REVIEW

A survey of botnet detection based on DNS

Kamal Alieyan[1] · Ammar ALmomani[2] · Ahmad Manasrah[3] · Mohammed M. Kadhum[1]

Received: 8 September 2015 / Accepted: 16 November 2015 / Published online: 12 December 2015

Abstract Botnet is a thorny and a grave problem of today's Internet, resulting in economic damage for organizations and individuals. Botnet is a group of compromised hosts running malicious software program for malicious purposes, known as bots. It is also worth mentioning that the current trend of botnets is to hide their identities (i.e., the command and control server) using the DNS services to hinder their identification process. Fortunately, different approaches have been proposed and developed to tackle the problem of botnets; however, the problem still rises and emerges causing serious threat to the cyberspace-based businesses and individuals. Therefore, this paper comes up to explore the various botnet detection techniques through providing a survey to observe the current state of the art in the field of botnet detection techniques based on DNS traffic analysis. To the best of our knowledge, this is the first survey to discuss DNS-based botnet detection techniques in which the problems, existing solutions and the future research direction in the field of botnet detection based on DNS traffic analysis for effective botnet detection mechanisms in the future are explored and clarified.

Keywords Botnet · Machine learning · DNS-based · Botnet detection · Neural Network

1 Introduction

The increasing reliance on the Internet in our daily lives adds a lot of challenges in terms of managing the Internet and the application usage, such as protecting the user data, privacy, integrity and availability. In the last years, the Internet plays a main role in our lives, especially in communication, education, government services, banking and e-commerce [1]. Unfortunately, the increasing demands on the user's applications become a threat to his privacy and data security [2]. Botnet is a software program that manipulates computers for malicious purposes, known as bots. Bots are small scripts built to carry out specific automated tasks [3]. These bots are controlled by one or a small collaborative group of attackers known as "botmaster" [4]. Based on McAfee Labs statistics, the number of newly discovered malware samples has reached 50 millions in the fourth quarter of 2014 and expected to reach half a billion by the end of 2015 [5]. Moreover, the Internet traffic consisted of up to 80 % of botnets traffic related to spam e-mails originating from known botnets such as Grum, Cutwail and Rustock [6]. Currently, a large scale of botnets can be more than one million PCs launching cyber attacks [7].

Botnets differ from any other types of malware through utilizing a communication channels to receive commands

✉ Ammar ALmomani
ammarnav6@gmail.com

Kamal Alieyan
Kamal_alian@nav6.usm.my

Ahmad Manasrah
ahmad.a@yu.edu.jo

Mohammed M. Kadhum
kadhum@nav6.usm.my

[1] National Advanced IPv6 Centre (NAV6), Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia

[2] Department of Information Technology, Al-Huson University College, Al-Balqa Applied University, P.O. Box 50, Irbid, Jordan

[3] Faculty of Information Technology and Computer Sciences, Yarmouk University, Irbid 21163, Jordan

and report their current status to their operator(s) [8]. According to the cybersecurity in the Golden State Report in 2014, cyber attacks on Christmas Eve against the Web site of a regional California bank helped to disrupt the work of the bank officials from recovering an account takeover of one of their clients, netting a cyber theft of more than $900,000 [9]. In addition, the FBI in 2013 reported that 10 international hackers were arrested for using botnets to steal more than $850 million through a group of compromised computers; they use the personal financial information of the people to steal such amount [10].

In fact, botnets have specific characteristics as compared to other types of malware. For instance, the botmaster can control the infected machines and send commands without directly communicating with them. There are also a lot of bots working in a coordinated way and taking instructions from the botmaster to instantiate coordinated attacks such as the distributed denial-of-service (DDoS) attacks, spam distribution and click fraud [11]. Also, the botnet provide these frauds as a service form botnet operator which are consider part of the botnet economy [12].

Domain Name System (DNS) is a fundamental element of the Internet functionality, which converts domain names to their corresponding IP addresses. However, the security of the DNS system is the responsibility of the whole Internet collaboratively [13]. The distributed and global system of the DNS motivates the cyber criminals to attack on a global scale [14]. To commit their crimes, attackers make use of DNS services to operate malicious networks, such as botnets and other types of malwares [15].

In addition, studies have demonstrated the challenges in tracking malicious domains using web content analysis or human observation due to the huge number of available domains within the cyberspace [14]. Unfortunately, botnets use the DNS traffic as any other legitimate host, which makes differentiating the legitimate DNS traffic from the illegitimate one a very challenging problem [16]. Moreover, botnet owners attempt to hide their communication with the bots to obstruct any deployed botnet detection processes [17]. The attackers or botmasters use the DNS services to hide their command and control (C&C) IP address to make the botnet reliable and easy to migrate from server to another without being noticed [18].

Generally, according to Bilge et al. [19] and Davuth and Kim [14], the use of DNS traffic characteristics to detect the botnet is directed to two different tracks. The first track is to detect the domains that are part of malicious activities aiming to recognize the infected hosts by monitoring the DNS traffic [19, 20]. The second track is to focus on the behavior of a group of machines requesting the same domain name frequently in a coordinated manner [14].

This survey identifies and classifies the various DNS-based methods and techniques for botnet detection. Moreover, this survey provides comparisons of the various known method in the field and scrutinizes their characteristics, weaknesses and strengths. To the best of our knowledge, this paper is the first to provide an up-to-date outline of the existing DNS-based botnet detection methods. The paper presents a systematic overview of the contemporary detection methods, with the goal of contributing to the better understanding of capabilities, limitations and opportunities of using DNS methods for identifying botnet traffic. An overview of botnet phenomena, life cycle of botnet, classification of botnet and a new classification of botnet detection techniques based on DNS is also presented.

The remaining sections of this paper are organized as follows: Sect. 2 contains a background and an overview of the botnet. Section 3 discusses the classification of botnet detection techniques based on DNS traffic analysis. Section 4 presents the summary and discussion. And Sect. 5 concluded this work and highlights the future research directions.

2 Backgrounds and botnet overview

Botnet is one of the most significant threats to the cybersecurity as they are considered a launching pad for a number of several illegal activities such as distributed denial of service (DDoS), click fraud [21], phishing, identity theft [22], spamming [23] and malware distribution [24]. Until now, there exists no permanent solution for the detection or mitigation of botnets threats because their techniques and methods keep changing over time [6]. The botnet detection process stands as an ongoing challenge for researchers and organizations. Therefore, understanding the botnet life cycle and their architecture may yield to better detection mechanisms.

2.1 The botnet life cycle

Generally, botnets apply similar set of steps to recruit members and form the zombie army. These sets can be considered as a life cycle of botnet and illustrate the steps of any botnets life cycle. The typical bots can be created and preserved in four phases.

2.1.1 Exploitation phase

This phase is the first step in the botnet life cycle. The botmaster makes a remote infection by exploiting an existing vulnerability of software running on the victim host. The botmaster defrauds the victim user to execute a malicious code on his machine, such as opening an e-mail attachment [25]. In this phase, the bots need to connect to a remote server to download the bot binaries. The connection to a remote server is established only after a DNS lookup command is issued by the compromised machine to map a domain name to its corresponding IP address [26]. This behavior of issuing a DNS lookup query is the dominant behavior of almost all botnets that are existed in the cyberspace [16].

2.1.2 *Rallying phase*

In this phase, the bots are connecting back to their botmaster through porting to a C&C server. The botmaster intends to make his botnet portable and stealth at the same time. Therefore, the botmaster equips his bots with a DNS lookup functionality to be able to perform DNS queries to locate the command and control (C&C) server. Unfortunately, botmasters have learnt that a static IP address of the C&C is not effective and vulnerable to be identified and blacklisted. Therefore, they start to misuse the DNS services to hide the location of the C&C server behind a domain name rather than a static IP address. As a result, bots will rally to connect back to the C&C server as soon as they obtain the location of the needed server [18]. This phase is considered very vital to the success of the botnet stealth nature and power [27].

2.1.3 *Attack execution phase*

In this phase, the group of bots performs malicious activities on target machines as instructed by the botmaster through sending the needed commands to the C&C servers. Bots will then grab the command from the C&C server to start the malicious activities. For instance, the group of bots may receive commands from the C&C server to redirect users' requests to certain malicious Web sites through capturing the users' DNS queries [17].

2.1.4 *Update and maintenance phase*

The last phase of the botnet life cycle is updating and maintaining the bots of the botnets. The botmaster needs to keep his bots up to date through instructing the bots to update their binaries from time to time for better coordination and patching [6]. Moreover, botmasters may require migrating his C&C server location frequently to evade the various detection techniques [6]. Understanding this phase is very important, because botnets can be identified through observing the same network behaviors and communication patterns/frequency from the bots to their C&C server (Fig. 1).

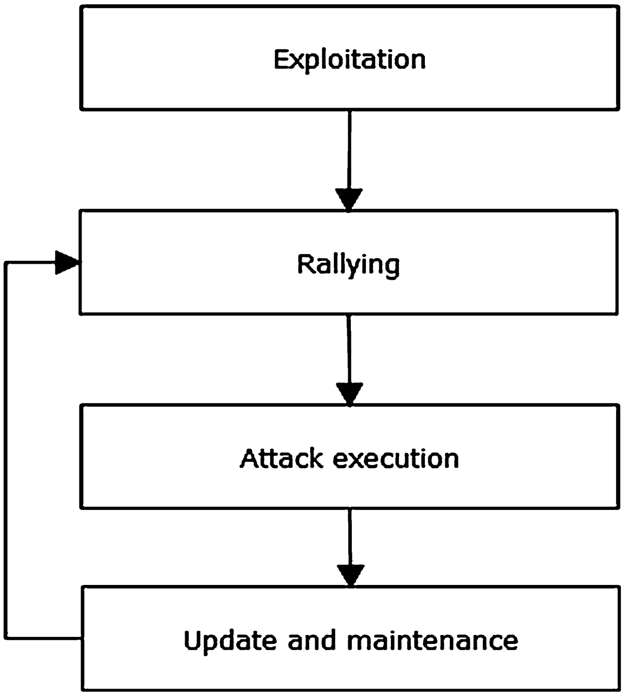
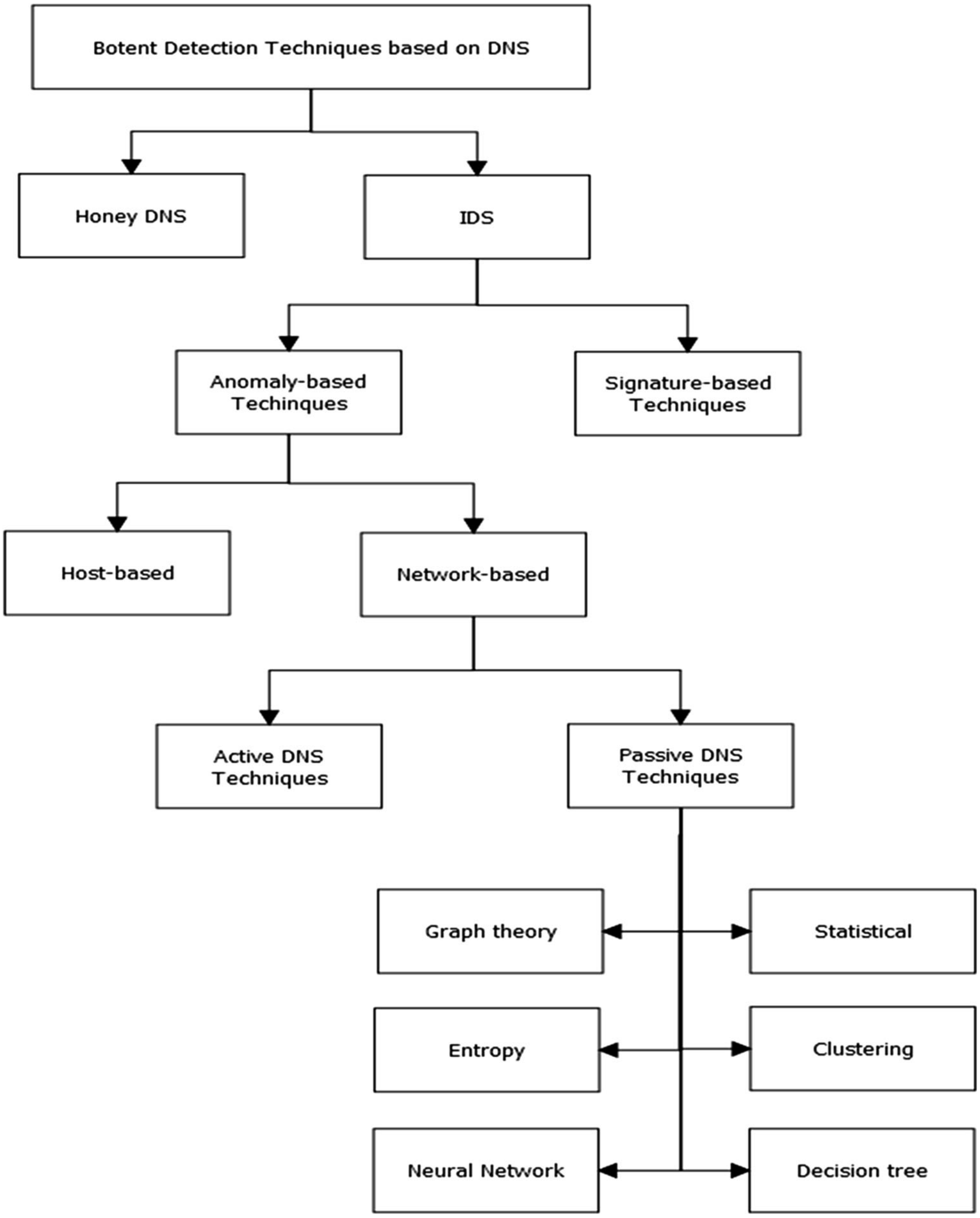


Fig. 1 Life cycle of a botnet

3 Classification of botnet detection techniques based on DNS traffic analysis

For the purpose of detecting botnets, different architectures and techniques were proposed to eliminate the danger of botnets attack. Moreover, researchers have produced different classifications for better understanding the botnet phenomena and its structure [6, 17, 28–31]. The botnet detection techniques were mainly classified into two types of approaches: those who are based on installing and configuring a honeynet within the monitored network and the intrusion detection system (IDS) [24, 26, 29, 31]. This section, however, provides a new classification of botnet detection techniques that are based on analyzing the DNS traffic characteristics and usage as well as their implantation. In recent years, the Domain Name System (DNS) has been misused by attackers as they succeed in keeping their malicious systems alive, stealth and organized. Consequently, current researches start to move toward approaches on botnet detection that leverage the distinctive features between benign and malicious DNS traffic and their usage [19]. Despite the many classifications on botnet detection, there is no comprehensive classification for botnet detection techniques that are based on DNS traffic characteristics yet. Figure 2 portrays the classification of the botnet detection techniques that are based on DNS traffic analysis.

Fig. 2 Classification of botnet detection techniques based on DNS traffic characteristics



Weimer [32] proposed a system that collects domain names passively from the DNS traffic and stores them into a database for malicious behavior analysis. Similarly, Zdrnja et al. [33] applied the concept of passive monitoring to detect DNS traffic anomalies. The authors disputed in their work, the potential in differentiating unusual or anomaly DNS from benign DNS behaviors. However, the authors did not mention the DNS features that need to be captured and stored [33].

The detection of botnets based on DNS traffic analysis has the potential to spot the real-world botnets without any prior knowledge of their communication protocols and structures [28]. The botnet detection techniques that are based on DNS traffic analysis are considered a promising research direction toward combating botnet threats [28], through which the attacks can be evaded before they happen [34]. The main purpose of carrying out this survey is to provide an understanding of the proposed researches in the field of botnet detection based on DNS traffic characteristics. With that in mind, this survey is tracking and providing a review of the most well-known botnet detection techniques.

3.1 Honeynet-based approaches

Honeynet-based approaches are used mainly to analyze and understand the behaviors and the characteristics of botnets. Honeynets emulate known software and network vulnerabilities to be infected by botnets [30]. The honeynets are prepared to be self-contained and thwart the extension of botnets. In addition, the honeynets are used to discover the capabilities of unknown attacks, the C&C system, the attackers' tools, techniques and motivation [6]. Different techniques and approaches were proposed based on honeynet systems to capture the botnets characteristics as in [25, 35–39]. The honeynets are substantial to understand botnet characteristics and technology [24, 31]. One of the known works in honeynet that uses the DNS queries is the one proposed by Oberheide et al. [36] who apply some basic statistics over the collected DNS queries. This work dealt with DNS queries targeting unused (i.e., darknet) address spaces and developed the honeydns system concept to assist honeypots to prevent the attackers from initiating their attacks [40, 41].

Despite all of that, the honeynet-based systems are easy to build and deploy with minimum cost and resource requirements. Nevertheless, there are some drawbacks for the honeynet systems including limited scalability and interaction with malicious activities. Moreover, attackers may use the honeynet to learn new evasion techniques [6]. As a result, honeynets are aimed to recognize the features and mechanisms of botnet, but cannot detect bot infections all the time [24].

3.2 Intrusion detection system (IDS)

The intrusion detection system (IDS) for botnet detection can be classified into two techniques: signature-/behavior-based IDS [26, 42] and anomaly-based IDS [29, 31, 43–46]

3.2.1 Signature-based IDS

The signature-based techniques detect only known bots through signature matching using the IDS such as Snort [47]. A DNS-based blacklist (DNSBL) approach proposed by Ramachandran et al. [48] is an example of a signature-based botnet detection system. The DNSBL-based approaches look for known bot signatures within the monitored DNS traffic. The DNSBL-based approaches are also used to publish malicious and spamming activities online through collecting IP addresses of server machines or networks related to these activities. DNSBL-based approaches attempt to recognize the botmasters' address and identify their location as shown in Fig. 3. However, the limitation of the DNSBL-based approaches resides in maintaining an up-to-date database of known malicious addresses. Unfortunately, one of the basic defense lines against DNS abuses is domain name blacklisting [49, 50].

Similarly, Antonakakis et al. [51] built a dynamic DNS reputation system called "Notos" that uses the passive DNS query data and analyzes the network and zone feature of a domain name as shown in Fig. 4. The *Notos* system assumed that the malicious DNS query has distinctive characteristics that are distinguishable from the benign DNS query [51]. Thus, observing DNS queries and building models of known malicious and benign domains are feasible and might lead to a good result. A reputation score for the new domain observed was computed by models that give low scores for malicious domains and high score for benign domains to differentiate between them. The Notos system has achieved high accuracy and low false-positive rate, and it can recognize the new domains before they get released to the public blacklist. However, the system needs a lot of history for a given domain name to reach a correct reputation score, and it is inaccurate against frequently changing C&C domains like the hybrid botnet architecture that uses many master C&C nodes to distribute its commands [52].

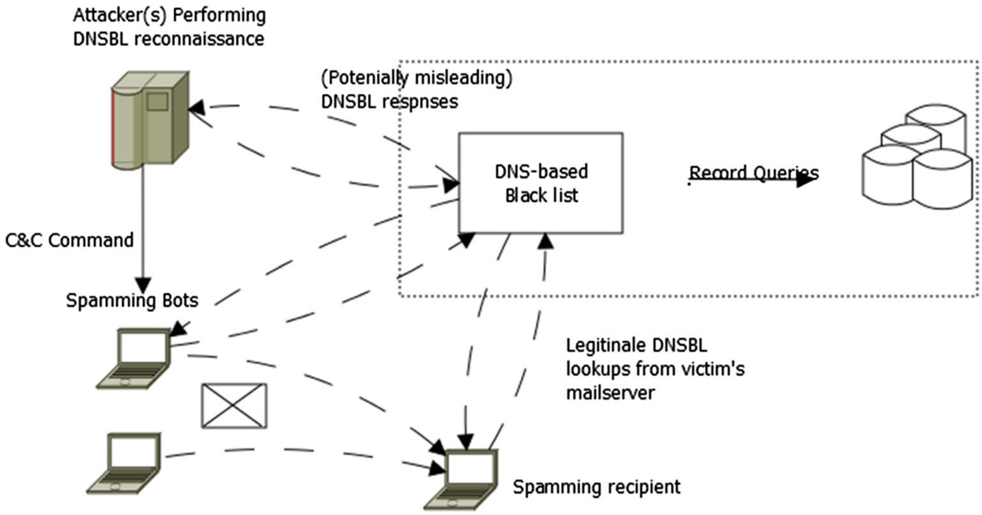
In contrast to previous work, the mentor system that was proposed by Kheir et al. [52] depended on removing the legitimate domains from the botnet C&C blacklisted domains, to reduce the false-positive ratio during the detection process as shown in Fig. 5. The mentor system implements scalable, positive DNS reputation system that automatically removes benign or harmless domains recorded inside a blacklist of botnet C&C domains [52]. The mentor system collects statistical features about a suspected domain name such as DNS properties and web content to build a DNS pruning model by applying a supervised learning technique into a labeled set of known benign and malicious domain names. The result of the mentor system was effective after tested over public blacklist, and it removed benign domain names having a very low false-positive rate.

Yadav et al. [53] proposed an approach to detect "domain fluxes" in DNS traffic through searching for algorithmically generated patterns in domain names and different from the domains generated by humans [53]. They observed a distribution of alphanumeric characters together with the bigrams of all the domains mapped to the same set of IP addresses. However, this system is limited in the detection of C&C domains for only known malware samples that are correctly performed in the training phase, so it cannot identify unknown botnets [52]. Table 1 shows the summary of DNS signature-based botnet detection techniques. Generally, the signature-based detection techniques have many limitations; mainly, they require a constant updating to detect the new botnets or zero-day botnet attacks in which the signatures are evolving [31].

3.2.2 Anomaly DNS-based botnet detection techniques

The anomaly-based or behavior-based techniques attempted to detect botnets through analyzing network traffic for anomalies like a sudden vast amount of traffic, traffic to unusual ports, high network latency and anomalous behavior that may indicate the existence of bots in the network [24].

Fig. 3 DNSBL-based spam mitigation architecture [2]



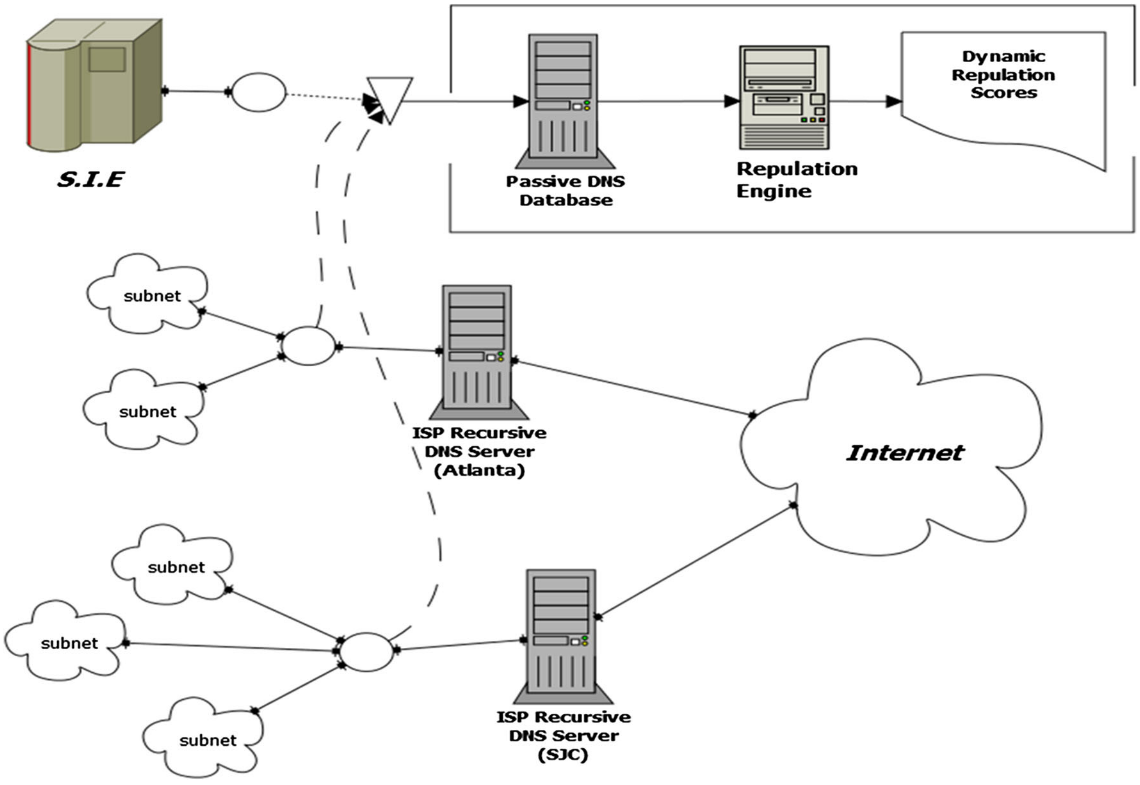


Fig. 4 Building a dynamic reputation system for DNS (Notos) [2]

These approaches have the ability to identify new bots. The anomaly-based technique can be categorized into host-based and network-based detection techniques [6, 24, 31].

3.2.3 *Host-based anomaly detection techniques*

In the host-based approaches, the monitoring and the analyzing process are made locally at each individual computer to detect any malicious activities through monitoring system processes, access to kernel-level routines and system calls [6]. An example of the host-based technique is BotSwat which was proposed by Stinson and Mitchell [54]. BotSwat focuses on the way bots respond to data received over the network through monitoring the execution of an arbitrary Win32 binary. However, the main limitation of the host-based botnet detection technique is that it is not scalable and limited to only bots within the monitored hosts. Moreover, to cover a wider view of the

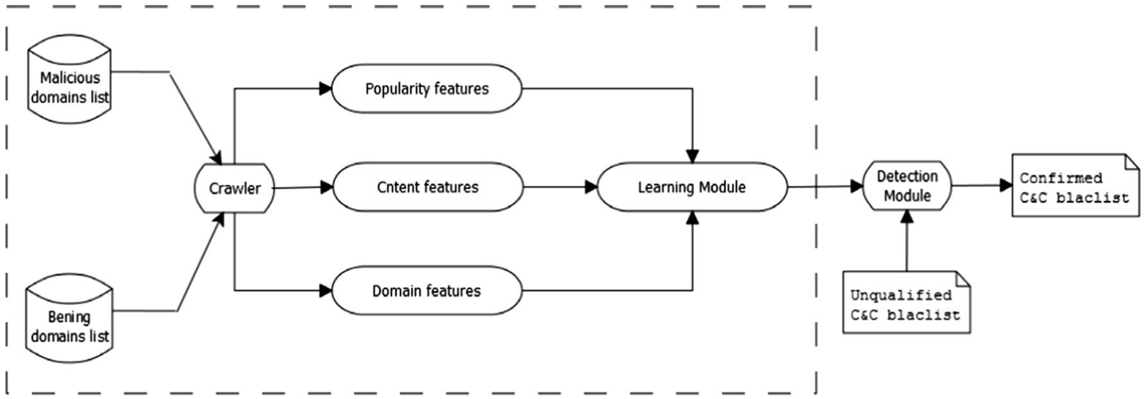


Fig. 5 Mentor system overview [2]

Table 1 Summary of DNS signature-based botnet detection techniques

Proposed model	Mechanism	Weakness
DNSBL (Ramachandran et al. [48])	Collecting published IP addresses of server machines	Update DNS-based blacklist
		Hard to design evasion techniques
Notos (Antonakakis et al. [51])	Dynamic DNS reputation system	Need lots of history for the given domain name to make reputation score
	Uses passive DNS query data to analyze the network feature of a domain name	Unreliable with hybrid botnet
Mentor (Kheir et al. [52])	Removing the legitimate domains from the botnet C&C domains blacklist	Need frequent feeding of information to the system
Yadav et al. [53]	Detect domain fluxes in DNS traffic	Limited to known botnets
	Seeking for algorithmically generated patterns inherent to domain names	Attacks evaded detection during analysis

network, each individual host should be equipped with powerful monitoring tools that work collaboratively with others [24].

One of the key techniques that focus on monitoring DNS traffic at the host/network level is the EFFORT framework that was proposed by Shin et al. [55]. This framework aims for effective and efficient detection through applying the multi-module approach that correlates bot-related indications from different clients and network-level aspects as shown in Fig. 6.

EFFORT framework used a supervised machine learning algorithm to distinguish the queried domain names for being benign or malicious domains. EFFORT framework is independent of topology, deployed communication protocol and capable of detecting encrypted protocols. However, EFFORT framework is limited in its scope to botnets that depend on DNS services for identifying the address of their C&C servers. To the best of our knowledge, EFFORT is the only DNS-host-based botnet detection framework that is available in the literature as consolidated in [55].

3.2.4 Network-based anomaly detection techniques

The second type of anomaly detection techniques is the network-based detection techniques. These detection techniques monitor the network traffics to identify the existence of botnets [56]. These techniques can be classified into active and passive monitoring methodologies [6, 24, 30, 31, 57].

3.2.4.1 Active monitoring techniques For the active monitoring methodology, special crafted packets are injected into the monitored network to stimulate the network to respond. The responses are then captured and analyzed for any performance- or malware-related evidences. BotProbe [58], Strayer [59] and Xiaobo [60] are examples of active monitoring tools [60]. The active monitoring techniques can leverage active DNS probing methods to identify malicious domains that might relate to botnet activities (i.e., spam) [60–62]. Accordingly, Ma et al. [61] extracted URL features from spam e-mails, and they used statistical and machine learning methods to

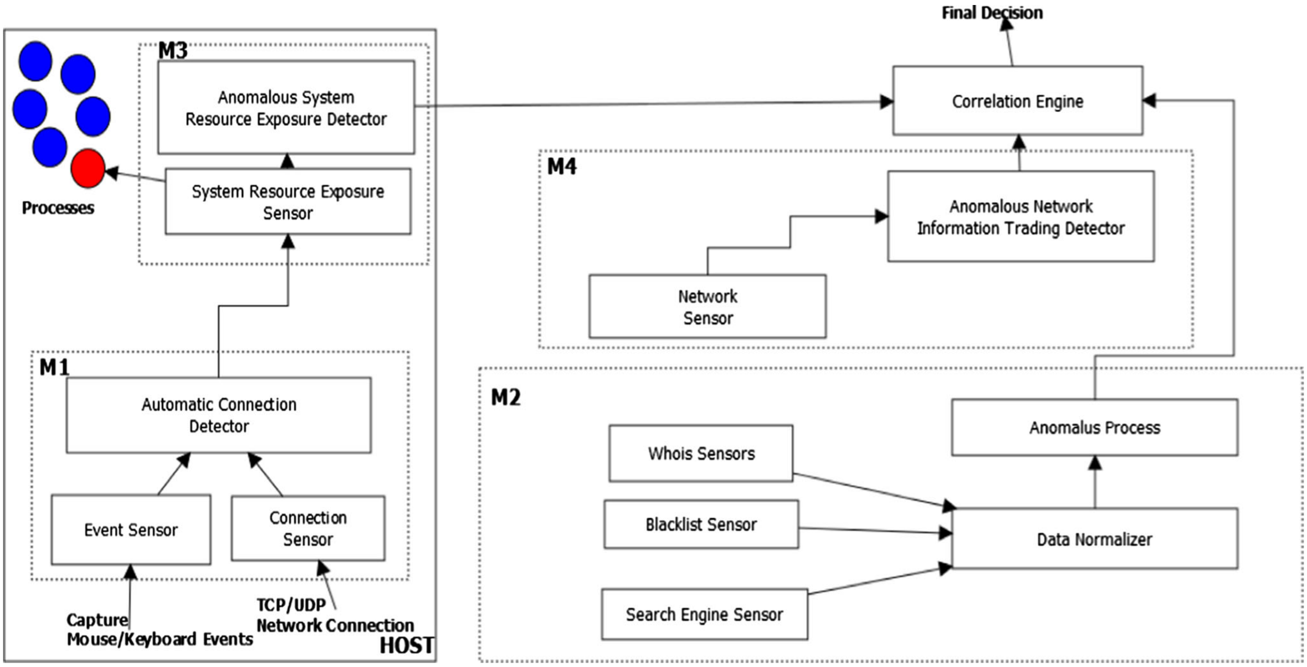

Final Decision
Correlation Engine
M3
Anomalous System Resource Exposure Detector
System Resource Exposure Sensor
Processes
M4
Anomalous Network Information Trading Detector
Network Sensor
M1
Automatic Connection Detector
Event Sensor
Connection Sensor
Capture Mouse/Keyboard Events
TCP/UDP Network Connection
HOST
M2
Whois Sensors
Blacklist Sensor
Search Engine Sensor
Anomalus Process
Data Normalizer

Fig. 6 EFFORT system overview [1]

classify malicious Web sites. This approach analyzes spam URLs based on lexical construction and the host name information. To gain the information from the host name, the authors carry out a DNS active probing to retrieve the number of IP addresses linked to the domain [61]. In particular, the approach takes into consideration the domain name properties such as time to live (TTL) and the value for DNS records attached to the host name. However, the main drawback of this approach is that it is targeting only domains known for spam campaign [19], but it cannot observe the class of malicious domains that are related to malware activities such as botnet C&C servers.

The recent approach by Ma et al. [61] aims to estimate certain DNS query characteristics based on the DNS cache activities. The characteristics can be obtained through active DNS probing on a large scale with minimum management cost and low privacy concerns. The approach integrates the renewal theory-based DNS caching formulation and the hyper-exponential disseminate model [60]. In addition, the approach performs a large-scale real-world DNS trace measurement and has reached high accuracy. This approach also functions in remote management networks to identify the infected host (Table 2).

However, all active monitoring techniques share the same limitation that summarized as injecting extra unwanted payloads to the network [31, 57]. On the other hand, the active DNS probing and analysis have a high possibility of being detected by the perpetrator who controls the domains under analysis [19]. Moreover, active monitoring techniques are subject to the point of investigation only, which makes it limited to the issues it can identify at one time. Table 3 shows the summary of botnet detection techniques based on active DNS probing and monitoring.

3.2.4.2 Passive monitoring techniques The Domain Name System (DNS) stores and provides information to the queried domain name. When the DNS entry expires, the DNS system will discard the expired information and fetches the updated one. Unfortunately, botmasters have also learnt this facet and become proficient in hiding their tracks. Therefore, it is essential for organizations to follow and analyze domain name tracks accumulatively over time. As a result, passive DNS or passive DNS replication that was first proposed by Weimer [32] aims to capture the inter-server DNS messages by sensors and forward them to a collection point for analysis [32]. The collected DNS messages and their history give the ability to study and keep track of malicious domain names even though they are removed or expired. Monitoring and tracking domain names is a problem that requires a lot of resources, especially in large distributed networks. Accordingly, the use of passive DNS analysis starts to emerge, and researchers adopt it to detect malicious activities [19].

Botnets use the DNS services as any legitimate software or program. Bots issue DNS queries to locate the IP address of a patching server to fetch binary updates [6] as well as the location of the C&C server. Therefore, analyzing the

Table 2 DNS-host-based botnet detection techniques

Proposed model	Mechanism	Algorithms	Weakness
EFFORT (Shin et al. [55]	Applied multi-module to correlate bot-related indications from different client/network aspects	SVM (support vector machine)	Vulnerable to different evasion techniques Limited to botnets relay on DNS

Table 3 Summary of active botnet detection techniques based on DNS traffic analysis

Authors	Mechanism	Weakness
Ma et al. [61]	Analyzes spam URLs based on lexical construction and data from the host name	Hard to observe other classes of malicious domains
Ma et al. [60]	Active probing on a large scale to assess DNS query characteristics based on DNS cache activities	High probability of being detected by attackers and raises a privacy concerns

issued DNS queries may unveil the existence of anomalous activities within the monitored network, which can be part of a botnet-generated traffic [18, 63]. Further inspection of the anomalous DNS queries reveals information related to botnet presence and the C&C server identification [64, 65]. Unfortunately, the new trend of command and control methodologies is attested by using the DNS servers as transient stores of payloads. By using the domain generation algorithms (DGAs) and fast-flux techniques to evade the detection through providing a large number of IP addresses, few of them are linked to the C&C domain name [66, 67]. The DNS queries and their response traffic are considered part of the network traffic composition. Therefore, the DNS-based detection methods can be avoided and considered a promising research direction toward identifying botnets. Moreover, the DNS-based detection methods do not require any prior knowledge about the botnet protocols, communications or signatures because most botnets utilize domain names to locate their command and control servers (C&C), or as rendezvous points for collecting stolen information from infected hosts [68]. As a result, botnet DNS traffic can be possibly identified and monitored through identifying the anomalies within the DNS traffic [24]. In this regard, Cranor et al. [64] apply the graphs analysis and passive/active measurement to identify clients, local DNS servers and authoritative DNS servers as a directed graph. The directed graph has the nodes as the IP addresses of the DNS server machines and the edges as queries commonly formed by clients. However, one of the issues of this scheme is its incapability to handle large datasets as claimed in [6].

Since DNS traffic can be characterized, Dagon et al. [35] pointed out some key metrics to characterize the botnet traffic in different topological structures used during the attack phase. In addition, they assumed a probability to consider different response techniques to stop or obstruct botnets [35]. Toward the end of their research, the authors conducted a comparison of a DNS density rates for botnet traffic against the DNS request rates, and they concluded that the density request rate of the botnet DNS traffic compared with the responses is almost the same. However, a key deficiency of this approach is that botmasters can generate a massive fake DNS queries to interrupt this scheme, hence generating a high rate of false-positive/false-negative alerts [6, 31].

Villamarín-Salomón and Brustoloni [63] evaluated two approaches for identifying botnet C&C servers based on anomalous DDNS traffic. The first approach is based on monitoring domain names with abnormally high or temporally concentrated query rates, while the second approach is based on monitoring abnormally recurring DDNS replies (i.e., the query is for an inexistent name “NXDOMAIN”). The first approach is based on the assumption that botmasters frequently move their C&C servers to avoid being detected and/or blacklisted. This migration may yield high DDNS query rates. On the other hand, the recurring DDNS replies are triggered if the queried domain name is not available any more. Such queries may correspond to bots trying to reach their C&C servers that have been taken down or migrated to new server location. The authors finally concluded that botnet detection techniques that are based on the recurring DDNS queries yield better results compared with the one based on sudden or concentrated DNS traffic rate [63]. However, distinguishing DDNS queries from other DNS queries is a difficult task, especially with large networks. Therefore, the authors attempted to make the distinguishing based only on responses' TTL values. Unluckily, the low TTL value is essential, but it is becoming increasingly common for domain names, apart from DDNS. For instance, many legitimate domains, such as google.com, yahoo.com and weather.com, use low TTL values for DNS-based load balancing. Similarly, some legitimate domain names, such as mozilla.com, are also hosted by DDNS providers. Finally, the proposed approach yields a significant number of false positives (i.e., legitimate names considered anomalous) [6, 31]. Table 4 illustrates a summary of botnet detection techniques that are based on the DNS traffic monitoring.

Various methods and techniques were proposed under the passive monitoring techniques that share the same goal of monitoring DNS traffic for botnet traces. These approaches can be categorized as statistical-, graph-, entropy-, clustering-, neural network- and decision tree-based approaches.

Statistical-based botnet detection techniques A statistical technique based on monitoring DNS traffic to detect botnet was proposed by Marko and Vilhan [69]. They proposed an innovative technique that observes DNS queries in local area networks, and through statistical analysis techniques, they concluded the existence of botnets within the monitored traffic [69]. The conducted experiment reveals that three spammers were identified by observing the large number of DNS mail exchange (MX) queries and five nodes participated in querying pseudorandomly generated domain names. However, this technique can identify botnets after they receive instructions from their C&C server such as sending out spam instruction [6].

In the RB-Seeker system (Redirection Botnet Seeker), Hu et al. [70] proposed a botnet detection approach to detect botnets irrespective of their structure. The redirection botnet means a huge number of compromised computers controlled by the botmaster were used as a redirection or proxy infrastructure [70]. Therefore, the RB-Seeker system collects information related to bots redirection activities such as spatial and temporal activities including the DNS query probing on domain names over time. The authors employ a statistical analysis methodology to distinguish between legitimate and malicious domains. The RB-Seeker system is considered one of the efficient tools to detect the stealthy and aggressive botnets [29]. Unfortunately, the RB-Seeker system is mainly targeting spam botnets [71]. Sanchez et al. [72] developed a support vector machine (SVM)-based classifier to disengage end-user machines from the legitimate mail server (LMS) through analyzing a set of machine features that cannot escape spam initiators easily [72]. This approach has high detection accuracy up to 99.27 %, with 0.44 % false-positive rate and 1.1 % false-negative rate. However, this approach targets small networks, and it is inconvenient for small business e-mail servers. Besides, if the host changes his name, then the whole detection method will become vulnerable to various evasion techniques [2].

Antonakakis et al. [73] proposed a detection system for malware-related domain names through monitoring the flows of a DNS questions and replies from the upper DNS hierarchy, called the Kopis system [73]. The Kopis system relies on some features extracted from the information obtained from the upper DNS hierarchy [51]. Kopis can detect the DNS malware-related domains independently even when there is no IP reputation. As a result, the Kopis system can detect the emerging new botnets, yet it cannot be deployed in local networks as real-time botnet detections systems [19]. Table 5 shows the summary of the passive statistical techniques for botnet detection based on DNS traffic features analysis.

Graph-based botnet detection techniques In this category, the researcher worldwide attempted to model the failed DNS queries to certain domain names as a directed graph. In this context, Jiang et al. [74] proposed an approach that aimed to spot malicious activities via a DNS-based failure graph. This approach employed an algorithm for graph decomposition based on a tri-nonnegative matrix factorization to progressively extract coherent subgraphs from

Table 4 Summary of botnet detection techniques based on DNS traffic

Proposed model	Mechanism	Weakness
Cranor et al. [64]	Characterizing large DNS traces using graphs analysis Identify clients, local DNS servers, DNS servers	Unable to deal with large datasets
Dagon et al. [35]	Using key metrics for measuring botnets on multi-topological structures	Botmaster can be avoided High false positive
Villamarín-Salomón and Brustoloni [63]	Comparing two approaches for identifying botnet C&C server Depend on anomalous DDNS traffic	The TTL of DNS queries can be distinguished incorrectly with famous Web sites

Table 5 Summary of passive statistical techniques for botnet detection based on DNS traffic

Authors	Mechanism	Weakness
Marko and Vilhan [69]	Observed DNS queries in LAN to monitor botnet nodes	Identify bots during processing instructions from the C&C
Hu et al. [70] (RB-Seeker system)	Detect botnet in any structures Collect information related to bots redirection activities such as spatial and temporal features	Only focused on the redirection botnets
Sanchez et al. [72]	SVM-based classifier to distinguish end-user machines from the legitimate mail server	Not scaled to large-scale network
	Used group of features that hard evaded by spam initiators	Vulnerable to evasion techniques
Antonakakis et al. [73] (Kopis system)	Controlling DNS questions and replies from the upper DNS hierarchy	Scaled to local network as real-time system

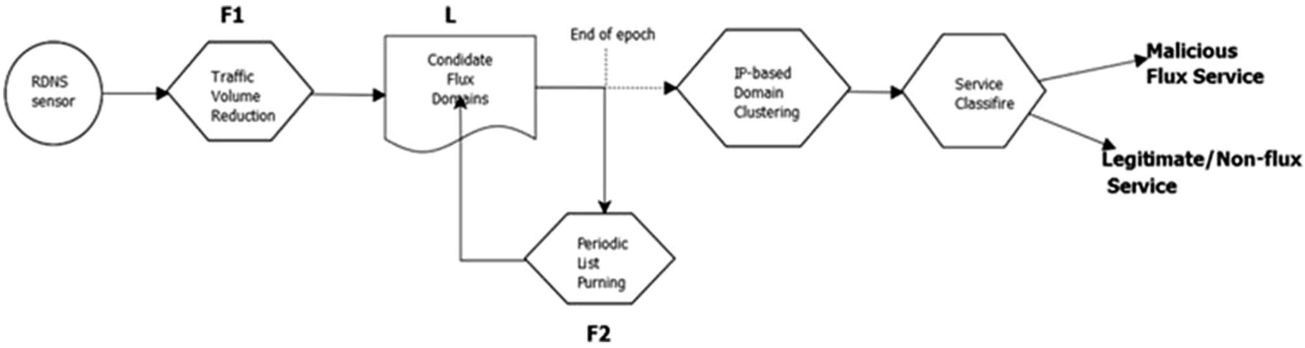


Fig. 7 Overview of the malicious fast-flux service networks detection [75]

the failure of the DNS queries [74]. This approach is based on unpredictable DNS traces which does not always lead to malicious activities [6].

Clustering-based botnet detection techniques Cluster analysis, or clustering in botnet detection, is the task of grouping a set of nodes based on some parameters (i.e., traffic volume and DNS traffic features) in a way that the grouped nodes within the same cluster are more similar to each other than to those in other clusters to a level that can lead to a conclusion of botnet existence. In this regard, Perdisci et al. [75] proposed an anomaly detection approach to track and detect malicious fast-flux service networks (FFSN). Their approach (see Fig. 7) focuses on passive analysis of recursive DNS (RDNS) traces obtained from large networks. The authors assume that botmasters usually operate malicious flux services using a number of fast-flux domain names that all point to flux agents related to the same flux service to evade domain blacklisting. This motivates them to group candidate flux domains based on the similarities in their resolved IP sets (i.e., the common set of resolved IP addresses) [75]. The author groups the similar domains using the single-linkage hierarchical clustering algorithm adopted from [76]. Not only the proposed approach is meant for the suspicious domain names from spam e-mails or domain blacklists, but also it can detect malicious flux services networks from different forms of spam behavior. However, the proposed approach is limited in its scope to botnets that generate spam e-mails and adopt the fast-flux technology to retrieve the proper C&C server location [6].

BotGAD is a Botnet Group Activity Detector that was proposed by Choi et al. [77]. BotGAD is based on monitoring group behaviors that appear in the DNS traffic of the monitored network. BotGAD extracted certain features from the monitored DNS traffic to distinguish between legitimate and illegitimate DNS queries that might be part of botnet traffic if it appears as a group of hosts showing the same behavior. For instance, bot tries to look up a C&C server or a victim address in a coordinated behavior. This behavior will appear as a group of hosts trying to look up certain domain names at different time intervals [77]. Botnets that have encrypted communication channels can be easily traced by gathering information from the IP header because IP header is the source to obtain the DNS information. However, the drawback of this approach is its incapability to identify botnets that employ a fast-flux algorithm or DGA which may yield high false-positive rates [28]. Moreover, the proposed approach does not scale well to large network with huge amount of traffic in terms of processing time.

Entropy-based botnet detection techniques In Huang et al. [78], the authors proposed the Spatial Snapshot Fast-Flux Detection system (SSFD). This solution (illustrated in Fig. 8) relied on spatial distribution estimation and spatial service relationship evaluation [78]. It uses botnet node time zones to distinguish between different geographic system spaces. The distinguished spaces are combined with information entropy to measure how the nodes are equivalently distributed within each time zone. The authors further noticed that benign domains tend to be distributed in the same time zone, while fast-flux nodes are widely distributed across multiple time zones. The authors further noted that if all the hosts of a botnet were to be located in the same time zone, time zone-based entropy would not be effective for the detection process if the hosts belonged to a benign or fast-flux domain. However, the SSFD is sometimes down because the requirement of the geographic information is not available [79]. Moreover, the SSFD system is efficient with dynamic DNS (DDNS) services [6].

Another study that adopts the entropy measures is that of Yadav and Reddy [80] who utilize the successfulness of a domain queries (NXDOMAIN) to identify the C&C server. The detection process employs temporal correlation and entropy information [81–83]. The authors deploy their approach off-line using a Tier-1 ISP dataset from Asia [80]. However, this approach needs access to DNS traffic with visibility over the IP addresses of the querying host [84]. Moreover, the approach needs the collection of precise time lines of the DNS queries deployed by all users [84].

Decision tree-based botnet detection techniques A framework for DNS-based detection and mitigation of malware infection on a network was proposed by Stalmans and Irwin [85]. Their IDS-based framework detects botnets based on malicious DNS queries. Their proposed work depends on certain features extracted from the DNS query responses such as A Record, NS record, Number IP ranges, Number ASNs, time to live (TTL) and some textual features [85]. They applied the C5.0 decision tree classifier and a Bayesian statistical approach at the core edge of the network to detect fast-flux domains. They labeled the traffic to be positive as malicious domains and negative as legitimate traffic. However, the proposed framework detects malicious domain names with the degree of accuracy at about 87 % [6].

Recently, Bilge et al. [20] proposed the EXPOSURE system that adopts large-scale passive DNS analysis techniques to detect malicious domain activities [20]. The authors used 15 features extracted from the DNS traffic to characterize different properties of DNS domain names and their query pattern. The EXPOSURE system assembles the features into the following group: DNS answer-based features, time-based features, domain name-based features and TTL value-based features. In addition, the EXPOSURE system has a data collector module to record the DNS traffic that is captured from the monitored network. It also has a feature attribution component to assign the recorded domains to the database along with its extracted features as illustrated in Fig. 9. Finally, the EXPOSURE system adopts a learning module to train the labeled set to construct a malicious domain detection models to be fed into a classifier to produce the final decisions. (i.e., label domains as being benign or malicious). The proposed classifier was built as a J48 decision tree algorithm (J48) which is an implementation of the C4.5 algorithm that is designed for generating a C4.5 decision trees. It uses the concept of information entropy to construct a decision tree

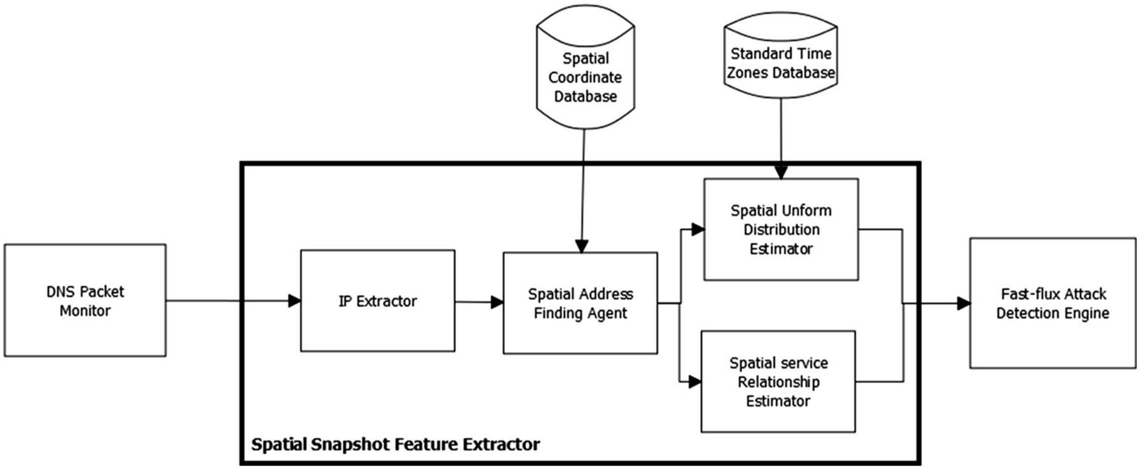

Spatial Coordinate Database
Standard Time Zones Database
DNS Packet Monitor
IP Extractor
Spatial Address Finding Agent
Spatial Unform Distribution Estimator
Spatial service Relationship Estimator
Spatial Snapshot Feature Extractor
Fast-flux Attack Detection Engine

Fig. 8 Spatial snapshot fast-flux detection system (SSFD) [78]

from a set of labeled training set. However, the EXPOSURE system was evaluated passively over recursive DNS (RDNS) traffic. Moreover, EXPOSURE system requires a huge number of RDNS sensors at different locations within the monitored network [73].

Neural networks-based botnet detection techniques Generally, neural networks-based approaches prove its efficiency in various domains and applications that may require parallel processing of information, classification of the information, adaptability to system dynamics and recognizing patterns of information with the existence of background noise [86]. Additionally, the neural networks (NN) applications proved to be successful in the field of intrusion detection [86]. Therefore, Wang et al. [87] proposed a behavior detection system to identify bots based on fuzzy pattern recognition techniques [87]. The proposed system uses the three steps illustrated in Fig. 10 to identify the malicious domain names and IP addresses from network traces. The authors developed a traffic reduction algorithm to reduce the amount of network traffic to be processed. From the reduced traffic, the authors extracted certain features related to bots' behavior such as failed DNS queries, similar DNS query intervals, failed network connections and frequently similar payload sizes for network connections. They also adopt a membership function that can be adapted it to the best values to enhance the capability of the proposed system. Finally, a pattern recognition technique was employed to identify bots through computing the probability of having bot-like activity from the reduced DNS and TCP traffic. However, the simple membership functions used in this work produce a high rate of false positives and low detection accuracy as it is not easy to adapt the membership function to real-world bot examples [88].

4 Summary and discussion

Our focus in this paper is to describe the botnet phenomena and botnet life cycle and to give a comprehensive view on the current state of the art in the field of botnet detection based on DNS traffic characteristics. Botnet is considered one of the highest destructive and prevalent attacks on the Internet. Botnets have many features which make it difficult to analyze and detect, because they are developed to be dynamic and flexible [89]. The current detection techniques are good but not enough to detect or mitigate the threat of botnet attacks. Even though the DNS-based botnet detection techniques show a promising direction toward mitigating the threat of botnets, the problem is still an open challenge. For instance, various methods being proposed with different algorithms that are used in diverse ways, there is no standard path or direction for these proposed algorithms to be used. Each serves special purpose from different angle. Moreover, these approaches were tested and evaluated using different datasets prepared or acquired from various sources. This is actually another issue of not having a standard dataset or test bed for botnet behavior analysis and techniques evaluations. Finally, yet there is no standard performance metrics to be used or adopted to evaluate the overall performances. What is suitable for one botnet or dataset may not be suitable for others. Therefore, this work comes to summarize the characteristics of the various methods and techniques and their methodologies, algorithms used and their limitations as a research direction in the field as summarized in Table 6. The study has a concise focus on DNS-based detection solutions or techniques (Table 6).

The honeydns methods such as the one proposed by Oberheide et al. [36] still suffer from the restrictions of

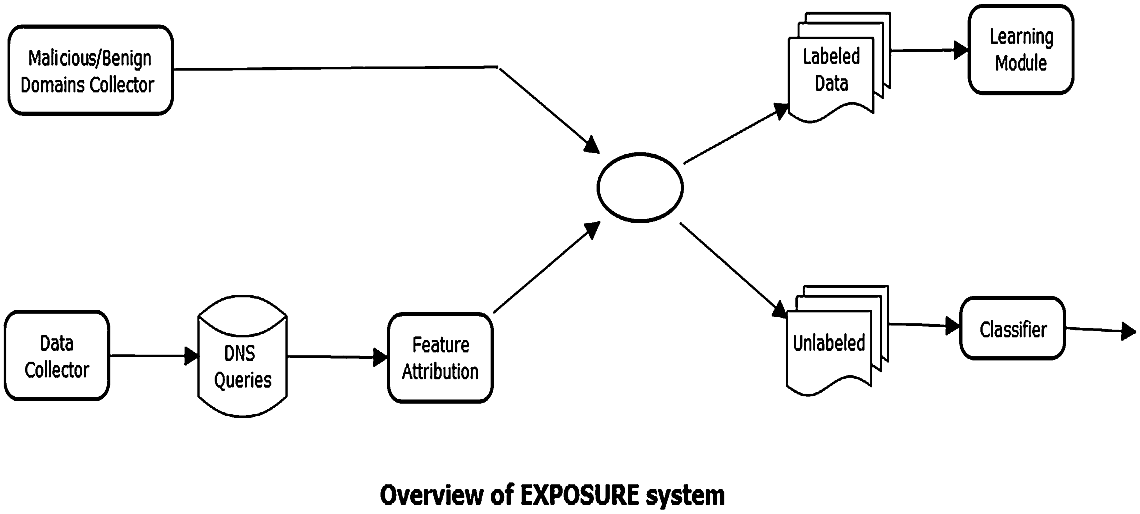


Fig. 9 Overview of EXPOSURE system

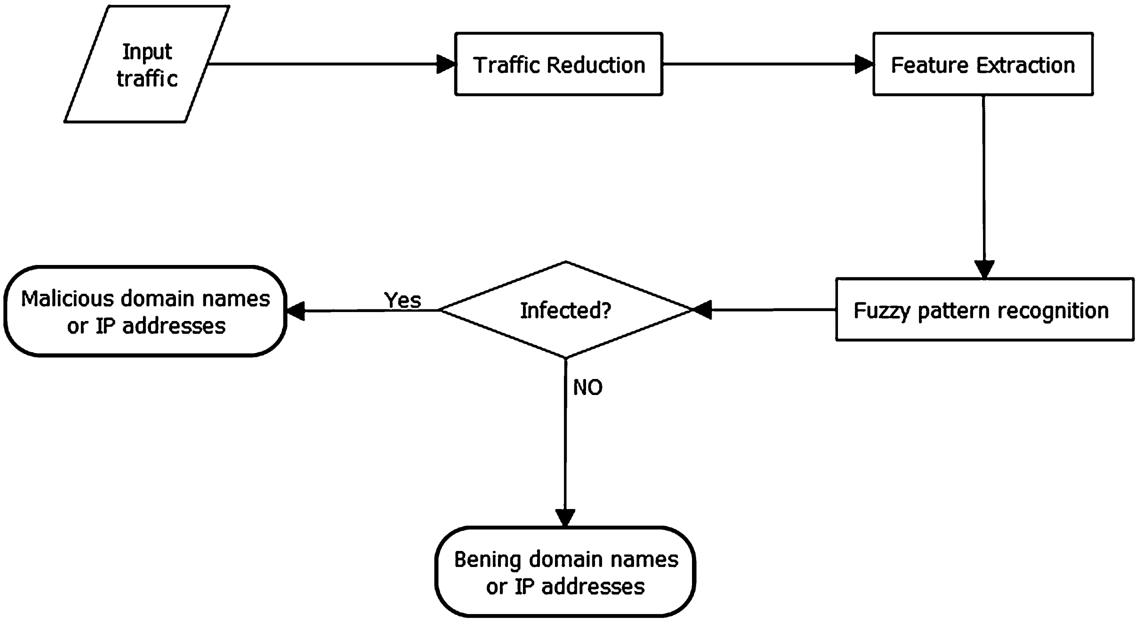

Input traffic
Traffic Reduction
Feature Extraction
Fuzzy pattern recognition
Infected?
Yes
Malicious domain names or IP addresses
NO
Bening domain names or IP addresses

Fig. 10 Fuzzy pattern recognition algorithm for identifying botnet domain names and IP addresses

deploying the system to certain places and period of times within the monitored network. In conclusion, the host-based botnet detection techniques are not efficient and limited in scope and view of the network, which leads to produce false positives and inaccurate results [36].

Recently, researchers worldwide focus on monitoring and detection techniques that are to be deployed at the edge of the monitored network to detect the attacks at an early stage before reaching the internal hosts. They also realize that adopting active techniques may expose their detection methodologies to the external world and get evaded. Therefore, adopting the passive techniques is more common and frequently used in network monitoring in general and botnet detection in particular. As stated before, different approaches have been proposed with different focus; for instance, many detection methods targeted singular bots or behavior of several bots in one botnet, while other methods targeted malicious commands and control servers (C&C). These kinds of approaches achieve the required objective, but not to the level that can lead to a mitigation solution of the botnet phenomena. On the other hand, some researchers start to observe the botnets as groups of bots that work collaboratively and synchronously. In fact, these collaborative or similar behaviors can help in identifying the malicious activities within the monitored network regardless of the botnet structure or type. For instance, BotCAD [77], EXPOSURE [20] and the work presented in [16] are some examples of the approaches that adopt the idea of the similar group behavior of botnet hosts. However, to the best of our knowledge, there are few works deployed to real-world environment in real-time mode. Therefore, the researches need to focus on different solutions to adapt the detection efficiency to the variations of botnets and the legitimate network traffic patterns. Considering the botnet life cycle, one can easily notice that botnets cannot thrive without the DNS services. Botnets still need to hide their C&C servers from being identified. Moreover, botnet writers start to learn the known detection methodologies and learn to evade them. For instance, botnet writers move the C&C server from having a static IP address to having a domain name that is mapped to various addresses. Moreover, to hinder the detection techniques that are aware of using domain names instead of static IP addresses, botmasters start to use the fast-flux services and the DGAs to generate a set of random domain name to add a sort of complexity to the detection mechanisms that are based on DNS traffic characteristics. Even though botmasters need to continuously avoid being detected and blacklisted through inventing different methods to evade the detection techniques, still they cannot thrive without the use of DNS service. Therefore, this study shed light on some existing techniques that targeting botnets through utilizing the fact that botnets cannot survive without DNS services. This in fact appears in the community to detect payload attack through analyzing the network traffic using anomaly detection techniques applied to application layer protocols such as DNS and HTTP [90].

An important work is needed in the future to define the requirements of timely detection or time efficiency, which implies that botnets have to be detected and mitigated at the boundary of the network or at least as early as possible. For instance, identifying botnets should start at the infection stage of its life cycle or at least before the command

Table 6 Summary of passive botnet detection techniques based on DNS

Authors	Passive technique	Mechanism	Weakness
Jiang et al. [74]	Graph theory	Identify malicious activities through DNS-based failure graph Employ graph decomposition algorithm based on a tri-nonnegative matrix factorization	Depend on unpredicted DNS traces
Perdisci et al. [75]	Clustering	Track and detect malicious FFSN Passive analysis of RDNS traces collected from large networks Applied in real-time environment	Slows down the network
Huang et al. [78]	Entropy	Spatial Snapshot Fast-Flux Detection system (SSFD) Detects the FFSNs during earning the geographic traffic patterns of network hosts	The SSFD cannot work for dynamic DNS (DDNS)
Stalmans and Irwin [85] (IDS-based framework)	Decision tree	Detect botnets based on malicious DNS queries Mitigate botnet activity to find malware infection on the network	Suffers from timely blacklist update problem
Leyla et al. [20] (EXPOSURE)	Decision tree	Passive DNS analysis to detect malicious domain activities	Rely on passive monitoring of recursive DNS traffic
		15 features extracted from DNS traffic to characterize different properties of DNS name	Need access to a very large number of RDNS sensors in many locations
Choi et al. [77] (BotGAD)	Clustering	Monitoring group behavior of DNS traffic Deploy small size of data from DNS traffic Relay on multiple observations of DNS traffic	High processing time
Wang et al. [87]	Neural networks	Detection is based on fuzzy pattern recognition techniques for malicious domain names and IP addresses used by botnets	High false positive

execution phase. Moreover, the performance of the detection approached can be affected by the selection of the right features (i.e., DNS-based features), so the type, quality and the optimization of the chosen features are considered important to the overall detection system accuracy which surely require particular attention in the future.

5 Conclusion

Botnet plays a key role as a major security threat to the Internet. The attackers may manipulate and take control of huge number of hosts to make illegal activities such as e-mail spam, DDoS attacks and click frauds. The botnet phenomena, architectures, life cycle and classification of botnet detection method based on DNS traffic analysis are discussed in this paper.

Nowadays, botnet has become more sophisticated and resistant to detection. There are a lot of detection techniques that have been proposed in the last decade. But the problem of botnets still emerges, and its threat is still rising. Therefore, in this work, a survey to improve the understanding of the botnet detection methodologies that are based on DNS traffic analysis is presented with a focus on the open issues within each category of the detection taxonomy provided earlier. The literature provides evidences to many cases where the existing approaches still have some limitations related to the accuracy and deployed location. Therefore, there is still a need for new approaches toward botnets detection. This paper shows that one of the main important factors of the botnet lifecycle is the DNS service that cannot avoid using it for stealthy purposes. Therefore, reviewing the existing techniques in botnet detection methodologies that are based on DNS traffic analysis may help the research community to produce better tools and techniques for mitigating the threat of botnets.

References

1. Stevanovic M, Revsbech K, Pedersen JM, Sharp R, Jensen CD (2012) A collaborative approach to botnet protection. In: Quirchmayr G, Basl J, You I, Xu L, Weippl E (eds) International Cross-Domain Conference and Workshop on Availability, Reliability, and Security (CD-ARES 2012), pp 624–638
2. Stevanovic M, Pedersen JM (2013) Machine learning for identifying botnet network traffic, Technical report, Aalborg University
3. Alomari E, Manickam S, Gupta B, Karuppayah S, Alfaris R (2012) Botnet-based distributed denial of service (DDoS) attacks on web servers: classification and art. arXiv preprint arXiv:12080403
4. Lu W, Rammidi G, Ghorbani AA (2011) Clustering botnet communication traffic based on *n*-gram feature selection. Comput Commun 34(3):502–514

5. McAfee. (2015) McAfee labs threats report. Accessed 18 May 2015. http://www.mcafee.com/us/resources/reports/rp-quarterly-threat-q4-2014.pdf
6. Karim A, Salleh RB, Shiraz M, Shah SAA, Awan I, Anuar NB (2014) Botnet detection techniques: review, future trends, and issues. J Zhejiang Univ Sci C 15(11):943–983
7. Yukonhiatou C, Kittitornkun S, Kikuchi H, Sisaat K, Terada M, Ishii H (2014) Temporal behaviors of Top-10 malware download in 2010–2012. In: 2014 International on electrical engineering congress (iEECON). IEEE, pp 1–4
8. Tiirmaa-Klaar H, Gassen J, Gerhards-Padilla E, Martini P (2013) Botnets: how to fight the ever-growing threat on a technical level. In: Tiirmaa-Klaar H et al. (eds) Botnets. Springer, London, pp 41–97
9. Harris KD, General A, Lookout A (2014) Cybersecurity in the Golden State. http://napi.net-flow.com/sananselmochamber.org/documents/CybersecurityReport.pdf
10. Botnets101- (2013) What they are and how to avoid them. http://www.fbi.gov/news/news_blog/botnets-101/
11. Emre Y (2011) A literature survey about recent botnet trends. http://geant3.archive.geant.net/Media_Centre/Media_Library/Media%20Library/botnet_trends_M2.pdf
12. Tiirmaa-Klaar H, Gassen J, Gerhards-Padilla E, Martini P (2013) Botnets, cybercrime and national security. In: Botnets, SpringerBriefs in Cybersecurity. Springer, London, pp 1–40
13. Shan G, Wang Y, Xie M, Lv H, Chi X (2014) Visual detection of anomalies in DNS query log data. In: 2014 IEEE Pacific visualization symposium (PacificVis). IEEE, pp 258–261
14. Davuth N, Kim S-R (2013) Classification of malicious domain names using support vector machine and Bi-gram method. Int J Secur Its Appl 7(1):51–58
15. He Y, Zhong Z, Krasser S, Tang Y (2010) Mining DNS for malicious domain registrations. In: 2010 6th International conference on collaborative computing: networking, applications and worksharing (CollaborateCom). IEEE, pp 1–6
16. Manasrah AM, Hasan A, Abouabdalla OA, Ramadass S (2009) Detecting botnet activities based on abnormal DNS traffic. arXiv preprint arXiv:09110487
17. Rodríguez-Gómez RA, Maciá-Fernández G, García-Teodoro P (2013) Survey and taxonomy of botnet research through life-cycle. ACM Comput Surv (CSUR) 45(4):45
18. Choi H, Lee H, Lee H, Kim H (2007) Botnet detection by monitoring group activities in DNS traffic. In: 7th IEEE international conference on computer and information technology, 2007 (CIT 2007). IEEE, pp 715–720
19. Bilge L, Sen S, Balzarotti D, Kirda E, Kruegel C (2014) EXPOSURE: a passive DNS analysis service to detect and report malicious domains. ACM Trans Inf Syst Secur (TISSEC) 16(4):14
20. Bilge L, Kirda E, Kruegel C, Balduzzi M (2011) EXPOSURE: finding malicious domains using passive DNS analysis. In: NDSS
21. ALmomani A, Gupta B, Wan T-C, Altaher A, Manickam S (2013) Phishing dynamic evolving neural fuzzy framework for online detection zero-day phishing email. arXiv preprint arXiv:13020629
22. Al-Mo AAD, Wan T-C, Al-Saedi K, Altaher A, Ramadass S, Manasrah A, Melhiml LB, Anbar M (2011) An online model on evolving phishing e-mail detection and classification method. J Appl Sci 11(18):3301–3307
23. Kirubavathi G, Anitha R (2014) Botnets: a study and analysis. In: Krishnan GSS, Anitha R, Lekshmi RS, Senthil Kumar M, Bonato A, Graña M (eds) Computational intelligence, cyber security and computational models. Springer, India, pp 203–214
24. Zeidanloo HR, Shooshtari MJZ, Amoli PV, Safari M, Zamani M (2010) A taxonomy of botnet detection techniques. In: 2010 3rd IEEE international conference on computer science and information technology (ICCSIT). IEEE, pp 158–162
25. Abu Rajab M, Zarfoss J, Monrose F, Terzis A (2006) A multifaceted approach to understanding the botnet phenomenon. In: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement. ACM, pp 41–52
26. Abdullah RS, Abdollah MF, Noh ZAM, Mas'ud MZ, Selamat SR, Yusof R, Melaka UTM (2013) Revealing the criterion on botnet detection technique. IJCSI Int J Comput Sci Issues 10(2):208–215
27. Liu L, Chen S, Yan G, Zhang Z (2008) Bottracer: execution-based bot-like malware detection. In: Wu T-C, Lei C-L, Rijmen V, Lee D-T (eds) Information security. Springer, Berlin, Heidelberg, pp 97–113
28. Feily M, Shahrestani A, Ramadass S (2009) A survey of botnet and botnet detection. In: Third international conference on emerging security information, systems and technologies, 2009 (SECURWARE'09). IEEE, pp 268–273
29. Jing L, Yang X, Kaveh G, Hongmei D, Jingyuan Z (2009) Botnet: classification, attacks, detection, tracing, and preventive measures. EURASIP journal on wireless communications and networking, IEEE Computer Society, Vol. 2009, pp 1184–1187
30. Khattak S, Ramay NR, Khan KR, Syed A, Khayam SA (2014) A taxonomy of botnet behavior, detection, and defense. In: Hossain E (ed) Communications surveys and tutorials, 16(2). IEEE, pp 898–924
31. Silva SS, Silva RM, Pinto RC, Salles RM (2013) Botnets: a survey. Comput Netw 57(2):378–403
32. Weimer F (2005) Passive DNS replication. In: FIRST conference on computer security incident, p 98
33. Zdrnja B, Brownlee N, Wessels D (2007) Passive monitoring of DNS anomalies. In: Sommer R, Hammerli B (eds) Detection of intrusions and malware, and vulnerability assessment. Springer, Berlin, Heidelberg, pp 129–139
34. Janbeglou M, Naderi H, Brownlee N (2014) Effectiveness of DNS-based security approaches in large-scale networks. In: 2014 28th International conference on advanced information networking and applications workshops (WAINA). IEEE, pp 524–529
35. Dagon D, Zou CC, Lee W (2006) Modeling botnet propagation using time zones. In: NDSS, pp 2–13
36. Oberheide J, Karir M, Mao ZM (2007) Characterizing dark DNS behavior. In: Hämmerli BM, Sommer R (eds) Detection of intrusions and malware, and vulnerability assessment. Springer, Berlin, Heidelberg, pp 140–156
37. Li Z, Goyal A, Chen Y, Paxson V (2009) Automating analysis of large-scale botnet probing events. In: Proceedings of the 4th international symposium on information, computer, and communications security. ACM, pp 11–22
38. Rieck K, Schwenk G, Limmer T, Holz T, Laskov P (2010) Botzilla: detecting the phoning home of malicious software. In: Proceedings of the 2010 ACM symposium on applied computing. ACM, pp 1978–1984
39. Pham V-H, Dacier M (2011) Honeypot trace forensics: the observation viewpoint matters. Future Gen Comput Syst 27(5):539–546
40. Aiello M, Mongelli M, Papaleo G (2014) DNS tunneling detection through statistical fingerprints of protocol messages and machine learning. Int J Commun Syst 28(14):1987–2002
41. Aiello M, Mongelli M, Papaleo G (2014) Supervised learning approaches with majority voting for DNS tunneling detection. In: International joint conference SOCO'14–CISIS'14–ICEUTE'14. Springer, Berlin, pp 463–472
42. Panimalar P, Rameshkumar K (2014) A review on taxonomy of botnet detection. In: 2014 International conference on advances in engineering and technology (ICAET). IEEE, pp 1–4

43. Li C, Jiang W, Zou X (2009) Botnet: survey and case study. In: 2009 Fourth International Conference on Innovative computing, information and control (ICICIC). IEEE, pp 1184–1187
44. Vania J, Meniya A, Jethva H (2013) A review on botnet and detection technique. Int J Comput Trends Technol 4(1):23–29
45. Gu G, Porras PA, Yegneswaran V, Fong MW, Lee W (2007) BotHunter: detecting malware infection through IDS-driven dialog correlation. In: Usenix security, pp 1–16
46. Nechaev B, Gurtov A (2013) Classification of botnet detection techniques. Helsinki Institute for Information Technology HIIT
47. SNORT. www.snort.org
48. Ramachandran A, Feamster N, Dagon D (2006) Revealing botnet membership using DNSBL counter-intelligence. In: Proceedings of the 2nd USENIX steps to reducing unwanted traffic on the Internet, pp 49–54
49. Oro D, Luna J, Felguera T, Vilanova M, Serna J (2010) Benchmarking IP blacklists for financial botnet detection. In: 2010 Sixth international conference on information assurance and security (IAS). IEEE, pp 62–67
50. Sinha S, Bailey M, Jahanian F (2008) Shades of grey: on the effectiveness of reputation-based "blacklists". In: 3rd International conference on malicious and unwanted software, 2008 (MALWARE 2008), pp 57–64. doi:10.1109/MALWARE.2008.4690858
51. Antonakakis M, Perdisci R, Dagon D, Lee W, Feamster N (2010) Building a dynamic reputation system for DNS. In: USENIX security symposium, pp 273–290
52. Kheir N, Tran F, Caron P, Deschamps N (2014) Mentor: positive DNS reputation to skim-off benign domains in botnet C&C blacklists. In: Cuppens-Boulahia N, Cuppens F, Jajodia S, El Kalam AA, Sans T (eds) ICT systems security and privacy protection. Springer, Berlin, Heidelberg, pp 1–14
53. Yadav S, Reddy AKK, Reddy A, Ranjan S (2010) Detecting algorithmically generated malicious domain names. In: Proceedings of the 10th ACM SIGCOMM conference on Internet measurement. ACM, pp 48–61
54. Stinson E, Mitchell JC (2007) Characterizing bots' remote control behavior. In: Hämmerli BM, Sommer R (eds) Detection of intrusions and malware, and vulnerability assessment. Springer, Berlin, Heidelberg, pp 89–108
55. Shin S, Xu Z, Gu G (2012) EFFORT: efficient and effective bot malware detection. In: 2012 Proceedings IEEE INFOCOM. IEEE, pp 2846–2850
56. Rahim A, Bin Muhaya FT (2010) Discovering the botnet detection techniques. In: Kim T-H, Fang W-C, Khurram Khan M, Arnett KP, Kang H-J, Ślęzak D (eds) Security technology, disaster recovery and business continuity. Springer, Berlin, Heidelberg, pp 231–235
57. Raghava NS, Sahgal D, Chandna S (2012) Classification of botnet detection based on botnet architecture. In: 2012 International conference on communication systems and network technologies (CSNT), pp 569–572. doi:10.1109/csnt.2012.128
58. Gu G, Yegneswaran V, Porras P, Stoll J, Lee W (2009) Active botnet probing to identify obscure command and control channels. In: Annual computer security applications conference, 2009 (ACSAC'09). IEEE, pp 241–253
59. Strayer WT, Lapsely D, Walsh R, Livadas C (2008) Botnet detection based on network behavior. In: Lee W, Wang C, Dagon D (eds) Botnet detection. Springer, USA, pp 1–24
60. Ma X, Zhang J, Li Z, Li J, Tao J, Guan X, Lui JC, Towsley D (2015) Accurate DNS query characteristics estimation via active probing. J Netw Comput Appl 47:72–84
61. Ma J, Saul LK, Savage S, Voelker GM (2009) Beyond blacklists: learning to detect malicious web sites from suspicious URLs. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 1245–1254
62. Holz T, Gorecki C, Rieck K, Freiling FC (2008) Measuring and detecting fast-flux service networks. In: NDSS
63. Villamarín-Salomón R, Brustoloni JC (2008) Identifying botnets using anomaly detection techniques applied to DNS traffic. In: 5th IEEE consumer communications and networking conference, 2008 (CCNC 2008). IEEE, pp 476–481
64. Cranor CD, Gansner E, Krishnamurthy B, Spatscheck O (2001) Characterizing large DNS traces using graphs. In: Proceedings of the 1st ACM SIGCOMM workshop on internet measurement. ACM, pp 55–67
65. Wills CE, Mikhailov M, Shang H (2003) Inferring relative popularity of internet applications by actively querying DNS caches. In: Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement. ACM, pp 78–90
66. Gardiner J, Cova M, Nagaraja S (2014) Command & control: understanding, denying and detecting. arXiv preprint arXiv:14081136
67. Qi C, Chen X, Xu C, Shi J, Liu P (2013) A bigram based real time DNS tunnel detection approach. Procedia Comput Sci 17:852–860
68. Kang BBH (2011) DNS-based botnet detection. In: Encyclopedia of cryptography and security. Springer, USA, pp 362–363
69. Marko P, Vilhan P (2012) Efficient detection of malicious nodes based on DNS and statistical methods. In: 2012 IEEE 10th international symposium on applied machine intelligence and informatics (SAMI). IEEE, pp 227–230
70. Hu X, Knysz M, Shin KG (2009) RB-Seeker: auto-detection of redirection botnets. In: NDSS
71. Choi H, Lee H (2012) Identifying botnets by capturing group activities in DNS traffic. Comput Netw 56(1):20–33
72. Sanchez F, Duan Z, Dong Y (2012) Blocking spam by separating end-user machines from legitimate mail server machines. Secur Commun Netw. doi:10.1002/sec.587
73. Antonakakis M, Perdisci R, Lee W, Vasiloglou N II, Dagon D (2011) Detecting malware domains at the upper DNS hierarchy. In: USENIX security symposium, p 16
74. Jiang N, Cao J, Jin Y, Li L, Zhang Z-L (2010) Identifying suspicious activities through DNS failure graph analysis. In: 2010 18th IEEE international conference on network protocols (ICNP). IEEE, pp 144–153
75. Perdisci R, Corona I, Dagon D, Lee W (2009) Detecting malicious flux service networks through passive analysis of recursive DNS traces. In: Annual computer security applications conference, 2009 (ACSAC'09). IEEE, pp 311–320
76. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. ACM Comput Surv (CSUR) 31(3):264–323
77. Choi H, Lee H, Kim H (2009) BotGAD: detecting botnets by capturing group activities in network traffic. In: Proceedings of the fourth international ICST conference on COMmunication system softWAre and middlewaRE. ACM, p 2
78. Huang S-Y, Mao C-H, Lee H-M (2010) Fast-flux service network detection based on spatial snapshot mechanism for delay-free detection. In: Proceedings of the 5th ACM symposium on information, computer and communications security. ACM, pp 101–111
79. Lin H-T, Lin Y-Y, Chiang J-W (2013) Genetic-based real-time fast-flux service networks detection. Comput Netw 57(2):501–513
80. Yadav S, Reddy AN (2012) Winning with DNS failures: strategies for faster botnet detection. In: Rajarajan M, Piper F, Wang H, Kesidis G (eds) Security and privacy in communication networks. Springer, Berlin, Heidelberg, pp 446–459
81. Sharifnya R, Abadi M (2015) DFBotKiller: domain-flux botnet detection based on the history of group activities and failures in DNS traffic. Digit Investig 12:15–26
82. Zhang Y, Zhang Y, Xiao J (2014) Detecting the DGA-based malicious domain names. In: Yuan Y, Wu X, Lu Y (eds)

Trustworthy computing and services. Springer, Berlin, Heidelberg, pp 130–137

83. Manadhata PK, Yadav S, Rao P, Horne W (2014) Detecting malicious domains via graph inference. In: Kutyłowski M, Vaidya J (eds) Computer security-ESORICS 2014. Springer International Publishing, pp 1–18
84. Schiavoni S (2013) Finding, characterizing and tracking domain generation algorithms from passive DNS monitoring. http://hdl.handle.net/10589/78505
85. Stalmans E, Irwin B (2011) A framework for DNS based detection and mitigation of malware infections on a network. In: 2011 Information security South Africa (ISSA). IEEE, pp 1–8
86. Nogueira A, Salvador P, Blessa F (2010) A botnet detection system based on neural networks. In: 2010 Fifth international conference on digital telecommunications (ICDT). IEEE, pp 57–62
87. Wang K, Huang C-Y, Lin S-J, Lin Y-D (2011) A fuzzy pattern-based filtering algorithm for botnet detection. Comput Netw 55(15):3275–3286
88. Wang K, Huang CY, Tsai LY, Lin YD (2014) Behavior-based botnet detection in parallel. Secur Commun Netw 7(11):1849–1859
89. Eslahi M, Salleh R, Anuar NB (2012) Bots and botnets: an overview of characteristics, detection and challenges. In: 2012 IEEE international conference on control system, computing and engineering (ICCSCE). IEEE, pp 349–354
90. Davis JJ, Clark AJ (2011) Data preprocessing for anomaly based network intrusion detection: a review. Comput Secur 30(6):353–375