Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jisa

Detecting botnet by anomalous traffic☆

Chia-Mei Chen[*], *Hsiao-Chung Lin*

Department of Information Management, National Sun Yat-sen University, Kaohsiung 804, Taiwan

ARTICLE INFO

Article history:
Available online 11 June 2014

Keywords:
Botnet detection
Intrusion detection
IRC

ABSTRACT

Botnets can cause significant security threat and huge loss to organizations, and are difficult to discover their existence. Therefore they have become one of the most severe threats on the Internet. The core component of botnets is their command and control channel. Botnets often use IRC (Internet Relay Chat) as a communication channel through which the botmaster can control the bots to launch attacks or propagate more infections. In this paper, anomaly score based botnet detection is proposed to identify the botnet activities by using the similarity measurement and the periodic characteristics of botnets. To improve the detection rate, the proposed system employs two-level correlation relating the set of hosts with same anomaly behaviors. The proposed method can differentiate the malicious network traffic generated by infected hosts (bots) from that by normal IRC clients, even in a network with only a very small number of bots. The experiment results show that, regardless the size of the botnet in a network, the proposed approach efficiently detects abnormal IRC traffic and identifies botnet activities.

1. Introduction

Botnets have become a common channel for developing cybercrimes. The growth of botnets has attracted a lot of attention on the security research and research community. According to the research reports (McAfee, 2012; Trend Micro, 2009), botnets have played a big dangerous threat to the Internet, responsible for various malicious activities from distributed denial of service (DDoS) to spamming, phishing, information harvesting, and identity theft. The survey paper (Freily et al., 2009) also highlights that the most prevalent botnets are IRC-based.

A typical botnet consists of four phases: infection, command and control connection, attack and post-attack (Leonard et al., 2009). Hacker may apply various strategies to explore and infect machines, such as malicious web pages, spam mails, viruses, or worms. Botnet is a collection of infected hosts (bots) and is controlled remotely by a botmaster through a command and control (C&C) channel. Utilizing botnet, botmaster can launch attacks or recruit more members. Bots update their functions and binaries in the post-attack phase. Unlike virus or worm, botnet is equipped with a form of communication for the botmaster to maintain complete control of the botnet. To prevent further damage, this study intends to propose a detection scheme to identify botnets during the control and command connection stage, before the botnet attack.

According to Cooke et al. (2005), the control mechanism of botnet can be classified into centralized and Peer-to-Peer (P2P). A centralized topology, including IRC (Internet Relay Chat) and HTTP, has a central point for the botmaster forwarding commands and messages to the bots, while its weakness is the single point of failure. P2P topology employs P2P protocol as the C&C channel overcoming the former weakness. The majority of botnets use IRC as a control

☆ This article belongs to the special issue Security, Privacy and Trust in Future Networks and Mobile Computing.

[*] *Corresponding author.*
E-mail address: cchen@mail.nsysu.edu.tw (C.-M. Chen).

http://dx.doi.org/10.1016/j.jisa.2014.05.002

channel, because IRC is a form of real-time communication vehicle and provides an efficient and stealthy channel for botmaster to control bots. HKCERT (2013) security report indicated that half of C&C servers found are IRC-based (HKCERT, 2013) and TACERT found 20% of all the incidents are IRC-based in the first three quarter of 2012 (TACERT, 2012). Even web-based botnets rely on IRC botnets heavily to recruit new members (Trustwave, 2013). In addition, with the increase in Bitcoin value (peaking at 1200 USD), IRC-based botnets increase for Bitcoin mining (Dell, 2013). IRC-based botnets are still prevailing. Therefore, this paper focuses on IRC-based botnet problem. An anomaly score based botnet detection approach is proposed to monitor and analyze network traffic and then to identify suspicious bot machines.

Akiyama et al. (2007) propose three metrics for determining the botnet behaviors: relationship, response, and synchronization. The relationship presents the connection between botmaster and bots over one protocol, such as IRC, HTTP, or P2P. The response means that bots respond immediately and accurately after they receive commands from the botmaster. The synchronization means bots simultaneously carry out programmed activities, such as DDoS attack, reporting their status, or sharing information, based on the botmaster's commands. Even though most researches identify botnets in the attack phase, as attack activities from a group of bot machines exhibit synchronized actions. However, this work also implies that botnet has similar network behaviors during command and control connection stage and might be able to identify before the botnet attack.

From our observation, there are the following unique characteristics of the IRC-based botnet traffic: abnormal direction of PING-PONG messages, homogeneous response, and group activity. As bots are programs, not interactively with human beings, the IRC server will send “PING” to check if it is alive (Auzzie, 2009; Bashcripts.org 2011; Karasaridis et al., 2007; Shake, 2007; Stack Overflow, 2012). On the contrary, a normal IRC client sends a “PING” message to IRC server to avoid idling long time and the server replies a “PONG” message. Bots respond similar messages immediately after they receive a command from their botmaster, which exhibit homogeneous response (Akiyama et al., 2007; Gu et al., 2008a; Yen and Reiter, 2008). Over a long duration, those hosts (bots) collectively carry out the action as the botmaster commands. For example, bots simultaneously report their status and system information, carry on an attack, or receive a command. As PING-PONG anomaly might be easily rewritten, the proposed detection does not take this into account. Only the anomaly on homogeneous response and group activity is used to identify anomalous hosts when they communicate with botmaster in communication stage. In this paper, the proposed detection attempts to identify suspicious bot machines in the network before they launch attacks.

The real network traffic, observed in our campus network for over three years, shows that the number of infected bot machines in a class B network might be small comparing with that of normal one. The proposed system is able to identify the anomalous traffic over a large amount of normal network data during the communication stage with the botnet server without a priori knowledge of the botnets or server. The rest of the paper is organized as fellows. The related work is introduced in Section 2. Section 3 describes the detection mechanism in detail based on the abnormal communication behaviors. The proposed detection system is evaluated in Section 4 and the conclusion is drawn in Section 5.

2. Related work

Botnet becomes one of the major threats on the Internet. Some researchers analyzed botnet behaviors to understand the architecture of botnet, command and control (C&C) channel, and the capabilities of bots (Barford and Yegneswaran, 2006; Stinson and Mitchell, 2007).

Lu et al. (2011) classified botnet detection and tracking methods into three categories: honeypots, traffic application classification, and passive anomaly analysis. Deploying honeypots or honeynets is an effective approach to capture and observe botnet. Freiling et al. (2005) present a DDoS prevention scheme based on the observed behaviors from the deployed honeypots. Rajab et al. (2006) track IRC-based botnet propagation patterns from honeynets. Their research results show that botnets may generate a massive volume of unwanted Internet traffic due to spreading action. In traffic classification research, some work focuses on centralized botnet communication protocols, namely IRC and HTTP, and adopts data mining method. Livadas et al. (2006) applied a variety of machine learning techniques to classify IRC/non-IRC traffic with 2.5% false negative rate and 15% false positive. Later, Strayer et al. (2006) improved the previous work by filtering out bulk flows and correlating the remaining flows with the similarity characteristics found in command and control flows. The authors (Strayer et al., 2008) focused on improving flow correlation by temporal relation and concluded that the interdependency exists among the botnet traffic. The flow attributes used in the studies mostly are average or covariance of size or time related features. The above three studies were evaluated using the same bot program and the bot traces consist of one hour traffic. More traces were experimented on the latest one (Strayer et al., 2008). BotMiner (Gu et al., 2008b) identifies botnet activities and control channels from flow information and packet payload. It divides traffic into normal and malicious groups and performs cross cluster correlation to identify the hosts that share similar malicious communication patterns and activity patterns. Masud et al. (2008) propose a temporal correlation algorithm to aggregate multiple log files and the detail network logs are required, produced by *tcpdump* and *exedump*, in order to detect bot activities. A variety of data mining algorithms are used to model the C&C traffic. Lu et al. (2011) employ n-gram, decision tree and clustering algorithms to classify network traffic into different application communities. The proposed detection system analyzes the temporal-frequent characteristics of the 256 ASCII bytes on the payload over a predefined time interval to distinguish malicious bot traffic from normal one. Zhao et al. (2013) use a decision tree with Reduced Error Pruning algorithm (REPTree) to classify network traffic behavior and detect botnet activity based on the proposed flow analysis model. The proposed classification model identifies the presence of existing and unknown botnets activity with high accuracy even with very small time windows.