Chapter 19
Detecting Botnet Traffic from a Single Host

Sebastián García
Universidad Nacional del Centro (UNICEN University), Argentina & Czech Technical University (CTU University), Czech Republic.

Alejandro Zunino
Universidad Nacional del Centro (UNICEN University), Argentina

Marcelo Campo
Universidad Nacional del Centro (UNICEN University), Argentina

ABSTRACT

The detection of bots and botnets in the network may be improved if the analysis is done on the traffic of one bot alone. While a botnet may be detected by correlating the behavior of several bots in a large amount of traffic, one bot alone can be detected by analyzing its unique trends in less traffic. The algorithms to differentiate the traffic of one bot from the normal traffic of one computer may take advantage of these differences. The authors propose to detect bots in the network by analyzing the relationships between flow features in a time window. The technique is based on the Expectation-Maximization clustering algorithm. To verify the method they designed test-beds and obtained a dataset of six different captures. The results are encouraging, showing a true positive error rate of 99.08% with a false positive error rate of 0.7%.

INTRODUCTION

In the last decade botnets have evolved from being used as a personal activity platform to becoming a financially aimed structure controlled by malicious groups (Wilson, 2007). We consider a botnet as a network of remotely controlled, compromised computers, used for malicious purposes. The hosts in a botnet are called 'Bots' and the owner of a botnet is called 'Botmaster'. Botnets have become the technological backbone of a growing community of malicious activities (Clinton, 2008), from small DDoS (Distributed Denial of Service attacks) to worldwide spam campaigns. They still remain as the most significant threat to the Internet today.

The first attempts to control a malicious programs remotely first appeared in late 1999. Since then the primary goal of the owners has been to

DOI: 10.4018/978-1-4666-7381-6.ch019

obtain financial gain. This forced the development of several botnet detection technologies trying to cope with the attacks, but botnets resisted besiege security measures resting on the infection of home based computers, circumventing security methods (Stone-Gross, Cova, Cavallaro, Gilbert, Szydlowski, Kemmerer, Kruegel & Vigna, 2009), encryption algorithms and anti-reverse engineering techniques. Although the IRC (Internet Chat Relay) protocol has been the most used command and control (C&C) channel in the last decade, nowadays the trend is towards decentralized networks, such as P2P (Peer to Peer) (Yan, Eidenbenz & Nago, 2009) (Kang, Zhang, Li & Li, 2009).

A wide diversity of methods had been proposed to detect botnets. In the survey presented by Garcia, Zunino and Campo (2013) there is a basic classification of these network-based detection methods. It shows that there is still a large amount of signature-based methods and protocol-dependant feature analysis. While these techniques may work under certain conditions, they are usually not enough to capture new botnets that significantly deviate from those signatures. More important, the survey shows a growing amount of algorithms making use of the behavior of the botnets. These techniques are more dynamic and therefore have a better chance to detect new behavior.

Since a real and large network usually generate a huge amount of diverse traffic, instead of detecting a botnet, some proposals focused only on detecting a single infected host. The survey shows that there are some proposals in this area, because it may be easier to differentiate a single bot from a single normal computer.

Our proposal is based on the idea of detecting a single bot by means of its network-based behavior. The benefits of analyzing the traffic of one bot, compared to analyzing a network, are that there is considerably less traffic to process and that the traffic may tend to be more homogeneous. On the other hand, the disadvantages are that less traffic may mean less behavior available to detect the bot and that to capture the traffic of one host we usually need the authorization of the owner.

The detection model that is proposed in this paper was created after a thorough analysis of the most inherent characteristics of the behavior of the bot. This analysis showed that the most typical characteristics of the bot are maliciousness (attacking and infecting, sending SPAM, DDoS, etc.) and being remotely managed. During some botnet life cycle phases, a single bot computer usually generates high network flow rates within very short time periods (Gu, 2008), for example, during Spam sending, DDoS attacks, network scanning and botnet distribution. Therefore, we focus on the relationship between the amount of IP addresses and ports in a time window. We hypothesize that our group of features can be used to detect the traffic of a single bot.

Our proposal first separates the bots' flows into time windows, then it extracts some aggregated features and then it applies a clustering algorithm (Baeza-Yates, 1999) to detect the bot. With this method it is possible to detect bots using encrypted traffic, without using static details of the protocols and within the first stages of infection.

The validation of the method was done by capturing a real dataset that contains labeled botnet traffic and labeled non-botnet traffic. This dataset allowed us to achieve a verified and more robust algorithm. For verification purposes the dataset was made public and can be downloaded from the website http://mcfp.felk.cvut.cz. This site is the Malware Capture Facility Project were more botnets are being captured, labeled and published (Garcia, 2013).

The contributions of our proposal are:

- It separates the aggregated network flows in time windows using behavioral features.
- It clusters instances using the Expectation-Maximization algorithm.
- It evaluates the algorithm with botnet, non-botnet and manual attacks labeled data.