Analysis of Botnet Classification and Detection Techniques: A review

1 Blessing Nwamaka Iduh, 2Obikwelu Raphael Okonkwo
1Academic Researcher, 2Professor of Computer Science
1[1][2]Computer Science Department,
[1][2]Nnamdi Azikiwe University, Awka, Nigeria.
1bn.iduh@unizik.edu.ng, 2ro.okonkwo@unizik.edu.ng

Abstract: Botnets have in recent times, become a very major challenge in the cyberspace. The Global Internet has experienced tremendous attacks designed mainly to disable internet infrastructure on one hand, while in most other cases people and organizations are targeted. At the center of these attacks are a group of compromised computers that have been infested and are now controlled by a Botmaster. These systems are usually located in schools, business premises, homes and government agencies which, unknown to their owners, are infested and controlled by Botmasters for malicious activities. This paper presents an analysis of Botnet with respect to its architectural representation, classification and characterization in order to help coordinate the development of new technologies to face this serious security threat.

Index Terms - Peer to Peer (P2P), Botnet, Command and Control Channel (C&C), Botnet Detection, Cyber Security

1.0. INTRODUCTION

The word Botnet, is a combination of the words robot and network. It is used to describe a group of compromised computer systems that are usually connected to a central controller called a Botmaster. The Botmaster uses command and control (C&C) channels, to manipulate these infested computers. A single infested system is known as a bot, while a network of infested devices is referred to as a Botnet. Botnets are created by the Botmaster for communication infrastructure to perform malicious activities like spamming, click fraud, identity theft, phishing attacks and distributed denial of service attacks. Characterizing existing Botnets will help to coordinate and develop new technologies to face this serious security threat. Systems that are connected to the internet have the chances of getting infested and become part of a Botnet. Several Researchers like [1], [2], [3] amongst others, have worked extensively on Botnets with respect to its classification and functionalities. According to [4] Botnets can be classified according to their attacking behavior, command and control (C&C) mechanism, rallying mechanisms, communication protocols, evasion techniques and other activities such as abnormal system calls and traceable DNS queries. [5], in their survey of the categories of Botnets, noted that Botnets can be classified into six basic types based on the C&C channel used. These include, according to them, IRC (Internet Relay Chat) Botnet, P2P (Peer-to-Peer) Botnet, HTTP (Hyper Text Transfer Protocol) Botnet, Mobile Botnet, Cloud Botnet and the hybrid Botnet which according to them, is the combination of all the types of Botnet Structures. These Botnets can be grouped into two categories, and these include, the Centralized Botnets and the Decentralized Botnets.

I. Centralized Botnets Architecture

Centralized Botnets usually have a centralized network topology. They consist of a Command and Control (C&C) server that a Botmaster uses to send commands to their bots. A Botmaster will issue a command by posting a message to this channel. The C&C server pushes the command to bots which then invoke the command with a relatively low latency. As long as the infested machine is on and has an active Internet connection, it will remain connected to the C&C server awaiting commands. In the survey presented by [6], they noted that the centralized Botnets are similar to a client server model. All the bots act as clients and connect to the centralized servers. The servers initiate commands to these bots which are connected to it. In the centralized Botnet architecture, the Botmaster can monitor all the bots and receive direct and accurate feedback along with the status of the Botnet. This tallied with the work of [7], where they noted that, the centralized Botnet, has one center point that is accountable for exchanging commands and malicious data between the Botmaster and Bots. They also recorded that in this centralized model, the Botmaster chooses a host which is usually a high bandwidth computer, to be the central point which will be the Command-and-Control server of all the Bots. Centralised Botnets belong to the first generation of Botnets where the Botmaster controls the bots through a single C&C server at a single point in form of a star topology. According to [8], three types of centralized Botnets architecture exit and these include the Internet Relay Chat (IRC) Botnets, Internet Messaging (IM) Based Botnets and the Hypertext Transfer Protocol (HTTP) Based Botnet architecture.

II. Internet Relay Chat (IRC) Botnets

Internet Relay Chat (IRC) is a text-based chat-system that organizes communication in channels. According to the survey done by [5], the Botmaster exploits the Internet Relay Chat (IRC) as the C&C Channel to communicate and control the bots. Also, [8] recorded that IRC is an on-line text-based instant messaging protocol that works on client-server architecture. They further stated that IRC is capable of connecting hundreds of clients through multiple servers, and that clients can be contacted using one-
to many or one-to-one relationships. This feature makes the IRC very suitable for use to create and control a Botnet. Also they added that each bot performs malicious actions, based on the commands received from the centralized IRC server. Furthermore, [9], stated that IRC bots operate in a typical star network. He explained that all the bots connect to a single IRC server to listen for command. IRC servers can work together by allowing IRC clients reach an IRC network through multiple IRC servers. This method is used by Botmasters of larger Botnets to be able to handle a large amount of bots. IRC is a client/server based model. A client is identified by its nickname. When a client sends a message to another client (nick), the message is sent to server with message and target nickname information and server delivers message to the target client. Due to this mechanism, a client could send message to any client that is connected to that IRC server and this makes a client to communicate with multiple clients simultaneously.

III. Internet Messaging (IM) based Botnets
Internet Messaging (IM) based Botnets are not a very common type of Botnets. They use Internet Messaging protocols such as AOL, MSN or Yahoo. The major strength of this type of Botnets is that there is an already-made platform which is usually used to carry out their activities, and on the other hand, the weakness is that a temporary link cannot be created, this means that there is usually the possibility of tracing back, through the vendors, to the source of the message and it usually needs to be registered manually by entering CAPTCHA code. Vendors like AOL, MSN and YAHOO have adequate measurements against CAPTCHA descriptors. Therefore IM-based Botnets have a limited number of login accounts as time is required to create an email account on AOL, MSN or YAHOO. Also a one ID cannot be logged into multiple computers that make IM based Botnets more limited.

IV . HTTP/Web Based Botnets
HTTP Botnet, according to [5], is a centralized Botnet that uses HTTP protocol as the C&C server. They recorded that the Botmaster uses the HTTP protocol to hide their activities among the normal web traffic. They have the capability to escape easily from the current detection methods that exist, like firewalls and other Intrusion Detection Systems. They also stated that, the bots use specific URL or IP address defined by the Botmaster to connect to specific web pages; HTTP is the protocol that is most commonly used for the delivery of data over the Internet. Contents like Videos, Audios, images and text files are shared and transported in uploads and downloads on daily basis and HTTP is available in nearly every network connected to the Internet since this is the primary protocol for web browsing and is rarely filtered. This is especially interesting for Botnet operators, because it makes the protocol viable as a command-and-control protocol. HTTP based Botnets use HTTP to logon to the website controlled by the Botmaster and leave information on the website that can be interpreted and executed by the bot.

2.0. Decentralized Botnets
According to [6], the decentralized model allows the bots to act autonomously by connecting several infested machines on a Botnet rather than to a command and control center. These links are useful for communicating with other bots within Botnets. Also, according to [4], a decentralized network topology has no single Command and Control (C&C) server that is known to all Botnet members. Some examples of the decentralized Botnets architecture include; mobile based Botnet architecture, email based, peer-to-peer (p2p) architecture.

2.1 Mobile Based C&C Botnet Architecture
According to [5], The Botmaster utilizes applications on smartphones and other mobile devices to propagate malicious scripts/commands without the knowledge of the user. A major strength of Mobile-based C&C is that the Botmaster can easily communicate with the root node because of tree topology form of communication. On the other hand, a major weakness of Mobile-based C&C is that it needs a node list which is to be operated on infested mobile phones. Some examples of Mobile based Botnets include, SMS based Botnet, Bluetooth Based Botnets and Android based Botnets. According to [10], SMS based Botnets, detect the C&C messages and makes them invisible to the mobile device owners. Moreover, it uses popular application names and icons like Google's search application. As reported by the Symantec research lab, it is particularly designed to be a spyware.
SMS Based Botnets have sophisticated capabilities of recording voice call conversations and even surrounding sounds. Bluetooth Based C&C Botmaster on the other hand utilizes the vulnerabilities of Bluetooth technology and make it as the Command and Control (C&C) channels. A Bluetooth based C&C is responsible for command transmission between the Bluetooth-enabled devices. It enables faster communication by simplifying the authentication and authorization process.

2.2. Email Based C&C Botnet Architecture
According to [9], the Email Botnet architecture is shown to feature a stealthy C&C channel, blending the commands with either Spam or Non-Spam flagged email. The Email Botnet is not very popular yet. Once providers introduce preventive measures that prevent bots softwares, Botmasters switch to non-Spam blending. Non-Spam blended email mimics normal usage behavior, making automatic detection harder. Experiments with a prototype confirmed the feasibility of hacking the users email account, and using this for encrypted C&C communication, as well as successfully manipulating Google’s, Yahoo’s and other Email Providers.

2.3 Peer-to-Peer (P2P) Botnets
According to [9], P2P networks works differently as compared to traditional networks with a centralized server. In a P2P network many devices, can perform server functions. Also, every device that performs such functions is capable of performing all other
server functions, avoiding the vulnerability of dependence on a central device. Decentralization gives P2P Botnets a high resistance against both targeted and random attacks. It is very difficult to detect P2P Botnets. According to [6], P2P Botnet is much harder to be suspended and it’s not easily manageable, because transferring command is slow as compared to centralized Botnet. P2P Botnets are very difficult for defenders to track because single point of failure in P2P Botnet does not create significant disruption. According to [4], peer-to-peer (P2P) Botnets, such as Trojan.Peacomm Botnet [11] and Stormnet [12] emerged due to the fact that attackers realized that there were several limitations in the traditional centralized Botnets.

Fig. 2.1 clearly illustrates the peer-to-peer Botnet architecture while fig. 2.2 shows the centralized Botnet architecture. In the peer-to-peer architecture, there is no centralized server, and bots are connected to each other topologically and act as both C&C server and client. P2P Botnets have shown more threats over traditional centralized Botnets. As the next generation of Botnets, P2P Botnets are more robust and difficult for security community to defend. Furthermore, [8], in his work on Peer-to-Peer Botnet detection based on bot behavior, explained that unlike centralized Botnets there is no single attacker in P2P Botnet architecture and that therefore, the detection of these Botnets becomes comparatively more difficult.

![Peer-to-Peer model](image1)

**Fig. 2.1. Peer to peer network model (Augument Systems, 2017).**

![Peer-to-peer Botnet architecture](image2)

**Fig. 2.2. Peer-to-peer Botnet architecture [13]**

[13] further explained that a P2P network is one in which two or more computers are connected and share resources like content, storage devices, CPU cycles, printers, etc by direct exchange, rather than going to a server or authority which manages centralized resources. They stated that one of the major uses of P2P networks today is file-sharing, and there are various P2P file-sharing applications, this includes;

(a) Instant Messaging Systems (Skype) - Skype is an instant messaging app that provides online text message and video chat services. Users may transmit both text and video messages, and may exchange digital documents such as images, text, and video. Skype allows video conference calls.

(b) Digital currency which is popularly known as Bitcoin. **Bitcoin** is the first decentralized peer-to-peer payment network that is powered by its users with no central authority or middlemen. It is a type of digital currency in which encryption techniques are used to regulate the generation of units of currency and verify the transfer of funds, operating independently of a central bank.
Bitcoin has become a hot commodity among speculators. Bitcoins can be used for online transactions between individuals according to a report by [14], it is also stated that nobody owns the Bitcoin network just as no one owns the technology behind email or the Internet. Bitcoin transactions are verified by Bitcoin Miners which have an entire industry and Bitcoin cloud Mining options. The decentralized nature of the bitcoin system, makes it a peer-to-peer network system.

(c) Netsukuku - Netsukuku is a mesh network or a peer to peer protocol that generates and sustains itself autonomously. It is designed to handle an unlimited number of nodes with minimal CPU and memory resources. This is due to the feature it has which is the fact that it can be easily used to build a worldwide distributed, anonymous and uncontrolled network, separated from the Internet, without the support of any servers, ISPs or authority controls. In this network, computers are linked physically to each other, therefore it is not built upon any existing network. Netsukuku builds only the routes which connects all the computers of the net. In other words, Netsukuku replaces the level 3 of the ISO/OSI model with another routing protocol.

(d) Currency Fair is another P2P network. It is an online peer-to-peer currency exchange market place with its headquarters in Ireland also having employees in UK, Australia and Poland (Brien, 2017).

(e) Torrent Applications is another P2P network. A torrent application is a program that allows peer to peer connection by users in such a way that users visit the site, upload files in bits. A torrent file is a computer file that contains metadata about files and folders to be distributed, and usually also a list of the network locations of trackers, which are computers that help participants in the system find each other and form efficient distribution groups called swarms. A torrent file does not contain the content to be distributed; it only contains information about those files, such as their names, sizes, folder structure, and cryptographic hash values for verifying file integrity. Depending on context, a torrent may be the torrent file or the referenced content. There are several Torrent applications that exist, like bitTorrent, Vuze, uTorrent, BitLord, TorrentFlux, BitTornado, rTorrent, etc.

(f) Gnutella DC++ - Gnutella is a file sharing protocol that defines the way distributed nodes communicate over a peer-to-peer (P2P) network. Like Napster, Gnutella is often used to share music files and has been an object of great concern within the music publishing industry (Gnutella, 2017).

### 3.0. Botnet Attacks

There are several types of Botnet attacks. The first attack we described is very familiar to anyone who uses email.

(a) Email spam: Email spam is an unsolicited email that is sent, in bulk, to a large number of accounts. Typical email spam may be directing users to phishing sites or distributing malware. Most of the email spam is sent by Botnets. The Botmasters will rent out their Botnets to perform a spam campaign. Email spam from Botnets may be more difficult to prevent than traditional email spam due to the number of different computers and email accounts being used in the attack.

(b) Another attack is the Denial-of-Service (DoS). This attack targets a system’s resources such as their central processing unit (CPU), memory or network connections. One common resource attack is to overload the capacity of a network device by sending a large number of packets. This prevents legitimate communications from occurring.

(c) Synchronization flood is another common attack, which affects a victim’s CPU, this is a situation whereby Transmission Control Protocol (TCP) SYN packets are continuously sent to a system. Each packet requires a lookup by the CPU for existing connection information and the creation of an entry if none is found. These incoming connection attempts are queued and can quickly prevent other legitimate connection attempts.

(d) Distributed Denial of Service (DDoS) attacks is another form of Botnet attack. It is an extension of a DoS attack that originates from many different systems simultaneously (i.e., distributed) and, which increases the effectiveness of the attack. Botnets are a perfect vessel to perform DDoS attacks. Botmasters make money from this type of attack either through extortion or as a paid service to a competing organization. Extortion may occur by threatening a company with shutting their servers down over a busy period such as the holiday season, which could result in a huge loss to online sales. A company may choose to pay the Botmaster rather than risk greater financial loss through reduced online sales and customer dissatisfaction.

(e) Click fraud is another P2P Botnet attack. It is an attack that targets the pay per impression (i.e., click) advertising market. According to (Wilbur & Zhu, 2009) Click fraud is the practice of deceptively clicking on search ads with the intention of either increasing third-party website revenues or exhausting an advertiser's budget. Search advertisers are forced to trust that search engines detect and prevent click fraud even though the engines get paid for every undetected fraudulent click. A Botmaster commands bots to click on advertisement banners, which pay a small amount of money per click on an account controlled by the Botmaster.

(f) Phishing attack is an attack that is based on redirecting users often by email or instant message, to a fake website that is meant to look like a real service that the user uses. When they log in or perform transactions they are inadvertently giving away their
credentials and private information. Botnets are used in phishing attacks in a few ways. They may be used to spam phishing emails, host phishing sites, or an infested system may control a user’s browsing experience such that they are redirected to a phishing site. This last attack is particularly problematic because what the user sees is under the complete control of the Botmaster.

(g) Information theft is yet another P2P attack. This is an attack just like Phishing attacks, that has the goal of stealing users’ confidential information either with the objective of selling the stolen information or using it directly. This can be performed in a number of ways for example: Keyloggers (Keykloggers, 2017) record and send a user’s keystrokes to an attacker’s database; and, form grabbers record information entered into online forms, such as credentials and credit cards, during a user’s browsing activity. Some security measures involve graphical keyboards. However, this type of malware often comes with screenshot capture capabilities; and the malware may also steal information from protected storage such as cookies and password files.

4.0. Botnet Detection
Botnet detection involves the identification of bots in the machine or network so that some sort of remedy can be done. In recent years Botnet detection has been a hot topic in the research community due to increase in the malicious activity. According to [6], the key features and characteristics of bots is considered as a critical step when dealing with Botnet detection. [13], stated in their survey of Botnet detection techniques, that detecting a Botnet often requires advanced analytic capabilities which involves related tasks performed by the Bots. They further explained, that an enterprise network might have access to Dynamic Host Configuration Protocol (DHCP) logs, Domain Name Service (DNS) resolver data, address allocation data, complete packet traces for each host, email server logs, policy data, as well as antivirus scanning logs. According to [14], DNS Data is Data regarding name resolution, and can be obtained by mirroring data to and from the local DNS servers or resolvers and it can be used to detect the behaviours of Botnet attacks such as email spam, as well as the communication behaviors of Botnets, such as DNS lookups for suspicious domains. Also, they explained that netflow data represents information gathered from the network by sampling traffic flows and obtaining information regarding source and destination IP addresses and port numbers. They further explained that such data is useful for identifying malicious communication patterns and course grained attacks and their visibility is often limited to the peering edge of a network, missing large amounts of backbone (ISP) or switched (enterprise) traffic. In addition, they explained Packet Tap Data provides a more fine grained view than netflow and offers an attractive deployment model and also, it is generally more costly in terms of hardware and computation. And finally, they explained the Address Allocation Data as data that enables Knowledge of where hosts and users are in the network which is a very powerful tool for identifying reconnaissance behaviors of bots and for tying them to specific machines or users. According to [6], the major characteristics of a bot malware are related to network activities since the bots require some sort of interaction with the command and control servers. Some of the common activities one could monitor to detect Botnets, according to them, are; opening of specific ports, establishing a number of unwanted network connections, downloading and executing files and programs, creating new processes with well-known names, disabling antivirus software and so on. Several researchers like [1], [2], [3], to mention but a few, have worked on the Botnet Detection. In general, two major approaches exist for detection of the Botnets; signature-based and anomaly based detection.

4.1. Signature Based Botnet Detection
This method is based on recognizing the characteristics of Botnet traffic which are also known as “signatures”. The signature-based methods rely on some set of predefined rules regarding anomalous traffic and packet level signatures. These techniques mainly perform packet level analysis by using deep packet inspection (DPI) to match signatures of malicious payloads. This class of detection techniques operate on all the three phases of Botnet life-cycle and is able to detect known Botnets with bounded number of false positives (FP). The main disadvantage of signature-based approaches is that they can only detect known Botnets, and to use the approach efficiently, constant update of Botnet traffic signatures is required. Also, these techniques are responsible for various evasion techniques that change signatures of Botnet traffic and malicious activities of bots, like encryption and obfuscation of C&C channel, Fast-flux and DGA techniques, etc. Researchers like [15] and [16] among others, have studied the signature-based detection and their results are applicable for known bots. In their approach, every packet is monitored and compared to the pre-configured signatures and attack patterns in the database. Even though their approach can detect some Botnets, the signature database needs to always be updated to detect the new bots. In addition, bot-masters obfuscate the bots by novel packers to avoid detection by the signature-based approaches. In the survey presented by [6], two approaches for signature based Botnet detection where presented. The first approach according to them, is based on locating Honeynets in the network and the other approach is Intrusion Detection System (IDS). The first step to mitigate different kinds of threats is to deeply understand the internal working of Botnets and the techniques employed by them. Thus, honeynet based approach is developed to the study and gathering of information about Botnets. After collecting information, it is possible to learn and understand the technology used and perform a complete analysis of the main Botnet characteristics. A honeypot according to them, can be defined as an environment where vulnerabilities have been deliberately introduced to observe attacks and intrusions. Honeynets can be used to obtain bot binaries and infiltrate those Botnets. The Honeynet project deploys an architecture that consists of a honeynets and the honeypot network. However, Honeynets can only track and capture activities that directly interact with them. They will not capture attacks against other systems. IDS Botnet detection approaches can be either a signature- or anomaly-based technique. A signature-based Botnet detection technique uses the signatures of current Botnets for its detection. This method has several advantages, such as (i) very low false alarm rate (ii) immediate detection (iii) easier to implement and there is better information about the type of detected attack. Furthermore, there is also the Network based Botnet detection technique.
4.2 Anomaly Based Botnet Detection
According to [17], anomaly-based detection technique focuses on detecting traffic anomalies that are associated with Botnet operation. These traffic anomalies vary from easily detectable changes in traffic rate and latency, to higher finite anomalies in flow patterns. Some of the most prominent anomaly-based approaches detect anomalies in packet payloads[7], DNS traffic, Botnet group behaviour[18], etc. The anomaly-based detection can be achieved through the use of different algorithms stretching from machine learning techniques, to the statistical approaches, graph analysis, etc. Unlike the signature-based detection technique, the anomaly-based detection is usually able to detect new methods of malicious activity that displays anomalous Botnet related characteristics. According to [17], one of the newest and very promising group of anomaly-based detection technique methods are the detection methods that use machine learning for detection of bot-related traffic patterns. Machine learning is used because it offers the advantage of automated recognition of bot-related traffic patterns. Also, it provides the capability of recognizing the patterns of malicious traffic without an initial knowledge about the characteristics of such traffic but simply inferring knowledge from the Botnet traffic traces that’s available.

4.3 Data Mining Based Botnet Detection
Some other Botnet detection techniques were presented by [19] in their study on Revealing the Criterion on Botnet Detection Technique. They presented these techniques as: Data Mining-based, which they explained as one of the effective technique for Botnet detection since it can be used efficiently to detect Botnet C&C traffic by using machine learning, classification and clustering approach.

4.4 Host based and Network based detection techniques
The host-based approach, monitors the network traffic for signs of bot-infested machines. The Network-based detection according to them, focuses on monitoring network traffic in: (i) trying to detect bots by checking the traffic patterns or watching out for contents that reveals some bot related activities. (ii) Analyzing the traffic to filter hosts that have similar patterns or that reacts to the same functions. Furthermore, Signature-based Detection was described by them, as Similar to anomaly-based techniques, it learn and gain knowledge of useful signatures or behaviors from existing Botnet. They also opined that the signature-based solution is useful for detection of known Botnet rather than the unknown bots. Network-based detection technique is based on analyzing network traffic to enable identification of the presence of compromised computers. It is classified as either signature- or anomaly-based methods. Also, it can further be classified as passive or active depending on the stealthiness of their operation. The passive approach operates only based on observation hence they do not interfere with Botnet operation which makes them stealthy in their operation and difficult for the attackers to detect. On the other hand, Active detection methods represent a more aggressive and intrusive methods that actively get in the way of Botnet operation by interfering with malicious activities or the C&C communication of the bots.

4.5 Machine Learning Based Botnet Detection
Machine learning techniques are also used in detecting bots[20]. Machine learning algorithms do not need explicit signatures to classify malware programs but rather is based on finding common features and correlating different activities of the malware. [21], proposed a machine learning technique for Botnet detection that uses network statistics. These statistics involve bytes per second, packet duration per second of some protocols used for chatting such as IRC protocol. [15], presented a detection approach that examines the use of network flow characteristics like bandwidth, packet timing, and burst duration to show evidence of Botnet activity by filtering traffic that are unrelated to Botnets, to reduce the amount of data that is being processed. The next step is to classify the rest traffic into a group that is likely to be part of a Botnet using three machine learning classification algorithms, namely J48 decision trees, Naive Bayes and Bayesian Network. Finally, they correlate the remaining traffic with each other to find clusters of flows that share similar timing and packet size characteristics which is part of the activity of a Botnet.

5.0 Conclusion
In this paper, an analysis of Botnets with respect to their architectural representation, classification and characterization was presented. The paper discussed centralized and decentralized Botnet architectures. The Internet Relay Chat (IRC), Internet Messaging (IM) and the Hypertext Transfer protocol (HTTP) based Botnets were reviewed as examples of the centralized Botnet architecture. While the Mobile Based, Email Based and the Peer-to-peer Botnet architectures were presented as examples of the decentralized Botnet. Some examples of P2P Botnet were given as; Instant messaging systems, digital currency popularly known as bitcoin, Netsukuku and Torrent application. Some known Botnet attacks were given as email spam, DDoS attacks, synchronization flood, click fraud, amongst others. Also Botnet detection techniques were shown to be; signature based, anomaly based, data mining based and machine learning based.

References
[3] M. A. S. &. P. M.Muthulakshmi, "BOTNET DETECTION BASED ON COARSE GRAINED PEER-TO-PEER


