A Survey of Botnet and Botnet Detection Methods

Mr. Sandip Sonawane
Assistant Professor, Department of Computer Engineering
R. C. Patel Institute of Technology, Shirpur

Abstract—Recent malicious attempts are intended to get financial benefits through a large pool of compromised hosts, which are called software robots or simply “bots.” A group of bots, referred to as a botnet, is remotely controllable by a server and can be used for sending spam mails, stealing personal information, and launching DDoS attacks. Among the various forms of malware, botnets are emerging as the most serious threat against cyber-security as they provide a distributed platform for several illegal activities such as launching distributed denial of service attacks against critical targets, malware dissemination, phishing, and click fraud. The defining characteristic of botnets is the use of command and control channels through which they can be updated and directed. Recently, botnet detection has been an interesting research topic related to cyber-threat and cyber-crime prevention. This paper is a survey of botnet and botnet detection. The survey clarifies botnet phenomenon and discusses botnet detection techniques. This survey classifies botnet detection techniques into four classes: signature-based, anomaly-based, DNS-based, and mining-base. It summarizes botnet detection techniques in each class and provides a brief comparison of botnet detection techniques. Our anomaly-based botnet detection mechanism is more robust than the other approaches so that the variants of bots can be detectable by looking at their group activities in DNS traffic.

Keywords: Botnet; Botnet Detection; Cyber-security

I. INTRODUCTION

The term “Bot” is derived from the word “Robot”; and similar to robots, bots are designed to perform some predefined functions in automated way. In other words, the individual bots are software programs that run on a host computer allowing the botmaster to control host actions remotely. A large pool of compromised hosts, called bots, communicate with a bot controller to coordinate the network of bots. Such a network is commonly referred to as a botnet. An attacker, called a botmaster, controls a botnet to perform various malicious activities. Malicious botnet is a network of compromised computers called “Bots” under the remote control of a human operator called “Botmaster”. Botnets pose a significant and growing threat against cyber-security as they provide a distributed platform for many cyber-crimes such as sending spam mails, theft of personal data, Distributed Denial of Service (DDoS) attacks against critical targets, malware dissemination, phishing, and click fraud. Botnet detection has been a major research topic in recent years. Researchers have proposed several approaches for botnet detection to combat botnet threat against cyber-security. In this survey, botnet phenomenon will be clarified and advances in botnet detection techniques will be discussed.

Two ways of detecting botnet are, first setting up Honeypots and Second Monitoring Traffic. This survey focuses on second way of detecting botnet traffic i.e. using Traffic Monitoring. This survey classifies botnet detection approaches into four classes: signature-based, anomaly-based, DNS-based, and mining-based. Furthermore, it summarizes botnet detection techniques in each class and provides a brief comparison of these techniques. The remainder of the paper is organized as follows: Section II describes botnet phenomenon. In this section, botnet characteristics and botnet life-cycle are explained to provide better understanding of botnet technology. Section III discusses botnet detection and tracking. In this section four classes of botnet detection approaches including signature-based, anomaly-based, DNS-based, and mining-based are discussed respectively. Section IV provides a brief comparison of botnet detection techniques. The survey concludes in Section V.

II. BOTNET PHENOMENON

Botnets are emerging as the most significant threat facing online ecosystems and computing assets. Malicious botnets are distributed computing platforms predominantly used for illegal activities such as launching Distributed Denial of Service (DDoS) attacks, sending spam, trojan and phishing emails, illegally distributing pirated media and software, force distribution, stealing information and computing resource, e-bussiness extortion, performing click fraud, and identity theft.

The high light value of botnets is the ability to provide anonymity through the use of a multi-tier command and control (C&C) architecture. Moreover, the individual bots are not physically owned by the botmaster, and may be located in several locations spanning the globe. Differences in time zones, languages, and laws make it difficult to track malicious botnet activities across international boundaries. This characteristic makes botnet an attractive tool for cyber-criminals, and in fact poses a great threat against cyber-security. In order to provide better understanding of botnet phenomenon, botnet characteristics and botnet life-cycle will be explained respectively.

A. Botnet Characteristics

Like the previous generations of viruses and worms, a bot is a self-propagating application that infects vulnerable hosts through exploit activities to expand their reach. Bot infection methods are similar to other classes of malware that recruit vulnerable systems by exploiting software vulnerabilities, trojan insertion, as well as social engineering techniques leading to download malicious bot
code. According to measurement studies bots are equipped with several exploit vectors to improve opportunities for exploitation. However, among the other classes of malware, the defining characteristic of botnets is the use of command and control (C&C) channels through which they can be updated and directed. The multi-tier C&C architecture of botnets provides anonymity for the botmaster. C&C channels can operate over a wide range of logical network topologies and use different communication protocols. Botnets are usually classified according to their command and control architecture.

According to their command and control architecture, botnets can be classified as IRC-based, HTTP-based, DNS-based or Peer to Peer (P2P) botnets. P2P botnets use the recent P2P protocol to avoid single point of failure. Moreover, P2P botnets are harder to locate, shutdown, monitor, and hijack. However, according to the analysis in [2] the most prevalent botnets are based on Internet Relay Chat (IRC) protocol with a centralized command and control mechanism. IRC protocol was originally designed for large social chat rooms to allow for several forms of communication and data dissemination among a large number of end-hosts. The great prevalence of IRC-based botnets is due to the inherent flexibility and scalability of this protocol. Furthermore, there are several open-source implementations that enable botmasters to extend them according to their demands.

B. Botnet Life-cycle

A typical botnet can be created and maintained in five phases including: initial infection, secondary injection, connection, malicious command and control, update and maintenance. This life-cycle is depicted in Fig. 1.

Despite the long presence of malicious botnets, only few formal studies have examined the botnet problem. To date, just very little is known about botnet malicious behavior. The Honeynet project was one of the pioneering informal studies of the botnet problem but using or setting up honeypot is suitable for understanding only botnet characteristics and not for detecting botnet. However, efforts are in progress to quantify the botnet problem, detect the presence of botnets, and design defenses against attacks by botnets.

Botnet detection and tracking has been a major research topic in recent years. Different solutions have been proposed in academia. There are mainly two approaches for botnet detection and tracking. One approach is based on setting up honeypots. For instance, solutions have been initial honeynet-based solutions. In addition, many papers discussed how to use honeypots for botnet tracking and measurement. However, honeypots are mostly useful to understand botnet technology and characteristics, but do not necessarily detect botnet.

The other approach for botnet detection is based on passive network traffic monitoring and analysis. Botnet detection techniques based on passive traffic monitoring
have been useful to identify the existence of botnets. These techniques can be classified as being signature-based, anomaly-based, DNS-based, and mining-based that will be described and summarized in this section respectively.

A. Signature-based Detection
Knowledge of useful signatures and behavior of existing botnets is useful for botnet detection. For example, Snort is an open source intrusion detection system (IDS) that monitors network traffic to find signs of intrusion. Like most IDS systems, Snort is configured with a set of rules or signatures to log traffic which is deemed suspicious. However, signature-based detection techniques can be used for detection of known botnets. Thus, this solution is not useful for unknown bots.

B. Anomaly-based Detection
Anomaly-based Detection base on finding similar communication patterns and behaviors among the group of hosts that are performing at least one malicious activity. The point that distinguishes our proposed detection framework from many other similar works is that there is no need for prior knowledge of Botnets such as Botnet signature. Anomaly-based detection techniques attempt to detect botnets based on several network traffic anomalies such as high network latency, high volumes of traffic, traffic on unusual ports, and unusual system behavior that could indicate presence of malicious bots in the network. Although anomaly detection techniques solve the problem of detecting unknown botnets, problems with anomaly detection can include detection of an IRC network that may be a botnet but has not been used yet for attacks, hence there are no anomalies. To solve this, Binkley and Singh proposed an effective algorithm that combines TCP-based anomaly detection with IRC tokenization and IRC message statistics to create a system that can clearly detect client botnets. This algorithm can also reveal bot servers. However, Binkley’s approach could be easily defeated by simply using a trivial cipher to encode the IRC commands.

In 2007, Karasaridis et al. presented an algorithm for detection and characterization of botnets using passive analysis based on flow data in transport layer. This algorithm can detect encrypted botnet communications. It helps to quantify size of botnets, identify and characterize their activities without joining the botnet. Recently, Gu et al. have proposed Botsniffer that uses network-based anomaly detection to identify botnet C&C channels in a local area network. Botsniffer is based on observation that bots within the same botnet will likely demonstrate very strong synchronization in their responses and activities. Hence, it employs several correlation analysis algorithms to detect patial-temporal correlation in network traffic with a very low false positive rate.

C. DNS-based Detection
DNS-based detection techniques are based on particular DNS information generated by a botnet. DNS-based detection techniques are similar to anomaly detection techniques as similar anomaly detection algorithms are applied on DNS traffic. As mentioned in Section II, bots typically initiate connection with C&C server to get commands. In order to access the C&C server bots perform DNS queries to locate the respective C&C server that is typically hosted by a DDNS provider. Thus, it is possible to detect botnet DNS traffic by DNS monitoring and detect DNS traffic anomalies.

In 2005, Dagon proposed a mechanism to identify botnet C&C servers by detecting domain names with abnormally high or temporally concentrated DDNS query rates. This technique is similar to the approach proposed by Kristoff in 2004. However, both techniques have the same weakness and could easily be evaded by using faked DNS queries. Furthermore, according to the evaluation in, this technique generates many false positives due to misclassification of legitimate and popular domains that use DNS with short time-to-live (TTL).

An alternative approach was proposed by Schonewille and Van Helmond in 2006. This approach was based on abnormally recurring NXDOMAIN reply rates. In order to classify anomalous reply rates, they use the algorithms similar to those Dagon used for classifying analogous query rates.

According to their observation DDNS responses indicating name error (NXDOMAIN) often correspond to botnet C&C servers that have been shut down by authorities. Hosts that repeatedly issue such queries may be infected with a bot and they may have the vulnerability to enable similar infection. According to, this approach is very effective to detect several suspicious domain names and there may be less false positive because NXDOMAIN replies are more likely to refer to DDNS than to other names. Ramachandran et al. proposed a set of techniques and heuristics to identify botnets using passive analysis of DNS-based Black-hole List (DNSBL) lookup traffic. This technique addresses the possibility of performing counter-intelligence that help us to detect DNSBL reconnaissance activity, whereby botmasters themselves must perform lookups against the DNSBL to determine their bots’ blacklist status. The goal in developing these models and heuristics is to distinguish DNSBL queries issued by botmasters from legitimate DNSBL traffic to identify likely bots. These heuristics could be used to detect reconnaissance activities in real-time and allows for active countermeasures. As botmasters usually perform reconnaissance lookups prior to the use of bots in an attack, this DNSBL counter-intelligence can be used for early warning to boost responses. Moreover, this detection technique does not require direct communication with any component of the botnet, and does not disrupt the botnet’s activity.

In 2007, Choi et al. proposed an anomaly-based botnet detection mechanism by monitoring group activities in DNS traffic, which form a group activity in DNS queries simultaneously sent by distributed bots. They have defined unique features of DNS traffic as group activity to distinguish botnet DNS queries from legitimate DNS queries. Since DNS traffic appears in several stages of botnet life-cycle, it is possible to detect botnet by using the group activity property of botnet DNS traffic. They also developed a mechanism that enables to detect C&C server.
migration. This anomaly-based botnet detection mechanism is more robust than the previous approaches and can detect botnet regardless of the type of bot and botnet by looking at their group activities in DNS traffic. Furthermore, it can detect botnets with encrypted channels since it uses the information of IP headers. Nevertheless, the main drawback of this approach is the high processing time required for monitoring huge scale of network.

Other proposals are also based on DNS monitoring. However, they are mostly useful for botnet tracking and measurement to understand botnet technology and characteristics, but do not necessarily detect bot infection.

D. Mining-based Detection
One effective technique for botnet detection is to identify botnet C&C traffic. However, botnet C&C traffic is difficult to detect. In fact, since botnets utilize normal protocols for C&C communications, the traffic is similar to normal traffic. Moreover, the C&C traffic is not high volume and does not cause high network latency. Therefore, anomaly-based techniques are not useful to identify botnet C&C traffic. Several data mining techniques including machine learning, classification, and clustering can be used efficiently to detect botnet C&C traffic.

Geobl and Holz proposed Rishi in 2007. Rishi is mainly based on passive traffic monitoring for unusual or suspicious IRC nicknames, IRC servers, and uncommon server ports. They use n-gram analysis and a scoring system to detect bots that use uncommon communication channels, which are commonly not detected by classical intrusion detection systems. However, this approach is quite limited, in that IRC nickname can be changed to resemble normal host. In addition, this method cannot detect encrypted communication as well as non-IRC botnets.

Masud et al. proposed robust and effective flow-based botnet traffic detection by mining multiple log files. They introduce multiple log correlation for C&C traffic detection. They classify an entire flow to identify botnet C&C traffic. This method does not impose any restriction on the botnet communication protocol and is therefore applicable to non-IRC botnets. Furthermore, this method does not require access to payload content. Hence, it is effective even if the C&C payload is encrypted or is not available.

Botmaster is the most recent approach which applies data mining techniques for botnet C&C traffic detection. Botminer is an improvement of Botsniffer. It clusters similar communication traffic and similar malicious traffic. Then, it performs cross cluster correlation to identify the hosts that share both similar communication patterns and similar malicious activity patterns. Botminer is an advanced botnet detection tool which is independent of botnet protocol and structure. Botminer can detect real-world botnets including IRC-based, HTTP-based, and P2P botnets with a very low false positive rate.

IV. COMPARISON OF BOTNET DETECTION TECHNIQUES

This section provides a brief comparison of botnet detection techniques. We have compared botnet detection approaches based on key features including: ability to detect unknown bots, capability of botnet detection regardless of botnet protocol and structure, and botnets with encrypted C&C channels, real-time detection, and accuracy. This comparison is summarized in Table 1.

<table>
<thead>
<tr>
<th>Botnet Detection Techniques</th>
<th>Unknown Botnet</th>
<th>Protocol &amp; Structure Independent</th>
<th>Encrypted Botnet Detection</th>
<th>Real-Time Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature Based</td>
<td>[24] X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Anomaly Based</td>
<td>[25] X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DNS Based</td>
<td>[27] X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Mining Based</td>
<td>[31] X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

As shown in this table, signature-based techniques can only detect known botnets, whereas the other classes are able to detect unknown bots. However, there are few botnet detection techniques [15, 33, and 34] that can detect botnet regardless of botnet protocol and structure. These techniques will be effective even though botmasters change their C&C communication protocol and structure. On the other hand, detection techniques that require access to C&C payloads are less effective as botmasters tend to use encrypted channels for C&C communications. Among all detection techniques, the only approach that allows real-time detection is a DNS-based detection which uses DNSBL counter-intelligence to detect reconnaissance in real-time. However, active countermeasures run the risk of false positives. The most recent botnet detection techniques based on data mining as well as DNS-based botnet detection approach in provide promising tradeoff. These methods are independent of botnet protocol and structure. Moreover, they are effective to detect encrypted C&C botnet communication. In overall, these techniques can detect real world botnets regardless of botnet protocol and structure with a very low false positive rate.

V. CONCLUSION
Botnets pose a significant and growing threat against cyber-security as they provide a key platform for many cyber crimes such as Distributed Denial of Service (DDoS) attacks against critical targets, malware dissemination, phishing, and click fraud. Despite the long presence of malicious botnets, in comparison to other kind of malware...
Botnets are harder to monitor and shutdown and detection of them become a challenging problem. Only few formal studies have examined the botnet problem and botnet research is still in its infancy. This paper surveys botnet and botnet detection.

As mentioned in this paper, the high light value of botnets is still in its infancy. This paper surveys botnet studies have examined the botnet problem and botnet of them become a challenging problem. Only few formal botnets are harder to monitor and shutdown and detection techniques work only on specific botnet C&C communication protocols and structures. Consequently, as botnets change their C&C communication architecture, these methods will be ineffective.

REFERENCES