Journal of Advanced Research (2014) 5, 435–448

Cairo University

Journal of Advanced Research

ORIGINAL ARTICLE

An efficient method to detect periodic behavior in botnet traffic by analyzing control plane traffic

Basil AsSadhan [a,*], José M.F. Moura [b]

[a] *Department of Electrical Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia*
[b] *Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA*

ARTICLE INFO

Article history:
Received 20 September 2013
Received in revised form 17 November 2013
Accepted 21 November 2013
Available online 28 November 2013

Keywords:
Botnet detection
Control plane traffic
Discrete time series analysis
Walker large sample test
SLINGbot

ABSTRACT

Botnets are large networks of bots (compromised machines) that are under the control of a small number of bot masters. They pose a significant threat to Internet's communications and applications. A botnet relies on command and control (C2) communications channels traffic between its members for its attack execution. C2 traffic occurs prior to any attack; hence, the detection of botnet's C2 traffic enables the detection of members of the botnet before any real harm happens. We analyze C2 traffic and find that it exhibits a periodic behavior. This is due to the pre-programmed behavior of bots that check for updates to download them every T seconds. We exploit this periodic behavior to detect C2 traffic. The detection involves evaluating the periodogram of the monitored traffic. Then applying Walker's large sample test to the periodogram's maximum ordinate in order to determine if it is due to a periodic component or not. If the periodogram of the monitored traffic contains a periodic component, then it is highly likely that it is due to a bot's C2 traffic. The test looks only at aggregate *control plane* traffic behavior, which makes it more scalable than techniques that involve deep packet inspection (DPI) or tracking the communication flows of different hosts. We apply the test to two types of botnet, tinyP2P and IRC that are generated by SLINGbot. We verify the periodic behavior of their C2 traffic and compare it to the results we get on real traffic that is obtained from a secured enterprise network. We further study the characteristics of the test in the presence of injected HTTP background traffic and the effect of the duty cycle on the periodic behavior.

Introduction

Botnets are large networks of bots (compromised machines) that are under the control of a small number of bot masters. In recent years, the threat posed by botnets toward Internet applications and communications has escalated. This is due to the fact that a bot master controls a large number of bots that ranges from hundreds of thousands to millions. This magnifies the impact of well-known network malicious activities such as scanning, E-mail spam and distributed denial-of-service (DDoS) attacks. Moreover, botnets increase the effectiveness of phishing, click fraud, identity theft, and espionage.

Due to the destructive capabilities of botnets, they have become a major threat to economy, information, and communication infrastructures. The Federal Bureau of Investigation (FBI) in the United States, in an initiative to detect bot masters

[*] Corresponding author. Tel.: +966 114676755; fax: +966 114676757.
E-mail address: bsadhan@ksu.edu.sa (B. AsSadhan).

Peer review under responsibility of Cairo University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jare.2013.11.005

and take them apart has identified over 1 million victim computers [1,2]. Many people have been indicted, pleaded guilty, or been sentenced for crimes related to botnet usage [1,2]. What increases the impact of the problem is that the majority of the owners of the compromised machines are not aware that their machines are a member of a botnet [1]. According to the April 2013 Symantec Internet Security Threat Report, Volume 18, 3.4 million distinct bot-infected computers were observed in 2012 [3]. According to the same report, botnets were responsible for about 69% of spam E-mail in 2012 [3]. The good news; there is a decrease in these numbers over the past years. For example, in 2009, there were 6.08 million distinct bot-infected computers and botnets were responsible for about 85% of spam E-mail [4]. Nevertheless, the numbers are still high; moreover, bot masters have begun linking mobile smart phones to form botnets of mobile devices to make monetary profits [3].

Botnets' traffic is different from the traffic of other types of malware in that it includes command and control (C2) communication channels traffic. A bot master relies on these channels to send commands to the members of its botnet to execute attack activities. In addition, a bot master relies on these channels to control botnet members to obtain the needed information and code to run their attacks. C2 communication channels traffic occurs before the execution of attack activities and can be considered as the intelligence communication between the different members of a botnet. This makes the detection of C2 communication channels traffic of interest as it means detecting bots before any targeted victim is attacked.

The detection of C2 traffic is difficult due to several reasons that was pointed out by AsSadhan et. al [5]. They include the following: (1) the low volume of C2 traffic; (2) C2's traffic is well behaved and does not violate any network protocol rules; (3) there may be only a few number of botnet members in the monitored network; and (4) the C2 traffic might be encrypted [5]. To tackle these difficulties, we look at one behavior that we, along with other researchers, observed in C2 traffic [5–7]. The behavior we observe is spatial-temporal correlation and similarities in the C2 communication traffic of the bots belonging to the same botnet.

In our work, we focus on temporal correlation in a single bot's traffic. We find that a bot's C2 traffic exhibits periodic behavior. This is due to the nature of the pre-programmed behavior of a bot, where in many variations of botnets each bot frequently contacts other bots every T seconds. This pre-programmed behavior is present in botnets with different structures and communication protocols and is done in order for bots to update their data, receive commands, and send keep-alive messages. We note that the periodic behavior is observed when looking at the traffic of the transport port number used by the bot for its C2 communication.

As a result, the detection of periodic behavior in a machine's traffic might be an indication that the machine is a member of a botnet. We exploit this observation in order to detect bots by detecting periodic behavior in the traffic of the network we monitor. To achieve this we present in this paper an efficient method to detect periodic behavior in botnet command and control traffic. The method is based on the evaluation of the traffic sequence's periodogram. A periodogram is used to view a periodic signal in the frequency domain to observe the peak located at the fundamental frequency of the signal. After the peak is located, we apply Walker's large sample test to decide whether or not the peak is significant enough compared to the rest of the periodogram's ordinates. In case the peak is significant, we declare that it is due to a periodic component with the frequency where the peak is located.

To increase the efficiency of the method further, we decompose enterprise LAN TCP traffic into control and data planes [8], and use the control plane traffic as a surrogate for the whole traffic (control and data planes combined). This is because data traffic generation is based on control traffic generation, which makes the behavior of the two traffic groups similar [8].

The rest of the paper is organized as follows, Section "Background and motivation: Detection of periodic behavior in botnet C2 communication channels traffic" reviews the command and control (C2) traffic of botnets and proposes how to detect botnets. Section "Approach: Discrete time series analysis of aggregate traffic" explains how we aggregate network traffic and decompose them into control and data planes traffic. Section "Methodology: Test network traffic for periodic behavior using periodograms" reviews periodograms and presents the Walker's large sample test. Section "Experimental setup: Evaluation and analysis" explains the experimental setup and presents our evaluation and analysis results of applying the test to several packets traces, and in Section "Conclusions" we give our conclusions.

Background and motivation: Detection of periodic behavior in botnet C2 communication channels traffic

Since a bot master controls a botnet via command and control (C2) communication channels. Our approach is to detect a botnet through the detection of its C2 communication channels traffic. This technique is effective as it detects bots before they engage in harmful malicious activities. This is because C2 traffic by itself is harmless, and its detection it will enable the detection of the bots that are transmitting/receiving it.

The C2 communication channels between bots and the C2 servers are based on either a pull or push mechanism [7]. Depending on the mechanism used, bots are pre-programmed to contact each other every T seconds to update bot's data, receive commands, and send *keep-alive* messages. This pattern of behavior is present in bots irrespective of the botnet's structure or the communication protocol being used between bots and the C2 server. This results in having a *periodic* behavior in the bot's traffic over a given transport port number. We note that in other botnet variants, C2 communication happens might occur in an aperiodic manner at arbitrary times. We briefly discuss this issue in Section "Experimental setup: Evaluation and analysis".

In our work, we exploit this periodic behavior to detect C2 communication traffic. In addition to our previous works [5,6], we are aware of a previous study, that exploits the periodic behavior of botnet C2 traffic to detect bots. In Gu et al. [7], the host's traffic autocorrelation function was computed in the time domain to examine whether the traffic has a periodic component or not. We, however, work in the frequency domain, as it involves less amount of computations, thus is faster in time. This is done thorough evaluating the periodogram [9] of the traffic and then applying Walker's large sample test [10] to the periodogram's maximum ordinate to detect periodic components.