsalleeh, M. Mannan, A. Youssef and M. Debbabi, "Detection of malicious payload distribution channels in DNS," 2014 IEEE International Conference on Communications (ICC), Sydney, NSW, 2014, pp. 853-858.

[15] Pengkui Luo; Torres, R.; Zhi-Li Zhang; Saha, S.; Sung-Ju Lee; Nucci, A.; Mellia, M., "Leveraging client-side DNS failure patterns to identify malicious behaviors," in Communications and Network Security (CNS), 2015 IEEE Conference on , vol., no., pp.406-414, 28-30 Sept. 2015.

[16] Ghafir, Ibrahim, and Vaclav Prenosil. "DNS traffic analysis for malicious domains detection." Signal Processing and Integrated Networks (SPIN), 2015 2nd International Conference on. IEEE, 2015.

[17] Zhao, G. U. O. D. O. N. G., et al. "Detecting APT Malware Infections Based on Malicious DNS and Traffic Analysis." IEEE Access 3 (2015): 1132-1142.

[18] Ichise, Hikaru, Yong Jin, and Katsuyoshi Iida. "Detection Method of DNS-based Botnet Communication Using Obtained NS Record History." Computer Software and Applications Conference (COMPSAC), 2015 IEEE 39th Annual. Vol. 3. IEEE, 2015.

[19] Tu, Truong Dinh, Cheng Guang, and Liang Yi Xin. "Detecting bot-infected machines based on analyzing the similar periodic DNS queries." 2015 International Conference on Communications, Management & Telecommunications (ComManTel). IEEE, 2015.

[20] Zhang, Han, et al. "BotDigger: Detecting DGA Bots in a Single Network." (2016).