TEST Engineering & Management

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

A Botnet Taxonomy and Detection Approaches

[1]Zahian Ismail, [2]Aman Jantan, [3]Mohd. NajwadiYusoff, [4]Muhammad Ubale Kiru
School of Computer Sciences, Universiti Sains Malaysia
[1]zahianismail@gmail.com,
[2]aman@usm.my,
[3]najwadi.usm.my,
[4]muhdujkiru@gmail.com

Article Info
Volume 82
Page Number: 3386 - 3408
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 20 January 2020*

Abstract:
This paper focuses on the study of Botnet and its Command and Control (C&C) structure. It also reviews the state of the art in machine learning-based Botnet detection system. In order to survive, Botnet implemented various evasion techniques, and one of the famousevasion technique is by manipulating an encrypted channel to perform their C&C communication. Therefore, we also look into the capabilities of machine learning approaches to detect these particular Botnet activities via encrypted channels. From the study, we show the effectiveness of machine learning in Botnet detection over an encrypted channel. The paper concludes by highlighting the limitations of the existing Botnet detection approaches and the way forward.

Keyword: *Botnet taxonomy, Botnet detection, encrypted channel, machine learning*

I. INTRODUCTION

Botnet is one of the significant concerns in the computer industry today, mainly because of the massive impacts caused by their operations. Botnet developed many capabilities, but unfortunately, it usesmost of those capabilities for attack purposes, such as performing a DDoS attack, spamming, malware campaign, and breaking down large networks. Fundamentally, Botnet acts as a carrier to amplify such attacks and cause massive Internet disturbance while utilizing substantial network resources in carrying out those attacks. In order to mitigate the Botnet attacks, we should be able to understand what Botnet is and how it works. Therefore, we discuss a detailed study of Botnet taxonomy and the detection techniques, particularly on detecting Botnet overan encrypted channel.

Bot is a shortform for robot[1], referring to the automated nature of Botnet operation and the fact that all bots follow the instruction of the botmaster. Botnetis also knownasa zombie network[2], [3], reflecting on how Botnet infects other hosts and turns them into infected machines. [4], [5] refer to Botnet as a well-organized network of infected machines thatare used to conduct malicious activities. These infected machines are controlled by a human handler known as both erder or botmaster. Going by the different Botnet definitions, one thing that is certain about Botnet is their enormous contribution to malicious activities.

Botnetoperations are based on three main components, namely bot, botmaster, and C & Cvserver, as depicted in Figure 1. Its Command and Control (C&C)structure differentiates Botnet with other malware. This C&C structure enables Botnet to remotely control all the bots and distribute their commands for malicious intent. Usually, the owner of the infected computers isnot aware of Botnet compromising their machines. Incredibly, each bot can compromise hundreds or even thousands of new machines. Therefore, botmaster uses Botnet to launch a massive attack and cause massive disruption of services.

We organized the remainder of this paper as follows. Section 2 highlights the severity of Botnet attacks and its evolution, where we provide the background of several notorious Botnet attacks. Understanding the Botnet life cycle is crucial because of the relationship with Botnet C&C; hence, we include it in section 3. Section 3 also discusses the essence of Botnet,which is Botnet C&C comprising its architecture, various communication protocols, communication initiation, and communication direction. Later in the same section, there are Botnet variants which show how Botnet manipulates the existing technology and implements different architecture and communication protocols to evade detection by introducing new variants. Lastly, section 3talks about evasion techniques employed by Botnet. Section 4 investigates Botnet command and control over the encrypted channel by focusing on SSL/TLS channel. Botnet detection focusing on machine learning is in section 5. Section 5 additionally provides limitations of Botnet detection,especially the detection over the encrypted channel. Finally, we draw some concluding remarks in section 6.

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

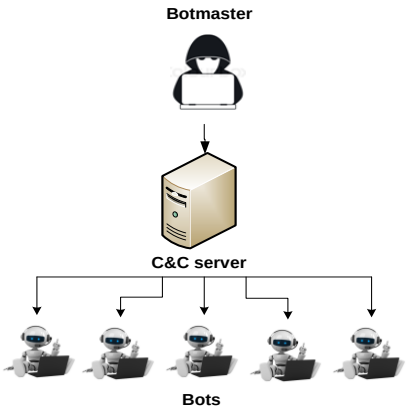

Figure 1.Coordination between botmaster, C&C server,and bots

II. Botnet Evolution and Attack History

This section is about the evolution of Botnet since Internet Relay Chat (IRC) until the recent IoT Botnet. This evolution shows how Botnet leverages different technologies to stay significant in the network and proves that each Botnet variant is different;thus, each requiresa dedicated approach. There is no absolute solution which can be applied on all Botnet variants. One of the Botnet survival mechanisms is by manipulating the SSL/TLS channel,and we provide several examples in this section. Other Botnet attacks discussedhere are meant to show how attackers launch attacks using Botnet. There are many Botnet attacks in history, and some of them are provided here.

Initially, bots offer legal services like simple games and messaging services in IRC. Bots originallyhave been utilized to assist the IRC channel management. Administering busy chat channels like IRC is time-consuming. Therefore, channel operators created bots to help manage the operation of the channel[6]. However,"malicious behavior evolved soon and resulted in the so-called IRC wars, one of the first documented DDoS attacks"[7]. IRC nuking (Nuke War) is a war among hackers in IRC. Eggdrop bot is one of the most powerful IRC war machines which was able to flood ICMP, nuke and quickly take over channels. Eggdrop is the oldest IRC bot and is still in active development. Furthermore, it is almost impossible to kill Eggdrop because of its robust features,and one of the features is SSL/TLS deployment to protect the Botnet traffic (eggheads.org).

In 2000, Michael Calce from Montreal,who is a well-knownMafiaboy,carried out a series of DDoS attacks at a massive scale. These DDoS attacks were targetedtowards large organizations' websites such as Fifa website, Amazon shopping site, Dell corporations,the email giant Yahoo, eBay online shopping website,andlastly CNN news network. The compromised computers in approximately 25 universities network and several private and company networks have been usedas zombies in launching the attacks. This case is considered to be the highest profile and high impact cybercrime case after Kevin Mitnik,which involved many investigators and resources from FBI and Royal Canadian Mounted Police (RCMP) at that time. Bill Clinton even organized a special meeting at White House, calling in chargeofficers, demanding explanations of the threat they have been facing at that time. On that state, he called that kind of attacks a digital Pearl Harbour [8]. The case of a 15-year-old Mafiaboyproves that to organize Botnet attacks did not requirehigh-level skills and sophisticated technology. The kid was not a brilliant student at school and hadlittle computer skills. Suchincidents indicate the ease of performing cyberattacks using Botnet.

Other Botnet attacks also recorded in 2008 was by an unknown attacker, who releaseda variant of worms known as Conficker which targetedMicrosoft operating systems.It was speculated that botmaster leased the Botnet infected by Conficker to infect millions of computers and established the attacks by sending spam, identity theft, phishing exploits,and other malicious agents. Conficker took advantage of Windows operating systems'vulnerabilities as well as dictionary attacks to launchan attack while forming a Botnet. It was reported that it infected thousands of systems,comprising business,government,and home computers in more than 190 countries, becoming the most highly known worm infection since 2003 [9], [10]. Cisco Security also observed that Conficker C&C traffic was using TCP port 443,whichSSL/TLS traffic typically use.

In 2009, an unknown attacker launched an attack targeting the government of the UnitedStates and South Korea official websites. The investigation conducted revealed thatthe virus had infected many personal computers which compelled them to visit government websites in the United States and South Korea simultaneously and overload the network with bogus requests, resultinginservice failure. The similarity between the two attacks on the government websites is that theywere launched through DDoS attacks, which involved massive compromised computers in launching the attack commands and are operated remotely by the botmaster[11].

In 2011, Microsoft and its partners worked aggressively to take down Botnet and by the US court order dismantled Rustock Botnet. Rustock was responsible for as much as 48% of all spam worldwide and was estimated to have more than one million compromised machines. Microsoft with US Marshall Service physically captured evidence on-site and took the affected servers from hosting providers for analysis.

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

With help from upstream providers, they successfully severed the IP addresses that control the Botnet,thereby cutting off the Botnet communications[12]. Symantec stated that 70% of spam sent from Rustock utilizedTLS.

According to [13], the most sophisticated criminal Botnet in existence is known as the Zeus Botnet. GameOver Zeus is a variant of the Zeus family identified in September 2011. GameOver downloaded the Gameover application over an SSL/TLS channel from an infected web server. Gameover Zeus Trojan takes advantageof encryption for both malware circulation and C&C communications. GameOver Zeus is a credential-stealing malware which was primarily used by cybercriminals to harvest banking information.

GameOver Zeus also installed Cryptolocker, a variant of malware identified as ransomware to the compromised computers. This malware encrypted computer files (for example photographs, vital business record, police investigation files) in the victims' computers and got them to payhuge amounts of money, usually in between $300 to $750 before they could receivea decryption key. Terrified of the idea of losing crucial data, most of the victims directly paid, but some victims like the Pittsburgh insurance company refused to compile with the payment and chose to take down the malware. They successfully restored the data from the backup. Nevertheless, they lost money worth $70,000 and had to send their employees home during the recovery process. To takedown Botnet, the inventors of the Zeus redesigned a novel and robust structure, which came in three distinct layers of C&C infrastructure that made the Botnetchallenging to mitigate[14].

By the end of 2016, Mirai Botnet infected approximately 100,000 Internet of Things (IoT) devices over the networks and performed a DDoS attack on DNS providers that resulted in several websites crashing, including Twitter and Netflix [15]. Mirai successfully rendered the Internet unavailable for hundreds and thousands of people by attacking Dyn, an Internet giant that provides a large amount of the internet service used in the US [16], [17]. IoT devices especially smart home appliances, are the primary targets because most of them use default usernames and passwords. Mirai performed brute force using a table that contains defaultlist of usernames and passwords and compromised the machine to perform massive DDoS attacks. Radware Security proved that by implementing SSL channel, Mirai was able to cause an advanced DDoS attack by stressing the server machines that block SSL renegotiation and swiftlysetting upa new TCP connection on each SSL connection.

In the Q1 of 2017, we have seen the largest number of Command and Control servers spotted in South Korea [18]. The distribution of Botnetraised from 58% in the Q4 of 2016 to 66%. According to [19], [20], two major Botnet variants were discovered, namely Miner Botnet and Xor Botnet. Both variants used secured encrypted channels to carry out DDoS attacks. According to Statistica, over 28,448,673 DDoS attacks were detected, all of which used Botnet to perpetrate the attacks.

Similarly, in Q4 2018, analysts at SecureList detected new families of Botnet which included a modified version of Mirai. We also saw an increase in activities of Chalubo bot whose attack was first reported in late August. It uses a persistent technique as used by Xor. Likewise, in October, a new variant called Torii Botnet also surfaced. Its central targets were IoT devices. Its attack styles differ as it uses a high level of anonymity to communicate with the C&C server. Moreover, DemonBot was also caught in December, and it has abilities to integrate the cloud and boosts its attack strength, then attacks Big data servers via encrypted channels. Within a short period, it infected up to 1 million machines in one day [21], [22].

A massive development is seen in 2019 with the introduction of Cayosin. It was seen in Early March assembled from codes of Mirai, Qbot,and Swam. It is important to mention that it was advertised on Youtube and sold on Instagram. Botnetters globally took the opportunity to buy. Therefore, lots of script-kiddies used it to cause a lot of damage. itis speculated that the release of Cayosin is what led to the attack on Albany University, in the US in March. Moreover, 17 more attacks were launched against Universities servers. The Bot,likeothers, used HTTPs to flood the networks with bogus traffic. According to Kaspersky [23], Botnets C&C server still holds their stand in US, Netherland, Russian and South Korea. Hence, Botnet will continue to flourish for a long time.

The Botnet attacks discussed in this section demonstrate that Botnet evolves and uses existing technologies to remain recent in the network. Botnet utilizes SSL/TLS channel as one of their evasion techniques to avoid detection. Therefore, conventional Botnet like Eggdrop resurface and employ SSL/TLS for them to be resilient and relevant in the network. The examples also show that it is easy to perform attacks using Botnet. In future, there might be other Botnets that use SSL/TLS to hide their C&C communication and enable them to become stealthy in the network.

III. Botnet Taxonomy

The discussion on Botnet taxonomy in this section studies Botnetunder the classifications of; the Botnet life cycle, C&C, variants,and evasion techniques.

Life Cycle

"For an infected host to become an active bot and part of a Botnet, the host must go through a cycle of phases"[24].

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

There aresome models for Botnet life cycle available, for example,[25]–[29]. Some models are quite similar, some are concise,and some models are more complicated than others. Nonetheless, Figure 2 illustratesthe similar phases of the Botnet life cycle. Thus, this section summarizesthe Botnet life cycle for a comprehensive understanding of Botnet C&C operations.

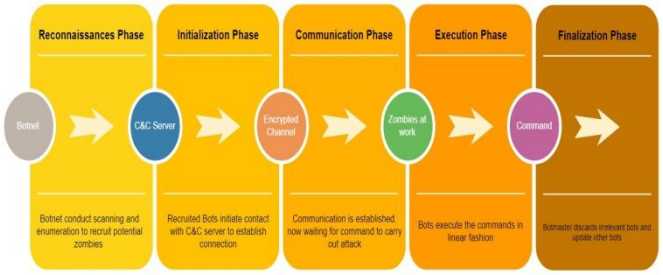

Figure 2. The life cycle of Botnet

Initially, Botnet scans for vulnerable hosts and recruit them. Scanning is a powerful active propagation method which is designed to locate and infect other hosts automatically with less human support[30]. Since it is the most successful propagation method, most Botnet uses the scanning method in looking for vulnerabilities to exploit. They then compromise the hosts and initiate communication with the Command and Control server while establishing the connection. This phase is also referred to as connection or rallying. After that, new bots listen to communication channels and wait for the commands, which they execute as given by the botmaster. Finally, the botmaster will discard the irrelevant bots and update other bots to evade detection. This cycle will continue until the Botnet is disassembled.

Bot life cycle comes in different ways, depending on the implementation structure. According to [27], [31]comprehending Botnet life cycle can increase the chances to detect and adequately respond to Botnet, there by developing a well-articulated strategy comprising of all different levels corresponding to each attack phase. Table 1 shows each phase in a Botnet life cycle and how we can build Botnet taxonomy from it. Therefore, understanding the Botnet life cycle helps in providing a solution for Botnet detection.

Table 1 Botnet life cycle

Botnet Life Cycle			
Phase			**Instance**
1. Botnet scanning and recruit			Distribution of malicious email
			P2P file-sharing network
			Software vulnerabilities
2. Rallying	C&C	A. Architecture	Centralized and decentralized
3. Listen and wait for commands		B. Communication Protocol	IRC, HTTP, P2P
		C. Communication Initiation	Push and Pull method
4. Execute commands		D. Communication Direction	Inbound and Bidirectional
5. Maintenance and upgrading			Discard irrelevant bots
			Update other bots

Command &Control Server

Botmaster controls bots using C&C, and this differentiates between Botnetand other malware. [30]stated that "without C&C communication, a Botnet is just an incoherent, random collection of infected machines." Bots link to the Command and Control server to receive the command from botmaster [29] and maintain the communication link with botmaster through that Command and Control server[32]. Therefore, C&C is a mechanism for bots to communicate with the botmaster. However, there are differences between centralized and decentralized C&C server, which will be discussed further in this section. Command and Controlserver is relaying the intentions of botmaster remotely to the bots. Thus, it is a challenge to locate the botmaster. The C&C is further discussed in term of the architecture, communication protocol, communication initiation, and communication direction.

Architecture

Generally, [27], [33]classify the Botnet C&C architecture as centralized and decentralized. Some researchers classify Botnet C&C architecture into centralized, decentralized, and hybrid C&C [34]–[38]identify three Botnet C&C topologies, whichare centralized, decentralized / Peer to Peer (P2P) and unstructured.

i. Centralized Command and Control Server

This structure is similar to a typicalclient-server model. "All bots establish their communication channel through one, ora few connection points, which are usually C&C servers responsible for sending commands to bots and for providing malware updates"[34]. Centralized Botnetenables

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

commands to be delivered, and replies received quickly from bots. Moreover, it is easy to control bots under centralized C&C. However, the C&C server itself is prone to failure. For example, if someone manages to control and remove the C&C server, the entireBotnet can be disassembled or become inoperable. The main protocols that employ centralized architecture are IRC and HTTP[39]. Figure 1 is an example of centralized Botnet architecture.

ii. Decentralized Command and Control Server

Dismantling Botnet with decentralized architecture is more challenging compared to centralized Botnet because there is no central Command and Control server to be found and deactivated. Some Botnets have multiple C&C servers. Thus, the detection of several bots does not certainlyaffect the wholeBotnet. This Botnetis generally based on variousPeer to Peer protocols and work as an overlay network. Figure 3denotes the decentralized Botnet.

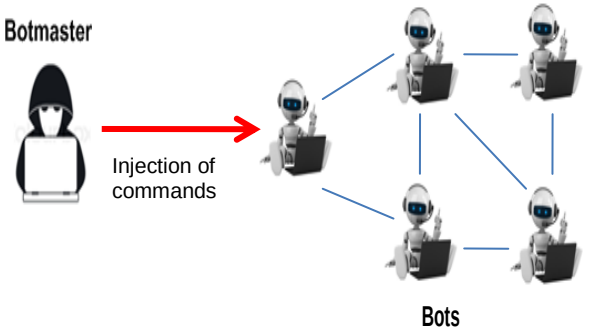

Figure 3. Decentralized Botnet

iii. Unstructured Command and Control Server

This Botnet architecture is more extreme than structured P2P Botnet. Bots encrypt their C&C channels [40] or use encrypted channels like SSL or TLS to evade detection. This architecture is based on the concept that no bot would see more than one other bot. As such, if a bot wants to send a message,it will encrypt it and arbitrarily scan the Internet and relay the message when it detects another bot.

Communication Protocols

The way Botnet communicates has changed accordingly with time and technology. Initially,Botnet has been created to help administrators manage IRC channels; ensure that the channel remains open, recognizes the channel operators, and give them control of the channels[41], [42]. Soon, however, cybercriminals realized the potential of Botnet and exploited it to create massive attacks. Furthermore, Botnet exploits P2P and Hypertext Transfer Protocol (HTTP) to launch the attacks. The Botnet also utilizes encrypted channels like SSL/TLS,where the initial purpose of this protocol is to securelegal communication.

i. IRC

The idea of Botnet originated from IRC. IRC is a text-based chat system that organizes communication via channels. The text-based protocol is easy to implement and customize. “This protocol remains as a significant technology for Botnet control and allowsfor centralized communication model”[33]. For instance, Eggdrop, one of the oldest IRC bot appeared in a later version where it uses SSL/TLS by default, allowing bots to establish SSL/TLS connection with IRC servers that support SSL/TLS.

An essential property of IRC is that technically there is no limitation to the number of possiblemembers within one channel, allowing many bots to be in thechannel and the ability for multicast communications. However, “IRC also enablesone-to-one conversations (private) that allows unicast communication”[43]. “Both multicast and unicast features allow the botmaster to have supple control of his Botnet, for example,selectinganexplicit group of bots to carry out an attack”[34].

Botmaster canmodify the protocol accordingly since there are numerous open-source applications for the IRC server. “Other advantages of IRC Botnet properties for the attacker are redundancy, scalability,and versatility, all of which allow for code reuse for bots and the re-creation of new bots”[44]. However, it is not difficult to detect and disturb the IRC Botnet process because IRC traffic is uncommon and rarely used in a corporate network. Network manager directlydetects, and blocks IRC traffics to prevent IRC Botnet activity. Furthermore, IRC is also inclined to single point of failure because of the centralized architecture itsupports. Therefore, IRC bots employ covert channels like SSL/TLS for their C&C to remain hidden and avoid detection.

ii. HTTP

Due to traffic constraints in IRC,whichwe discussed previously, HTTP became a simplified mechanism for implementing C&C communications[34]. Thisis because HTTP-based C&C communication is stealthier since web traffic is allowed in most networks [45], [46]and rarely blocked. Furthermore, it offers ease of development and deployment becausealmost all infected hosts are capable of communicating with HTTP C&C server [29].

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

HTTP is a popular and recognized standard used all over the Internet, and it isgenerally used for the conveyance of data (human-readable content such as websites and images) over the Internet. Hence HTTP is accessible by virtually every network device connected to the Internet and are rarely filtered by filtering protocols. This scenario has developed the interest for botmasters as it makes the protocol feasible via C&C protocol. The bots have to send a request to the C&C server in question periodically. According to [43], these requests usually comprise of a status report (the server decides to transfer, which commands to this particular bot). Despite the benefits it offers, like IRC, HTTP also suffers from a single point of failure because it primarily usesa centralized architecture.

iii. HTTPS

One of the recent Botnet evasion techniques is by leveraging an encrypted web base that uses SSL or TLS for their C&C communication (HTTPS).[47] states that Botnetis using SSL/TLS more frequently even though previously it used to be rare to see SSL/TLS used forthe C&C activity. It is because encryption is now ubiquitous and botmasters are relying on encryption more than ever to hide the implementation of an attack. The report provided by Desai also shows that new malicious payloads leveraging SSL/TLS for C&C activities have increased recently to about 60% for banking credential-stealing (Zbot, Vawtrak, Tickbot), 25% for ransomware families (for example Mirai), 12% for information stealer families and 3% areanotherBotnet. The most prevalent malware families leveraging SSL/TLS-based were Dridex and Emotet, the credential-stealing Botnet hadcontributed 34% of the total unique new payloads in 2017.

[48], [49] report that by 2017, more than 50% of Internet traffic had been protected by HTTPS.However, there are more than double the malicious content being delivered over SSL/TLS in the last six months. From this report, we can conclude that as the use of SSL/TLS encryption increases, so also the use of SSL/TLS for malicious purposes. The research by Blue Coat System (2016), a security solution provider owned by Symantec indicates that SSL/TLS will be increasingly implementedto hide attacks in the future.

iv. P2P

In P2P, the information about the order of participating partiesis distributed accordingly amongst the Botnetsthemselves. Consequently, knowledge about the entireBotnetchain cannot be directly received as commands have to be embedded into at least one member of the Botnetchain [43]. A Botnet with P2P communication protocol has leverage over IRC and HTTP protocol because P2P Botnet does not use centralized C&C and therefore, no single point of failure [44]. As such, it is much harder to suspend P2P Botnet. Furthermore, P2P Botnet is also hard to detect [35]. However, it is difficult to manage P2P Botnet because transferring command is slow without a centralized server [33].

Communication Initiation

[32] goes further to highlight at least 2severalmethods in which a bot can receive commands from the botmaster, namely: push mode and pull mode. Both methods require a remote C&C server as a medium for command dissemination. In the push mode, the C&C server ‘push’ the C&C message to the bots. Each bot receives the same message from the C&C server. In the pull mode, bots often send a request to the C&C server for the latest command within a stipulated time interval, very much like the browser’s request dispersed to the webserver. The C&C server will then respond if there are new commands available. Therefore, IRC is the protocol of choice for push mode, while HTTP is for pull mode.

Communication Direction

[27]“state that two techniques for communication initiation lead to two different models for communication direction. The two models are inbound and bidirectional communication. When the botmaster pushes commands to the bot, the communication channel could potentially be in-bound only. There is no need for the bot to send a message back to the botmaster or initiate communication by sending a request. When the bot initiates a check for commands, the communication channel used must be bidirectional. Having bidirectional communications is convenient to botmasters, as they can learn about the status of their bots.”

Variants

Botnet variants are primarily recognized by (i) Botnet C&C structure (for example centralized or decentralized), (ii) content (For example bots in similarBotnet cycles execute the same command), (iii) communication protocol (for example IRC, HTTP, P2P) and (iv) purpose (for example DDoS, spam, ransomware)[50]. Table 2 shows some popularBotnet variants that use SSL/TLS for their C&C communication. Some of the variants do not use SSL to cover their C&C at first, but later start to implement SSL, for example,the Zeus Botnet. The Botnet variants have been grouped according to their structure and communication protocol. Most of the Botnetvariants shown in Table 2 have centralized C&C and use encrypted C&C to send spam and steal banking credentials. Some Botnets which have been shut down previously re-emerge and

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

become stronger by applying encryption to mask their communications,forinstance, Kelihos which has been shut down by Microsoft and Kaspersky in 2011.

Table 2
Encrypted Botnet variants

Structure	Botnet (Year)	Family	Purpose	Protocol
Centralized	Eggdrop (1993)	-	flood ICMP, nuke and take over channels	IRC
	Torpig/ Sinowal/ Anserin	Mini bot (Specialized Botnet)	Steal bank credential info	HTTP or custom protocols to communicate
	Rustock (2006)	-	Send spam	HTTP
	Srizbi (2007)	-	Send spam	HTTP
	Kraken (2008)	Bobax	Send spam	HTTP
	Pushdo/ Cutwail	-	Send spam	HTTP
	WireX	-	DDoS attack	HTTP
	Shylocks	-	Intercept network traffic and inject code into banking websites.	HTTP
	Ramnit	-	Commit financial fraud	HTTP
	Stuxnet (2011)	-	Sabotage industrial process (infect nuclear plants for the enrichment of Uranium in Iran).	HTTP
	Duqu (2011)	Stuxnet	Steal information (hide in Kaspersky network).	HTTP
	Flame (2012)	Stuxnet	Cyber espionage	HTTP
	Mirai (2016)	IoT bot	DDoS attack	HTTP
Centralized & Decentralized	Social bot/ socbot Twitterbot	-	Bots that control social media accounts, performing phishing attacks, re-tweet storms, hashtag hijacks.	HTTP, P2P
	Koobface (2008)	Social bot	Target social network (for instance, Facebook, Skype,and Yahoo Messenger) and email (Gmail. Yahoo mail, AOL mail). Use for phishing.	HTTP, P2P
	Dridex/ Bugat/ Cridex	-	Steal bank credential info	HTTP, P2P
	Vawtrak/ Neverquest	-	Steal bank credential info	HTTP, P2P
Decentralized	Storm (2007)	-	Send spam	P2P
	Conficker (2008)	-	Send spam, identity theft, phishing exploits	P2P
	Kelihos/ Waledac/ Hlux (2010)	-	Send spam and theft of bitcoins	P2P
	Game Over Zeus (2011)	Zbot/ Kneber	Steal bank credential info	P2P
	Hajime (2016)	IoT bot - Mirai	The purpose still unclear and has a similar method of infection with Mirai, built massive P2P Botnet (almost 300 000 devices)	P2P

Botnet Evasion Technique

Evasion techniques have been designed to overcomethe detection mechanism,therebyallowing Botnet to have long operationperiods. Botnet continually addsa new mechanism to hide traces of communication. Some Botnetsare moving away from standard communication protocol like IRC and using a modified protocol like VPN or VoIP [51], and Botnet structure moved from centralized to decentralized [39]. Additionally, VoIP traffic has been used as a covert control channel. Botnet also employed IPv6 tunneling[52], and statistical patterns change, by utilizing dynamic Domain Name Service entries[33], fast-flux service network (FFSN)[53], tunneling throughICMP, HTTP or Voice over IP protocols

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

[54], randomizing bot communication patterns and passing different tasks to bots within the same infrastructure[39].

[55] view Botnet evasion strategies from the standpoint of the C&C server, botmaster, bot, and Command and Control communication channel. Evasion tactics by bots include binary obfuscation (to conceal bot binary), anti-analysis (refuse to run virtual machine or sandbox), security suppression (taking downthe running security software on the targeted machine) and rootkit technology (gain privileged access). Evasion tactics by botmaster are by using proxies where botmasters usuallycamouflet their realproperties by creating intermediate hosts (proxies) between the C&C server. Evasion methods by C&C server include IP flux (to evade IP based blacklisting and blocking) and rogue DNS servers (carrying out C&C covertly, effective redirection ofweb traffic to another malicioussite). Finally, C&C communication use encryption, protocol manipulation (for example, HTTP and IPv6 tunneling) and traffic manipulation (purposely create low traffic volume over a relativelysubstantialperiod). Figure 4 summarizesBotnet evasion strategies.

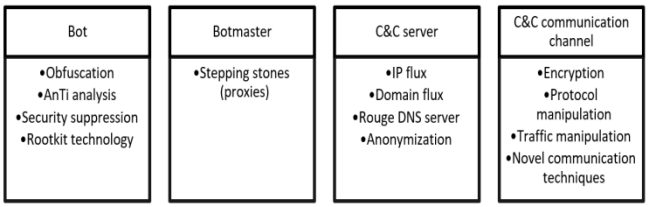

Figure 4.Botnetevasion strategies from C&C server, botmaster, bot,and C&C communication channel perspective

The Botnet also avoids detection by misusing the whitelist, evading protocol matcher, injecting malicious noisy packets, using very long response delay, random garbage injection in the packet, or employing random response delay [39]. Recently Botnet uses encryption such as SSL/TLS[56]especially over social media and covert communication protocols such as TCP and ICMP tunneling. The development of new avoidance techniques gives rise to the development of new detection techniques. Thus, there will be constant competition between attackers and defenders [34].

Botnetforms many variants mostly to evade detection. There are Botnet variants that use an encrypted channel like SSL/TLS to hide their activities, for example, GameOver Zeus. Encrypted Botnet requires extra effort for detection because of the additional layer of security to obfuscate their C&C. Therefore, appropriate techniques are needed to detect this kind of Botnet.

The observation of Botnet attacks in the recent years has shown similar pattern in Botnet evasion techniques [13], [15], [16], [47], [57]–[65]. This pattern is mostly based on the technologies available at the time. CurrentBotnets evade detection by using the technique as described below:

i. Manipulation of encrypted channels like SSL/TLS for their covert command and control communication,
ii. Use of social media and email for command and control (social bot/socbot) mostly to spread malware, spamming and gain credential info (it also happens that social media and email mostly deploy SSL/TLS to secure the communication), and
iii. Use of IoT devices for command and control, mostly to perform DDoS attack (DDoS of Things - DoT).

[13]reports that botmaster exploitsthe SSL blind spot to sneak past security control. Encryption is now ubiquitous,andmost legitimate sites support SSL/TLS. These have become significant factors for botmaster to currently employ SSL/TLS to hide their C&C. According to [47], C&C communication is an importantcomponent of Botnet attack. Therefore botmasters are relying on encryption more than ever to hide the implementation of attacks.

[66]states that there are several reasons botmaster are preferring to use social media and emails,such as Facebook Twitter and Gmail for their C&C. One of the reasons is thatthey have several users; Facebook alone has over a billion subscribers [65],and they use social networking sites and email on a routine basis. Since botmaster is targeting to compromise many hosts and turn them into bots to expand their C&C operation, these applications become an excellent platform for the botmaster. Another reason is that visiting social networking websites requires the use of HTTPsor HTTP connections,whichare not always blocked. This reason also related to the use of SSL/TLS, where social media and email application usually encrypt their communication. Botnet developers take advantage of SSL/TLS channel blind spot to place their C&C. Therefore, to detect Botnet activities, it is necessary to inspect SSL/TLS traffic as well.

According to Bernard Marr in [67],IoT devices have paved the way forthe massive growth of Botnet and its capability,whichis primarily used for DDoS attacks. It is because IoT allows botmaster touse a huge amount of Botnets of network-connected systems to overwhelm a number of websites or network resource with illicit requests.This often involves thousands of devices or systems with their unique IP addresses.Thus, it has almostbecome impossible to shutdown the attacksanddifferentiate between legitimate users and fake users. Mainly because of an enormous number of IoT devices, there is little or no built-in security in IoT devices due to the use of lightweight functions. Hence, ithasmade IoT to become easy targets for botmaster.[68]states that Botnet has turned IoT into IoV, Internet of Vulnerabilities. For example, in October

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

2016, Mirai partially took down an integral Internet infrastructure providerby using 100,000 unsecured IoT devices in the estimation, thus resulting in service disruption of websites receiving much traffic including Twitter and Netflix. To tackle this insecurity, IoT starts to deploy encrypted channel. However, this scenario mightgive a perfect hiding place to the attacker rather than securing the IoT.

Vectra Networks, a security company releases an article in 2016 which discusses the five ways cybercriminals conceal C&C communication. Cybercriminals use sophisticated avoidance techniques in an attempt to conceal their attack communications. The techniques they are using include:

i. Encryptionmethodswhich range from standard SSL/TLS to customized encryption schemes.
ii. Hidden tunnels in which communication is buried within multiple connections that use standard and commonly used protocols.For example, embedding hidden malware requests in regular HTTP traffic.
iii. Undercover in an allowed application where cybercriminals' preferred hiding place is within the vast amounts of Web traffic in a typical enterprise.
iv. External remote access (RAT) with which cybercriminals modify existing RATs or create their custom RATs to avoid detection; and appear as Virtual Network Computing (VNC), Remote Desktop Protocol (RDP), WebEx as well as other standard tools.
v. Anonymizing technologies such as The Onion Router (TOR), P2P networks and other proxies to obscure the location and identity.

The techniques reported by Vectra are similar to the three evasion techniques discussed previously. All these three evasion techniques are related to oneanother. A Botnet is not new, but the effects of encryption, a vast network of social media, emails,and IoT, have amplified the impact of Botnet attacks. Furthermore, anonymity services like TOR that utilize SSL/TLS also help to hide C&C servers.

Botnet Command and Control over the SSL/TLS

Botmaster hasa sturdy reason for turning to SSL/TLS. SSL/TLS provides a cohesive standard which is widely usedamongweb-based application.Hence, SSL/TLS guarantees robustness and reliability to Botnet infrastructures [56]. Furthermore, SSL/TLS implement encryption to ensure the security of the message sent through this channel. The Botnetuses these SSL/TLS features to hide within benign traffics.Therefore, it is tedious to identify or detectBotnet in the encrypted channel. Moreover, Botnet detection over SSL/TLS results in privacy issues. Thus, detection through payload analysis is challenging.

There are concerns about SSL/TLS presence in OSI [69] but mostly SSL is between the transport layer and application layer. The handshake protocol in SSL/TLS is used to perform authentication by utilizingan asymmetric key or public-key cryptography. However, Botnet intercepts this handshake by manipulating web browser's design flaw and lack of efficient verification system.

Another concern is that botmaster canobtain certificates either by stealing it or making fake certificates. Itis done by attacking the CA websites and database. In 2011, many organizations involved in issuing digital certificates had suffered from Botnetattacks. Some of the prominent organizations involved in these attacks included Comodo and DigiNotar.As soon as the private key relatedtothe trustedsystemis compromised, the malicious code can be assignedto the Botnet. Also, CAs are issuing improper certificates,and the certificates have been used for cyber-attacks. Companies like DigiCert unknowinglyleased a certificate to a fakeorganization that does not exist,and the certificates were assigned to Botnetsfor malicious activities.

[34]reports"that many attackers are using SSL to protect malicious traffic between C&C and infected machines",and this is mainly done by abusing the digital certificates. Stuxnet and Duqu in 2001 have been reported to have stolen the digital certificates.These Botnets digitally signed with the stolen certificates and appeared legitimate to conduct illicit activities. In 2016, Spymel usedthe stolen certificate to evade detection. The Spymel configuration data,includingC&C, is hardcoded within the Spymel executable file [70]. The digital certificates are used by botmasters to launch attacks through secure channels and trick unsuspected victims into believing they are on a legitimate website when their SSL/TLS traffic is being stealthilyaltered and hijacked. Therefore, Botnet was able to use SSL/TLS channel for the covert C&C communications.

The Botnet also can use SSL channel for their C&C communication by compromising the web server. A Botnet can manipulate the automatic open SSL channel to send itscommands and receive updates from its peersonce the web server has been compromised.Another method exploited by Botnetisa social media application. Botnet can stealthily distribute the C&C through social media posts, news feeds,and comments. This way,Botnet would camouflage within the benign SSL traffic. Figure 5 summarizes the four parties that Botnet can manipulate to go through SSL/TLS channel.

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

Figure 5. The four parties manipulated by Botnet

Having said that,Botnets either develop their illegal encrypted website through stolen certificate or other SSL/TLS-enabled applications for their C&C communication, for example, web server and social media site. By using a fake website, Botnet mainly lures the victims through phishing. Because of all the issues discussed above, botmasters can impersonate an encryptedwebsite, hide in it for a long time without being detected,and perform malicious activities through C&C communication over the SSL/TLS channel. The Botnet uses encrypted channel because the packet content (payload) sent over the SSL/TLS is encrypted and therefore increases the difficulties for detection.

IV. Botnet Detection

This section discusses Botnet detection and shows that Botnet evolves by using the existing technology of the time. Therefore Botnet detection techniques also evolveaccordingly [71], [72]. The detection techniques change as the Botnet changes communication technology and infrastructure. This trend can be observed by the papers published in Botnet detection. Between 2000 and 2010, researchers focus on finding detection solution for IRC-based and HTTP-based (centralized) Botnet[39], [73]–[79]. From 2010 and above the focus switched to decentralized architecture including P2P communication protocol[80]–[88]. Apparently, a lot of P2P Botnetdetection use machine learning-based detection. Aftera while, the focus of Botnet detectionis to find the solution for encrypted and covert Botnet[56], [89]–[95]. From 2015 onward, IoT bot became a hit,especiallyafter the Mirai attack [58], [61], [68]. Therefore, in this section, we study various Botnet detection techniques to detect different Botnet C&C communication protocols, and we also identify the detection techniques that are able to detect Botnet over the encrypted channel.

Over time, studies have proposed many architectures and a variety ofsolutions for Botnet detection [35]. [43]generally classifiesBotnet detection into passive and active techniques based on a research of experts working in the domain of Botnet mitigation. However, taxonomy dividing Botnet detection into Honeynet and IDS as the standard classifications for detection techniques[26], [34], [35], [82], [96]. Even though Honeynetswere not actively used in the specific bot detection system,[5], it is because Honeynet only collects the Botnet samples and needs to be integrated with other analysis tools such as antivirus and sandbox [1]. Furthermore, Honeynet is not effectivewhen it comes to detecting P2P and other decentralized Botnet[71]. Likewise, it is difficult for Honeynet to detect encrypted C&C. Furthermore, according to [26], Honeynet is mostly useful when performing analysis of Botnetsand their characteristics. Figure 6 depicts the taxonomy of Botnet detection techniques.

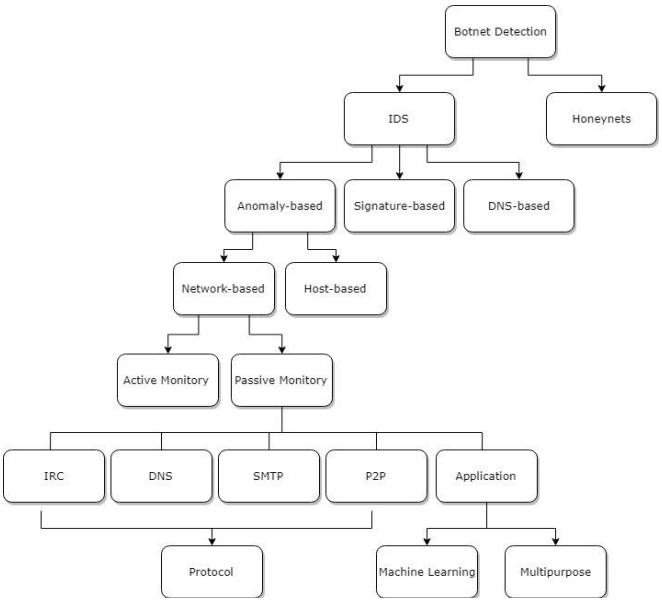

Figure 6. Basic classifications of Botnet detection techniques

Other than Honeynet, IDS has also been used for Botnet detection. Referring to Figure 6, IDS has beencategorizedbased onsignature-based, DNS based,andanomaly-based detection. Signature-basedBotnetdetectioncan detectBotnet immediately with a lowfalse positive rate, but only for known Botnet attacks. Unknown or new Botnet cannot be detectedusingsignature-based detection [96]. DNS based detection detect DNS traffic anomalies and Botnet DNS traffic [97]. This detection technique is the same as anomaly detection because of the fact that anomaly detection algorithms are appliedto DNS traffic [26].Furthermore, DNS based detection is only limited to Botnet that usesDNS and does not work on non-DNS-basedBotnet[98]. Therefore, researchers like [99], [100]have classified IDS into two broad categories,i.e.anomaly-based and signature-based detection. Anomaly-based detection consists of

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

host-based and network-based. Machine learning is part of anomaly-based detection approach under passive monitoring.

Anomaly-based detection identifiesBotnet without having any previous knowledge of signatures. Therefore, this detection technique effectively detects new Botnet attack compared to known attacks. Anomaly detection is built based on different traffic anomalies such as high traffic volumes, high network latency, traffic on unusual system behavior andunusual ports[52], that is enough proof to detect the presence of malicious bots in the network. For example, BotSniffer[39]is an anomaly-basedBotnet detection system designed to detect Botnet command and control traffic. Hence, this technique is suitable to detect encrypted Botnet C&C as encrypted traffic also produces anomaly,whichis used for detection. The combination ofanomaly and signature-baseddetection can overcome the limitation of the known attack in anomaly detection.For instance, BotHunter [39]integrated anomaly detection algorithm into Snort (signature-based detection).

Unlike integrated signature and anomaly techniques in BotHunter, [99]suggest an anomaly-based approach that needs no previous knowledge ofbot signatures, C&C server addresses,andBotnet C&C protocols. They clustered bots with the samenetflows and which carry the attacks in separate time windows toconducta correlation analysis to identify the bot-infected host. They built a prototype system and evaluated the prototype by usingreal-world traces,comprisingof normal traffic and numerousreal-worldBotnet traces. This approach produced high accuracy and low falsepositive rate.

There are other approaches employed by IDS-based Botnet detection,asindicatedin Figure 6. [101]present a host-based detection system for detecting and classifyingBotnet based on C&C communication (IRCHTTP, or P2P). This host-based system focuses on identifying bot on certain types ofhostthat typically make useof behavior or signature-based techniques to compare network traffic or system events with identified bot signatures or having similar behavioral information. Unlike many host-based IDS, this approach potentially discovers infections of previously unknown bots and producethe bestoutcomes in terms of false positive rate (0.078) and accuracy (0.929).

Active bot detection comprises ofpartaking in the Botnet operation, which involves spoofing a component of Botnet[30]. [102]It employs active Botnet detection by performing Botnet infiltration. They passively observe spam associatedwith commands and the information it distributes, and where necessary, actively changing specific elements of these messages in transit. Spamming activity observed on timecould help in detecting or blacklisting the Botnet.

Contrary to active detection, passive detection detects Botnet by covertly monitoring and studying their properties or attributes without making any mindful efforts to participate in the proceeding[30]. [103], [82]proposes a framework for Botnet detection based on traffic monitoring. This framework is based on finding familiar communication patterns and behaviors within the group of hosts that are doing at least one malicious activity.BotMiner[38], BotSniffer [45], BotDigger [101]exploit Botnet communication and behavior homogeneity by auditing the traffic behavior of the number of machines thereby identifying machines which are part of Botnet when they startto perform similar malicious activitiesinstantaneously.

Machine learning ispart of IDS-based detection. Machine learning approaches have been used extensively in analyzing various forms of network traffic data[105] and machine learning has more capability of handling new variants of Botnet compared to the conventional approach. It is because machine learning mainlyusesa knowledge-based approach focusing on pattern recognition, as Botnet produces distinct traffic patterns and behaviors[106]. These patterns,according to Stevanovic &Pedersen,can be proficiently detected with machine learning algorithms (MLA). Machine learning has proven to be powerful and accurate when it comes to detecting various forms of Botnet, including P2P Botnet[86], [107] and other decentralized Botnet and encrypted Botnet as well. Furthermore, it has been proven by previous researches that machine learning is the most suitable tool for detecting, predicting and preventing Botnet attacks[79], [86], [108]–[110]Table 3 summarizes various Botnet detection techniques.

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

Table 3
Botnet detection techniques

	Author	Detection Approach	Unknown Bot Detection	Protocol Structure Independent	Encrypted Bot Detection	Real-Time Detection	Low False Positive	Detection Rate (True Positive)	Features for Detection
Honeynet-based	[111]	virtual Honeynet (Nepenthes)	X	X (IRC & HTTP)	X	X	X	X	activity and message response command sequence from payload
	[112]	low interaction honeypot (Nepenthes)	X	X (IRC, DNS & HTTP)	X	X	X	X	source code
	[111]	virtual Honeynet (Nepenthes)	X	X (IRC, HTTP, P2P)	X	X	X	X	binaries
Signature-based	[113]	IDS-driven dialog correlation strategy	/	/	X	/	/	99.20%	calculated value of dialog correlation
	[114]	n-gram analysis	X	/	/	X	/	78%-100%	C&C protocol syntax
Anomaly-based	[73]	IRC mesh detection component with TCP scan detection heuristic	/	X (IRC)	X	X	X	X	TCP work weight
	[99]	X-means algorithm and hierarchical algorithm (ML approach)	/	/	/	/	/	3 out of 4 Botnets (100%)	features associated with NetFlows
	[115]	The Cooperative Adaptive Mechanism for NEtwork Protection (CAMNEP)	/	/	/	X	/	50% (accuracy)	features associated with NetFlows

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

DNS-based	[116]BotGAD	Generic metric model to measure group activities	/	/	/	/	/ (0.1)	X	Botnet group activity
	[109] - EXPOSURE	J48 decision tree algorithm	/	X (DNS)	X	X	/ (0.3%)	99.50%	15 features from 4 feature set; DNS answer-based, time-based, TTL value-based, domain name-based
Mining-based / Machine Learning-based	[83]	1.Nearest Neighbors Classifier, 2.Linear SVM 3.ANN, 4.Gaussian Based Classifier, 5.Naive Bayes Classifier	/	X (P2P)	/	X	X (1.NN-7%, 2.SVM-6% 3.ANN-8%, 4.GBC-20%, 5.NBC-12%)	1.NN-92%, 2.SVM-97.8% 3.ANN-94.5%, 4.GBC-96.2%, 5.NBC-89.7%	17 features (flow based and host based features)
	[117]	1. Adaboost, 2.Conjunctive rule, 3.J48, 4.Naïve Bayes, 5.Ripper, 6.SVM	/	X (HTTP, HTTPS)	/	X	X (Ada-5.5%, CR-5.1%, J48-3.2%,NB-44.3%, Ripper-2.9%, SVM-2.4%)	Ada-94.9%, CR-94.9%, J48-96.1%,NB-97.3%, Ripper-95.9%, SVM-96%	TLS features
	[114] - CoCoSpot	Average-Linkage Hirarchical Clustering	/	/	/	X	88% of Botnet families have less than 0.1%	>50% of Botnet families are over 95.6%	traffic features (carrier protocol distinction, message length sequences, encoding differences)
	[86]	Neural networks with Bayesian regularization	/	X (P2P)	/	/	X	99.2% (accuracy)	15 features selected using Information Gain Ranking Algorithm out of 44

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

								features extracted
	[118]	Decision Tree classifier + Reduced Error Pruning Algorithm (REPTree)	/	/	/	/	/ (0.01%)	98.30%	traffic flow behaviour analysis in a small time windows
	[106]	Adaptive Neuro Fuzzy Inference System (ANFIS)	/	X (DNS)	/	X	X	95.29%	DNS query
	[56]	Decision Tree Algorithms	/	/	/	X	/ (0)	99.96%	6 SSL features identified, only use 4 of them for detection
	[89]	Naïve Bayes	/	/	/	X	X	98.84% (accuracy)	flow duration, flow size, number of packets in flow, protocol, source IP
	[110]	Nearest Neighbour, Decision Tree, SVM	/	X (P2P)	/	X	X	NN-97.10%, DT-100%, SVM-100% (accuracy)	SrcIP, SrcPort, DstIP, DstPort, Protocol, Total Packets, Total bytes and Duration
	[119]	1. DecisionTree J48 Classifier without Pruning Algorithm 2. Naïve Bayes	/	/	/	X	X (Decision Tree-13.3047 %, NB-21.4592 %)	Decision Tree - 86.6953 %, NB-78.5408 %	traffic flow characteristics based on time intervals
	[120]	Boosted Decision Tree (AdaBoost + J48), Naïve Bayes,	/	/	/	/	/ (Ada+J48-0.0813%, NB-0.0481%, SVM-	Ada+J48-95.86%, NB-99.14%, SVM-92.02% (accuracy	Small_Packets, Packet_ratio, Initial Packet_length, Bot-response_pac

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

		SVM					0.0972%)	)	ket ratio
	[121]	SVM	/	/	/	X	X (15.1%)	100%	entropy

Even though most of the MLAs in the above table claim that they can detect Botnet over the encrypted channel, most of them are structure independent or HTTPS-based detection, the test that they conducted does not use appropriate datasets that include encrypted traffics. Furthermore, theresults provided do not include the detection of Botnet over the encrypted channel.

Detecting Botnet over the Encrypted Channel

MostBotnet detection relies on the detection features. [93],in their experiments, proved that the accuracy of Botnet detection highly depends on the features extracted. There are different features used for detection,as shown in Table 3. The features use for detecting Botnetover the encrypted channel might slightly differ from common Botnet; for example,Botnet that uses encrypted channel produces high entropy. There was also the algorithm available to measure the weight and reliability of the features used in detection.

Table 3 also shows the detection metrics used for detection by previous researchers. Accuracy, false positive rate,anddetection rateare the standard metrics used to evaluate the performance of their detection approaches. These metrics are vital to prove that the proposed approaches are relevant. Therefore, it is very crucial to achieve high accuracy and detection rates while achieving a lowfalse positive rate to indicate that the approaches areeffective. That is the issue faced by several approaches discussed in this section where they can provide a high detection rate, but at the same time, the false positive rate is also very high. In detection, real-time,or fast detection is also essential to measure the efficiency of the approaches.

Detecting Botnet is a challenge due to its dynamic nature but detecting Botnetoverthe encrypted channel is even more challenging. Previously, detecting this kind of Botnetrelies on payload analysis,which requires decryption,and this leads to privacy issues.Even though [122]propose PROVEX to improve payload-based detection, their approach faced privacy issue by decrypting and inspecting the payload.Instead of payload analysis, there were also other approaches available for detecting Botnetoverthe encrypted channel. For example, flow analysis and machine learning approach which have become prominent. Even the approachesdiscussed below providegood results, there are limitations, for instance, some of them are dependable onspecificBotnet structure, for example,[83], [86] focusedon detectingBotnet that uses P2P protocol. It limits the type of Botnet detected and causes invariants signature. Another limitation discussed in this section is the insufficient alarm mechanism.

There are other approaches capable of detectingBotnetoverthe encrypted channel. [123]introduce a unified framework for detecting the known bots and encrypted bots using signature-based classifier and anomaly-based detection. Here, Bots are detected based on the pattern of network flow. This technique is one of the examples of flow-based encrypted Botnet detection,eventhough this approach has certain flawswhen it comes to detection time efficiency.

Machine Learning in Encrypted Botnet Detection

As discussed previously, machine learning is a promising approachfor detectingBotnet in the encrypted channel, and it solves the limitation of existing approaches in detecting encrypted Botnet. Machine learning can select relevant features that are good for detection and learn from them[15]. Learning in this context meansthe ability to recognize intricatestructures and patterns,then make rational decisions based on known or existing data. According to [105], Machine learning algorithms are categorizedas supervised and unsupervised based on the expectedresult of the algorithm. [124]state that supervised learning is a class of well-structured MLA that generates a function that feedsthe input to the desired output. While [125],on the other hand, state that unsupervised learning is a type of MLA where training data comprisesa set of input without any target output values.

a. Supervised Learning

[126]describe supervised learning as a learning techniquethat performs predictions based on a set of examples.

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

Each instance used for training is labeled with the value of interest. Hence,the learning algorithm identifiesunique features in those value labels,and each training model looks for different types of features. [127] generally categorize supervised learning into classification and regression. Classification, on one hand, means predicting the class as one of a finite number of distinctlabels. On the other hand, Regression uses algorithms to predict the output value as one of a possibly infinite set of real-valued points. Neural Network, Expert System, Support Vector Machine, Dendric Cell Algorithm, K-Nearest Neighbors,and Genetic Algorithm are the example of the supervised learning algorithm.

b. Unsupervised Learning

Unsupervised learning takes a different turn here because data points do not possess labels that identify them in the class. [128],whendescribing unsupervised learning,says,"unsupervised learning means teaching machines to learn for themselves without having to be explicitly told what is right or wrong". Which means learning is independent and self-contained. Unsupervised learning is categorized into clustering and association. Self-Organizing Map and K-Means algorithm are examples of the unsupervised learning algorithm.

Machine learning has been extensively used in detecting various kinds of Botnet. [105] states that machine learning is one of the contemporary advances in network-basedBotnet detection for identifying patterns of malicious traffic. There are quite some researches done in this area discussing different MLA; either supervised, unsupervised,orsemi-supervised.

Some researchers performed comparative studies on different machine learning to show which machine learning provides the best performance. [83]propose a studyrelated to the ability of 5 commonly used MLAs to satisfynetworkBotnet detection requirements, known as novelty detection, adaptability,and early detection. All fiveMLAs provide high true positive value. However,the Support Vector Machine got the highest true positive value,which is 97.8%. [117]compares different techniques and based on the results; he proposes three novel techniques for detecting HTTPS and HTTP-based C&C channels. It shows Naïve Bayes got the highest true positive,which is 97.3%. [110]compares different supervised MLAs for determining peer to peer Botnet detection accuracy. Decision Tree and Support Vector Machine achieved 100% accuracy.

[129]put forth a detection techniquethat has been used topredict bot hosts from the benign host by studying traffic flow activities based on time series instead of payload inspection. They use the Decision Tree and Naïve Bayes for classification. Classification with decision tree gave better true positive of 86.69%. [120] propose an approach to detect Botnet irrespective of their structures. They try several MLAs to their approach,and Naïve Bayes has the highest detection rate of 99.14%.

[56], [118]implement Decision Tree to their approaches,and both provide very high detection rates which are 98.5 % and 99.96% with very lowfalse positive rates of 0.01 % and 0%.[130]develop CoCoSpot use Average-Linking Hierarchical Clustering. 50% ofBotnet families were detected at the rate of 95.6%. [89], [131]use Naïve Bayes and achieved 98.84% accuracy. Apparently, most of the MLAs discussed have very high detection rates.

Even though some techniques provide high detection rates, comparatively they also havea highfalse positive rate. For example,[121], [132]proposedan approach using Support Vector Machine and got a 100% detection rate. However, the false positive rate is more than 15%. The work by [88] also provided very high false positive,of more than 21%. Above all, [117] had the highest false positive value of 44.3% by using Naïve Bayes.

[79] presents a host-based behavioral technique for detecting Botnetby comparing different activities generated by bots through monitoring function calls within a given time series. Al-Hammadi used Dendric Cell Algorithm inspired by the Immune System. The evaluated results show that by analyzing different events generated by IRC/P2P bots within a giventime, high detection accuracywas achieved. However, using an intelligent correlation algorithm will reveal automatically whether or not an anomaly is detected while at the same time, revealing the location.

Among the most promising algorithm for Botnet detection is Neural Networks and one of the sub-categories of NN is Self-organizing Map. SOM hasbeen widely usedfor intrusion detection. Unfortunately, there are limitedworksthat discuss SOM for Botnet detection.However, SOM is a promising approach,especially for developing autonomous Botnet detection system. [80] used SOM toclassify and cluster peer to peer Botnet traffic and other suspicious network activity by analyzing firewall log entries. [100]adaptsHexagonal Self Organizing Map in an effort tocluster and predict unknown firewall log,which was later used to detectunknown bots in the network.

[86]propose and implemented a hybrid framework for detecting P2PBotnet in an ongoing network flow by juxtaposing Bayesian Regularization with Neural Networks todetectknown and newly discoveredBotnet. The statistical tests revealthat the trained Neural Network and Bayesian

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

Regularization model can generalize very well and can predict unknown bots' malicious activity. Thus,Botnet detection activities are successfully achieved with an accuracy of 99.2%. [133]extended the framework proposed by Salvador and developeda new system known as BoNeSSy. Nogueira develops a Botnet detection system that works by collecting statistics from network flow using Neural Network. The results obtained shows that it is practically possible to have anefficientand feasible system that can display high detection rates with less computational resources.

Many current approaches to the process of detecting intrusions utilize some forms of rule-based analysis. Expert System is the most common type of rule-based intrusion detection approach. Most concurrent behavior-based techniquesfail topredict or detectBotnet because they keep changing their structure and pattern to avoid detection. [106]presentsanAdaptiveNeuro-Fuzzy Inference System (ANFIS), a technique which trains the system for future prediction. However, the limitation of this work is the restriction of fuzzy rules and fuzzy sets for the comparison purpose. Therefore, the proposed work should be able to overcome the limitations by increasing the number of rules generated using the Botnet features and information gain.

Fuzzy pattern recognition proposed by [78]shows very promising results as it can detect both IP address and domain names of the bot by auditing the network channel. The algorithm developed involves traffic reduction, feature selection,and pattern recognition. Fuzziness in pattern recognition helps to detect bots which are hidden or camouflage. Results from performance evaluation reveal that the system has successfully reduced up to about 70% of raw packet input and has achieved a detection rate as high as 95% and low false positive rate of at least 0-3.08%. On the one hand, FPRFalgorithm has significantly identified inactive Botnet that shows a potential vulnerability on the host. Likewise, BotDigger proposed by [104]uses fuzzy logic as a tool for defining logical rules that are solely based on statistical facts retrieved fromessential features that recognizeBotnet activities. One majorgoal that was achieved by implementing this architecture is that it provides a platform for integrating a wide range of traffic specifications.

The discussion of the machine learning above mostly provides performance evaluation using the detection rate, accuracy,or false positive value. However, there are other vital metrics which are real-time and autonomous. Without the detection able to detect accurately, it is useless without fast detection or real-time detection. Researchers focus on developing real-timeBotnet detection system, for example,[78], [133], [134].

Autonomous mainly focuses on self-learning and self-managerial properties. [26] proposes the new autonomous model for Botnet detection using K-means algorithm, one of the most straightforward algorithms that analyze clustering problem. According to [30], the level of automation can be classified as semi-automated,manual,and automated. Semi-automated Botnet detection requires minimum human interaction,andin most cases, the detection is conducted automatically. Alternatively, fully automated Botnet detection works without human interactionduring initial development. Khattak also agreed that ideally, any detection method should be as generic and automated as possible.

V. SUMMARY

We have seen many Botnet attacks over the years and recognize the high impact of the attacks on our lives,such as money loss, data loss,disruption of services,and many more. Based on this study, it is clear that Botnet attacks will continue to grow exponentially over time. More variants of Botnet will evolve due to availability of the crime-as-a-service model, which allows both skilled and unskilled individuals to purchase Botnet codes and recreate new variants. According to Kaspersky Quarterly report 2019[23], [135], Botnet attacks will increase from 700 attacks per day to 7000 attacks per day by 2025. With the growth of IoT devices, the number is expected to double. So far, machine learning, especially neural networks, has proven to be the most promising technique for detecting and predicting Botnet attacks over an encrypted channel. In the future, we hope that more research will be done using powerful algorithms such as convolutional neural networks, which are yet to be appliedonBotnet detection. In summary, this paper contains a comprehensive survey on structured taxonomy of Botnet. This knowledge is essentialfor understandingBotnet technology and to aid in findingsolutions for Botnetmitigation. Likewise, the study provides well-established literature for existing Botnet detection approaches, Botnet life cycle, Botnet C&C, and evasion techniques. Another section of this study provides a detailed overviewof Botnet detection over an encrypted channel. Towards the end, this paper talks extensively on the application of machine learning being the most effective approach for mitigating Botnet attacks.

During the course of this study, some research gaps were identified which are very necessary to address. It is self-evident that there are no much available datasets sufficient enough to train a classifier to predict Botnet attacks over an encrypted channel. The existing features are not powerful enough to produce a good model. Hence, the need for building different datasets for different types of Botnet variants is paramount. Our study also identified that some variants of Botnet would take advantage of Windows PowerShell to be

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

used as bots in the next generation of Botnet attacks. These days, we have seen an increase in the use of Windows PowerShell application in carrying out RAT attacks with ransomware or trojan horse. Hence, more studies should be conducted to accommodate such misuse of trusted windows applications for perpetrating Botnet attacks. Another significant gap is to address the issue of lightweight architecture for IoT devices with delicate security infrastructure. According to [45], there will be over 25 billion IoT devices by 2020. Hence, researchers should work towards finding solutions that can work hand in hand with lightweight architecture to protect Botnet from exploiting such devices. Otherwise, we will not only be talking about the loss of data and money, but also we will be talking about death as a result of one device or another being hijacked by Botnet attack especially hospital appliances such as MRI machines, implants and chipsets inside the human body (such as heart monitors, ear devices for hearing enhancement).

VI. ACKNOWLEDGEMENT

This studyhas been sponsored by the Universiti Sains Malaysia through Research University (RUI) Grant, titled "Enhancing Botnet Detection Efficiency and Accuracy using Machine Learning Techniques" (account number 1001/ PKOMP/ 8014017).

VII. REFERENCES

[1] A. Jakalan, J. Barazi, and W. Xiaowei, "Botnet Detection Techniques," *Int. J. Comput. Sci. Commun. Secur.*, 2014.

[2] G. Khalil, "InfoSec Reading Room," 2014.

[3] S. M. Hasan, "Detection of P2P Botnets Based on Support Vector Machine: Case Study," vol. 32, no. 5, pp. 1227–1239, 2014.

[4] G. Kirubavathi Venkatesh and R. Anitha Nadarajan, "HTTP botnet detection using adaptive learning rate multilayer feed-forward neural network," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7322 LNCS, pp. 38–48.

[5] F. Haltaş, E. Uzun, N. Şişeci, A. Poşul, and B. Emre, "An automated bot detection system through honeypots for large-scale," in *International Conference On Cyber Conflict (CyCon 2014)*, 2014.

[6] E. Cooke, F. Jahanian, and D. McPherson, "The Zombie Roundup: Understanding, Detecting and Disrupting Botnets.," *SRUTI*, vol. 5, pp. 6–6, 2005.

[7] N. Provos and T. Holz., "Virtual honeypots: from botnet tracking to intrusion detection," in *Proceedings of The 5th Australian Digital Forensics Conference*, 2007.

[8] C. Şahin, "The predictive level of social media addiction for life satisfaction: A study on university students," *Turkish Online J. Educ. Technol.*, vol. 16, no. 4, pp. 120–125, 2017.

[9] Margaret Rouse, "What is Microsoft SQL Server? - Definition from WhatIs.com," *TechTarget*, 2014. [Online]. Available: https://searchsqlserver.techtarget.com/definition/SQL-Server. [Accessed: 05-Apr-2018].

[10] Margaret Rouse, "Data Collection," *TechTarget*, 2017. [Online]. Available: https://searchcio.techtarget.com/definition/data-collection. [Accessed: 27-Apr-2018].

[11] M. Weaver, "After US, cyber attackers target South Korea | World news |," *The Guardian*, 2009. [Online]. Available: https://www.theguardian.com/world/2009/jul/08/south-korea-cyber-attack. [Accessed: 12-Dec-2019].

[12] B. Sterling, "Microsoft Versus Rustock Botnet | WIRED," *WIRED*, 2011. [Online]. Available: https://www.wired.com/2011/03/microsoft-versus-rustock-Botnet/. [Accessed: 12-Dec-2019].

[13] P. Nicholson, "What Lies Beneath: Advanced Attacks that Hide in SSL Traffic," *A10*, 2015. [Online]. Available: https://www.a10networks.com/blog/what-lies-beneath-advanced-attacks-hide-ssl-traffic/. [Accessed: 10-Dec-2019].

[14] H. Security, "Recommended Practice: Defense in Depth," US.

[15] R. Mitchell, "Mirai: The Program That Makes IoT Botnet Zombies - News," *All about Circuits*, 2017. [Online]. Available: https://www.allaboutcircuits.com/news/mirai-the-program-that-makes-iot-Botnet-zombies/. [Accessed: 10-Dec-2019].

[16] N. Newman, "Mainstream media and the distribution of news in the age of social discovery," *Reuters Inst. Study Journal.*, no. September, p. 58, 2011.

[17] Isadore Newman and Carolyn R. Benz, *Qualitative-quantitative Research Methodology: Exploring the Interactive ... - Isadore Newman, Carolyn R. Benz - Google Books*. 2007.

[18] SecureList, "Ransomware Incident Response – Indicator of Compromise," *The Neek*, Aug-2017.

[19] R. Unuchek, F. Sinitsyn, D. Parinov, and A. Liskin, "IT threat evolution Q3 2017. Statistics," 2017.

[20] R. Unuchek, F. Sinitsyn, D. Parinov, and A. Liskin, "IT threat evolution Q2 2017. Statistics," 2017.

[21] Symantec, "An Internet Security Threat Report Special Report," 2017.

[22] Sophos, "SophosLabs 2018 Malware Forecast," 2018.

[23] Kaspersky, "Kaspersky Security Bulletin: Kaspersky Lab Threat Predictions For 2018," 2018.

[24] S. S. C. Silva, R. M. P. Silva, R. C. G. Pinto, and R. M.

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

Salles, "Botnets: A survey," *Comput. Networks*, vol. 57, pp. 378–403, 2013.

[25] C. Schiller, J. Binkley, D. Harley, G. Evron, and T. Bradley, "Botnets: The killer web app. Rockland, MA," 2007.

[26] M. Feily, A. Shahrestani, and S. Ramadass, "A survey of botnet and botnet detection," in *Proceedings - 2009 3rd International Conference on Emerging Security Information, Systems and Technologies, SECURWARE 2009*, 2009, pp. 268–273.

[27] Y. Ben Mustapha, G. Gonzalez, G. Atos, N. Hachem, G. G. Granadillo, and H. Debar, "Botnets: Lifecycle and Taxonomy," in *Conference on Network and Information Systems Security. IEEE, 2011.*, 2011.

[28] M. Eslahi, R. Salleh, and N. B. Anuar, "Bots and botnets: An overview of characteristics, detection and challenges," in *2012 IEEE International Conference on Control System, Computing and Engineering.*, 2012.

[29] R. A. Rodriguez-Gomez, G. Macia-Fernandez, and P. Garcia-Teodoro, "Survey and taxonomy of botnet research through life-cycle," *ACM Comput. Surv.*, vol. 45, no. 4, Aug. 2013.

[30] S. Khattak, N. R. Ramay, K. R. Khan, A. A. Syed, and S. A. Khayam, "A Taxonomy of Botnet Behavior ," vol. 16, no. 2, pp. 898–924, 2014.

[31] T. Teixeira, S. Hachem, V. Issarny, and N. Georgantas, "Service oriented middleware for the internet of things: A perspective (invited paper)," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6994 LNCS. pp. 220–229, 2011.

[32] A. Kak, "Lecture 29: Bots and Botnets Lecture Notes on " Computer and Network Security "," pp. 1–52, 2015.

[33] A. Tesfahun and D. L. Bhaskari, "Botnet Detection and Counter measures: A Survey," *Int. J. Emerg. Trends Technol. Comput. Sci.*, vol. 2, no. 4, 2013.

[34] S. S. C. Silva, R. M. P. Silva, R. C. G. Pinto, and R. M. Salles, "Botnets: A survey," *Comput. Networks*, vol. 57, no. 2, pp. 378–403, Feb. 2013.

[35] A. Karim *et al.*, "Botnet detection techniques: review, future trends, and issues," *J. Zhejiang Univ. Sci. C*, vol. 15, no. 11, pp. 943–983, 2014.

[36] M. Bailey, E. Cooke, F. Jahanian, Y. Xu, A. Arbor, and A. Arbor, "A Survey of Botnet Technology and Defenses," 2006.

[37] Y. Zeng, X. Hu, and K. G. Shin, "How to Construct a Mobile Botnet?," in *The 40th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2010)*, 2010.

[38] J. A. Jupin, T. Sutikno, M. A. Ismail, M. S. Mohamad, and S. Kasim, "Review of the machine learning methods in the classification of phishing attack," vol. 8, no. 4, pp. 1545–1555, 2019.

[39] G. Gu, R. Perdisci, J. Zhang, and W. Lee, "Botminer: Clustering analysis of network traffic for protocol-and structure-independent botnet detection," *USENIX Assoc.*, vol. 139, 2008.

[40] M. Bailey, E. Cooke, F. Jahanian, Y. Xu, and M. Karir, "A Survey of Botnet Technology and Defenses - IEEE Conference Publication," *IEEE*, 2009. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4804459. [Accessed: 29-Nov-2019].

[41] FBI, "Criminals Continue To Defraud And Extort Funds From Victims Using Cryptowall Ransomware Schemes," 2015.

[42] Z. Bu, P. Bueno, R. Kashyap, and A. Wosotowsky, "The New Era of Botnets," *White Pap. from McAfee*, 2010.

[43] D. Plohmann, E. Gerhards-Padilla, and F. Leder, "Botnets: measurement, detection, disinfection and defence," *ENISA Work.*, 2011.

[44] J. B. Grizzard, V. Sharma, C. Nunnery, B. Byung, H. Kang, and D. Dagon, "Peer-to-Peer Botnets: Overview and Case Study," *Comput. Based Learn. Unit, Univ. Leeds.*, vol. 33, no. 2, 2002.

[45] Z. Bu, P. Bueno, R. Kashyap, A. Wosotowsky, and M. Labs, "The New Era of Botnets White Paper," 2018.

[46] G. Gu, J. Zhang, and W. Lee, "BotSniffer: Detecting botnet command and control channels in network traffic," in *15th Annual Network and Distributed System Security Symposium*, 2008.

[47] D. Desai, "SSL/TLS-based Malware Attacks," *ZSCaler Blog*, 2017. [Online]. Available: https://www.zscaler.com/blogs/research/ssltls-based-malware-attacks. [Accessed: 29-Nov-2019].

[48] K. Finley, "Half the Web Is Now Encrypted. That Makes Everyone Safer," *WIRED*, 2017. [Online]. Available: https://www.wired.com/2017/01/half-web-now-encrypted-makes-everyone-safer/. [Accessed: 29-Nov-2019].

[49] G. Gebhart, "We're Halfway to Encrypting the Entire Web | Electronic Frontier Foundation," *Electronic Frontier Foundation*, 2017. [Online]. Available: https://www.eff.org/deeplinks/2017/02/were-halfway-encrypting-entire-web. [Accessed: 29-Nov-2019].

[50] G. Gu, J. Zhang, and W. Lee, "BotSniffer : Detecting Botnet Command and Control Channels in Network Traffic," *Proc. 15th Annu. Netw. Distrib. Syst. Secur. Symp.*, vol. 53, no. 1, pp. 1–13, 2008.

[51] TrendMicro, "What You Need to Know About the LockerGoga Ransomware - Security News - Trend Micro USA," 2019. [Online]. Available: https://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/what-you-need-to-know-about-the-lockergoga-

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

ransomware. [Accessed: 11-Jul-2019].

[52] B. Saha and A. Gairola, "Botnet: An Overivew. CERT-In White Paper," 2005.

[53] T. Holz, C. Gorecki, F. Freiling, and K. Rieck, "Detection and mitigation of fast-flux service networks," in *Proceedings of the 15th Annual Network and Distributed System Security Symposium (NDSS)*, 2008.

[54] A. Nappa, A. Fattori, M. Balduzzi, M. Dell'Amico, and L. Cavallaro, "Take a deep breath: A stealthy, resilient and cost-effective botnet using skype," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010, vol. 6201 LNCS, pp. 81–100.

[55] 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. Surv. tutorials*, vol. 16, no. 2, pp. 898–924, 2013.

[56] R. Bortolameotti, "C&C Botnet Detection over SSL," *Electrical Engineering, Mathematics and Computer Science*. 2014.

[57] D. Gooley, "The Rise in SSL-based Threats," *Zscaler Blog*, 2017. [Online]. Available: https://www.zscaler.com/blogs/research/rise-ssl-based-threats. [Accessed: 29-Nov-2019].

[58] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and other botnets," *Computer (Long. Beach. Calif).*, vol. 50, no. 7, 2017.

[59] M. H. H. Ichsan, W. Kurniawan, and S. R. Akbar, "UDP Pervasive Protocol Integration with IoT for Smart Home Environment using LabVIEW," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 6, p. 5342, Dec. 2018.

[60] G. Mayfield, "How to deal with the blind spots in your security created by SSL encrypted traffic," *Network World*, 2015. [Online]. Available: https://www.networkworld.com/article/3005646/how-to-deal-with-the-blind-spots-in-your-security-created-by-ssl-encrypted-traffic.html. [Accessed: 10-Dec-2019].

[61] J. Moon, J. J. Jang, and I. Y. Jung, "Bot Detection via IoT Environment," in *2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems*, 2015, pp. 1691–1692.

[62] C. Cimpanu, "Remaiten Is a New DDoS Bot Targeting Linux-Based Home Routers," *Softpedia News*, 2016. [Online]. Available: https://news.softpedia.com/news/remaiten-is-a-new-ddos-bot-targeting-linux-based-home-routers-502434.shtml. [Accessed: 29-Nov-2019].

[63] D. Goodin, "New, more-powerful IoT botnet infects 3,500 devices in 5 days," *ARS Technica*, 2016. [Online]. Available: https://arstechnica.com/information-technology/2016/11/new-iot-Botnet-that-borrows-from-notorious-mirai-infects-3500-devices/. [Accessed: 29-Nov-2019].

[64] L. Larinkoski and C. Leita Petros Elia, "Detecting Encrypted Command & Control Channels with Network Fingerprints," Aalto University, 2016.

[65] P. Muncaster, "Anti-worm 'Nematode' Could be Answer to Mirai Botnets - Infosecurity Magazine," *InfoSecurity Group*, 2019. [Online]. Available: https://www.infosecurity-magazine.com/news/anti-worm-nematode-could-be-answer/. [Accessed: 10-Dec-2019].

[66] Avira, "Drive-by download," *Avira Security Wordbook*, 2019. [Online]. Available: https://www.avira.com/en/security-term/t/drive-by-download/id/13. [Accessed: 10-Apr-2019].

[67] L. Columbus, "2017 Roundup Of Internet Of Things Forecasts," *Forbes*, 2017.

[68] K. Angrishi, "Turning Internet of Things(IoT) into Internet of Vulnerabilities (IoV) : IoT Botnets," *arXiv*, vol. 1, no. feb 2017, Feb. 2017.

[69] M. Gregg, "OSI: Securing the Stack, Layer 6 - Encryption," *TechTarget*, 2018. [Online]. Available: https://searchnetworking.techtarget.com/tip/OSI-Securing-the-Stack-Layer-6-Encryption. [Accessed: 29-Nov-2019].

[70] E. Kovacs, "'Spymel' Trojan Uses Stolen Certificates to Evade Detection," *Security Week Network*, 2016. [Online]. Available: https://www.securityweek.com/spymel-trojan-uses-stolen-certificates-evade-detection. [Accessed: 10-Dec-2019].

[71] T. Hyslip and J. Pittman, "A Survey of Botnet Detection Techniques by Command and Control Infrastructure," *J. Digit. Forensics, Secur. Law*, vol. 10, no. 1, pp. 7–26, 2015.

[72] K. Luechaphonthara and V. A, "IOT based application for monitoring electricity power consumption in home appliances," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 6, p. 4988, Dec. 2019.

[73] J. R. Binkley and S. Singh, "An Algorithm for Anomaly-based Botnet Detection," in *The Advanced Computing Systems Association*, 2006.

[74] W. T. Strayer, D. Lapsely, R. Walsh, and C. Livadas, "Botnet Detection Based on Network Behavior," *Springer*, vol. 3, no. 4, pp. 1–24, 2008.

[75] M. Akiyama, M. Shimamura, Y. Kadobayashi, S. Yamaguchi, T. Kawamoto, and T. Yokoyama, "A proposal of metrics for botnet detection based on its cooperative behavior," in *International Symposium on Applications and the Internet Workshops. IEEE*, 2007.

[76] J. Goebel and T. Holz, "Rishi: identify bot contaminated

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

hosts by IRC nickname evaluation," *HotBots'07 Proc. first Conf. First Work. Hot Top. Underst. Botnets*, p. 8, 2007.

[77] W. Lu and A. A. Ghorbani, "Botnets Detection Based on IRC-Community," in *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, 2008, pp. 1–5.

[78] W. Wang, X. Guan, X. Zhang, and L. Yang, "Profiling program behavior for anomaly intrusion detection based on the transition and frequency property of computer audit data," *Comput. Secur.*, vol. 25, no. 7, pp. 539–550, 2006.

[79] Y. Al-Hammadi, "Behavioural correlation for malicious bot detection," University of Nottingham, 2010.

[80] C. Langin, D. Che, M. Wainer, and S. Rahimi, "Visualization of Network Security Traffic using Hexagonal Self-Organizing Maps," in *International Conference on Computer Application and Industry and Engineering*, 2009.

[81] H.-J. Liao, C.-H. Richard Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," *J. Netw. Comput. Appl.*, vol. 36, no. 1, pp. 16–24, Jan. 2013.

[82] H. R. Zeidanloo, M. J. Zadeh, Shooshtari, P. V. Amoli, M. Safari, and M. Zamani, "A taxonomy of Botnet detection techniques," in *Proceedings - 2010 3rd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2010*, 2010, vol. 2, pp. 158–162.

[83] S. Saad *et al.*, "Detecting P2P botnets through network behavior analysis and machine learning," in *2011 Ninth Annual International Conference on Privacy, Security and Trust*, 2011, pp. 174–180.

[84] L. Zhang, S. Yu, D. Wu, and P. Watters, "A survey on latest botnet attack and defense," in *Proc. 10th IEEE Int. Conf. on Trust, Security and Privacy in Computing and Communications, TrustCom 2011, 8th IEEE Int. Conf. on Embedded Software and Systems, ICESS 2011, 6th Int. Conf. on FCST 2011*, 2011, pp. 53–60.

[85] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nat. Methods*, vol. 13, no. 1, p. 35, 2015.

[86] S. C. Guntuku, P. Narang, and C. Hota, "Real-time Peer-to-Peer Botnet Detection Framework based on Bayesian Regularized Neural Network," *arXiv Prepr. arXiv*, vol. 1307, no. 7464, Jul. 2013.

[87] Nemir Ahmed Al-Azzawi and Shatha Mizhir Hasan, "Detection of P2P Botnets Based on Support Vector Machine: Case Study," *Eng. Technol. J.*, vol. 32, no. 5, pp. 1227–1238, 2014.

[88] P. Narang, V. Khurana, and C. Hota, "Machine-learning approaches for P2P botnet detection using signal-processing techniques," in *DEBS 2014 - Proceedings of the 8th ACM International Conference on Distributed Event-Based Systems*, 2014, pp. 338–341.

[89] S. Buriya, D. Bhilare, A. Kumar Patel, and S. Singh Yadav, "Botnet Behavior Analysis Using Naïve Bayes Classification Algorithm Without Deep Packet Inspection," *Int. J. Comput. Eng. Appl.*, vol. IX, no. VIII, 2015.

[90] P. Burghouwt, "Detection of Botnet Command and Control Traffic in Enterprise Networks," The Hague University of Applied Sciences, 2015.

[91] A. Sanatinia and G. Noubir, "OnionBots: Subverting Privacy Infrastructure for Cyber Attacks," in *2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, 2015, pp. 69–80.

[92] J. Sherry, C. Lan, R. A. Popa, and S. Ratnasamy, "BlindBox," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 5, pp. 213–226, Aug. 2015.

[93] J. Jianguo, B. Qi, S. Zhixin, Y. Wang, and B. Lv, "Botnet detection method analysis on the effect of feature extraction," in *Proceedings - 15th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 10th IEEE International Conference on Big Data Science and Engineering and 14th IEEE International Symposium on Parallel and Distributed Proce*, 2016, pp. 1882–1888.

[94] S. Cha and H. Kim, "Detecting encrypted traffic: A machine learning approach," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10144, pp. 54–65.

[95] H. Zhang, "Detecting Advanced Botnets in Enterprise Networks," Colorado State University, 2017.

[96] N. S. Raghava, D. Sahgal, and S. Chandna, "Classification of Botnet Detection Based on Botnet Architechture," in *2012 International Conference on Communication Systems and Network Technologies*, 2012, pp. 569–572.

[97] A. Dange and P. Gosavi, "Botnet Detection through DNS based approach," *Int. J. Appl. or Innov. Eng. Manag.*, vol. 2, no. 6, 2013.

[98] P. Gibbs, "Botnet Tracking Tools," *SANS Inst. InfoSec Read. Room*, vol. 21, no. 2, 2019.

[99] S. Arshad, M. Abbaspour, M. Kharrazi, and H. Sanatkar, "An anomaly-based botnet detection approach for identifying stealthy botnets," in *ICCAIE 2011 - 2011 IEEE Conference on Computer Applications and Industrial Electronics*, 2011, pp. 564–569.

[100] C. Langin, H. Zhou, S. Rahimi, B. Gupta, M. Zargham, and M. R. Sayeh, "A self-organizing map and its modeling for discovering malignant network traffic," in *2009 IEEE Symposium on Computational Intelligence in Cyber Security*, 2009, pp. 122–129.

[101] G. Fedynyshyn, M. C. Chuah, and G. Tan, "Detection and classification of different botnet C&C channels," in

Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2011, vol. 6906 LNCS, pp. 228–242.

[102]C. Kanich *et al.*, "Spamalytics: An empirical analysis of spam marketing conversion," in *Proceedings of the 15th ACM conference on Computer and communications security*, 2008.

[103]H. R. Zeidanloo, A. B. Manaf, P. Vahdani, F. Tabatabaei, and M. Zamani, "Botnet detection based on traffic monitoring," in *ICNIT 2010 - 2010 International Conference on Networking and Information Technology*, 2010, pp. 97–101.

[104]B. Al-Duwairi and L. Al-Ebbini, "BotDigger: A fuzzy inference system for botnet detection," in *5th International Conference on Internet Monitoring and Protection, ICIMP 2010*, 2010, pp. 16–21.

[105] M. Stevanovic and J. M. Pedersen, "On the use of machine learning for identifying botnet network traffic," *J. Cyber Secur. Mobil.*, vol. 4, no. 2, pp. 1–32, 2016.

[106]R. S. Roshna and V. Ewards, "Botnet Detection Using Adaptive Neuro Fuzzy Inference System," *Int. J. Eng. Res. Appl.*, vol. 3, no. 2, pp. 1440–1445, 2013.

[107]V. Thomas and N. Jyoti, "Defeating IRC bots on the internal network." Virus Bulletin, 2007.

[108]P. Salvador, A. Nogueira, U. França, and R. Valadas, "Framework for zombie detection using neural networks," in *Proceedings - 2009 4th International Conference on Internet Monitoring and Protection, ICIMP 2009*, 2009, pp. 14–20.

[109]L. Bilge, S. Sen, D. Balzarotti, E. Kirda, and C. Kruegel, "EXPOSURE: A passive DNS analysis service to detect and report malicious domains," *ACM Trans. Inf. Syst. Secur.*, vol. 16, no. 4, 2014.

[110]C. Han and R. Dongre, "Q and A," *Acad. Manag. Proc.*, vol. 4, no. 10, p. 363, 2014.

[111]S. Kumar, P. Singh, R. Sehgal, and J. S. Bhatia, "Distributed Honeynet System Using Gen III Virtual Honeynet," *Int. J. Comput. Theory Eng.*, vol. 4, no. 4, 2012.

[112]A. Jakalan, G. Jian, and L. S. Dong, "Distributed low-interaction honeypot system to detect botnets," in *International Conference on Computer Engineering and Technology, 3rd (ICCET 2011)*, 2011.

[113]G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee, "BotHunter: Detecting Malware Infection Through IDS-Driven Dialog Correlation," in *16th USENIX Security Symposium*, 2007.

[114]C. J. Dietrich, C. Rossow, and N. Pohlmann, "CoCoSpot: Clustering and recognizing botnet command and control channels using traffic analysis," *Comput. Networks*, vol. 57, no. 2, pp. 475–486, 2013.

[115]S. García, M. Grill, J. Stiborek, and A. Zunino, "An empirical comparison of botnet detection methods," *Comput. Secur.*, vol. 45, pp. 100–123, 2014.

[116]J. Sherry, C. Lan, R. Ada Popa ETH Zürich, and U. Berkeley Sylvia Ratnasamy, "BlindBox: Deep Packet Inspection over Encrypted Traffic," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, pp. 213–226, 2015.

[117]M. Warmer, "Detection of Web-based Command & Control Channels," University of Twnte, 2011.

[118]D. Zhao *et al.*, "Botnet detection based on traffic behavior analysis and flow intervals," *Comput. Secur.*, vol. 39, no. PARTA, pp. 2–16, 2013.

[119]S. Thomas and J. M. Pittman, "A Survey of Botnet Detection Techniques by Command and Control Infrastructure," *J. Digit. Forensics, Secur. Law*, vol. 10, no. 1, 2015.

[120]G. Kirubavathi and R. Anitha., "Botnet detection via mining of traffic flow characteristics," *Comput. Electr. Eng.*, vol. 50, 2016.

[121]T. J. Richer, "Entropy-based detection of botnet command and control," in *ACM International Conference Proceeding Series*, 2017.

[122]C. Rossow and C. J. Dietrich, "ProVeX: Detecting botnets with encrypted command and control channels," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 7967 LNCS, pp. 21–40.

[123]A. A. Khan, C. Ahlawat, and A. Bijalwan, "A unified botnet detection framework," *Int. J. Adv. Electron. Comput. Sci.*, vol. 2, pp. 81–87, 2015.

[124]S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised Machine Learning: A Review of Classification Techniques," *Emerg. Artif. Intell. Appl. Comput. Eng.*, vol. 160, pp. 3–24, 2007.

[125]A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, 1999.

[126]A. S. Bist, "A Survey of deep learning algorithms for malware detection," *Int. J. Comput. Sci. Inf. Secur.*, vol. 16, no. 3, 2018.

[127]Y. Upendra and K. Jain, "Intrusion detection using supervised learning with feature set reduction," *Int. J. Comput. Appl.*, vol. 975, no. 8887, 2011.

[128]S. Raval, "Unsupervised Learning." YouTube Video, 2019.

[129]O. M. K. Alhawi, J. Baldwin, and A. Dehghantanha, "Leveraging machine learning techniques for windows ransomware network traffic detection," *Adv. Inf. Secur.*, vol. 70, pp. 93–106, 2018.

[130]A. M. Pirbazari, A. Chakravorty, and C. Rong, "Evaluating Feature Selection Methods for Short-Term

January - February 2020
ISSN: 0193-4120 Page No. 3386 - 3408

Load Forecasting," *2019 IEEE Int. Conf. Big Data Smart Comput. BigComp 2019 - Proc.*, pp. 1–8, 2019.

[131]M. U. Kiru and A. B. Jantan, "The Age of Ransomware: Uniderstading Ransomware and its countermeasures," in *Artificial Intelligence and Security Challenges in emerging networks*, R. Abassi, Ed. Pennsylvania: IGI Global, 2019, pp. 1–37.

[132]M. K. Ubale and S. M. Isyaku, "A Situation Analysis on Cybercrime and its Economic Impact in Nigeria," *Int. J. Comput. Appl.*, vol. 169, no. 7, pp. 975–8887, 2017.

[133]A. Nogueira, P. Salvador, and F. Blessa, "A botnet detection system based on neural networks," in *5th International Conference on Digital Telecommunications, ICDT 2010*, 2010, pp. 57–62.

[134]G. Kulkarni, V. Waykule, and H. Bankar, "Cloud Storage Architecture," in *2012 7th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, 2012, no. June, pp. 76–81.

[135]Statistica, "Countries with highest ransomware infection rates 2016-2017," 2017.

[136] S. Hartinah, S. Suherman, M. Syazali, H. Efendi, R. Junaidi, K. Jermsittiparsert,&R. Umam. (2019). Probing-Prompting Based on Ethnomathematics Learning Model: The Effect on Mathematical Communication Skill. Journal for the Education of Gifted Young Scientists, 7(4), 799-814.