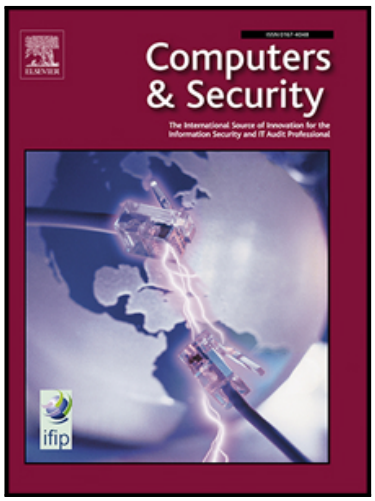

An Adaptive Framework for the Detection of Novel Botnets

Javier Álvarez Cid-Fuentes, Claudia Szabo, Katrina Falkner

PII: S0167-4048(18)30980-5
DOI: https://doi.org/10.1016/j.cose.2018.07.019
Reference: COSE 1391

To appear in: *Computers & Security*

Received date: 27 November 2017
Revised date: 9 May 2018
Accepted date: 16 July 2018

Please cite this article as: Javier Álvarez Cid-Fuentes, Claudia Szabo, Katrina Falkner, An Adaptive Framework for the Detection of Novel Botnets, *Computers & Security* (2018), doi: https://doi.org/10.1016/j.cose.2018.07.019

An Adaptive Framework for the Detection of Novel Botnets

Javier Álvarez Cid-Fuentes, Claudia Szabo, Katrina Falkner

School of Computer Science, The University of Adelaide
Adelaide, Australia

Abstract

Detecting and disrupting botnet activities is critical for the reliability, availability and security of Internet services. However, despite many efforts in this direction, key challenges remain. These include the high computational requirements of processing large amounts of network information, the similarity between botnet and normal traffic, and the constant creation of new botnet mechanisms to bypass current detection approaches. Because of these challenges, existing detection approaches have difficulties in detecting novel botnets with high accuracy and low false positive rate. In this paper, we address this problem with an scalable and decentralized framework. Our framework creates a complete characterization of the behavior of legitimate hosts that can be used to discover previously unseen botnet traffic. Moreover, our framework dynamically adapts to changes in network traffic, and is capable of detecting novel botnets without any assumption on their architecture or protocols employed. This is crucial to nullify the constant efforts by botnet managers to adapt to current detection techniques. Through an experimental analysis using the most realistic and varied publicly available botnet dataset, we find that our framework can detect bots in a network with 1.00 TPR and 0.082 FPR or, alternatively, can detect half of the malicious hosts with a FPR as low as 0.0017. These results significantly improve the results reported by similar works in the area, with the added value of not relying on historical botnet data or specific architectures and protocols.

1. Introduction

Botnets are networks of illegally controlled computers (i.e., *bots*) typically employed for malicious activities [1]. These networks are created by infecting a large number of computers with *malware* (i.e., malicious software) by means of operating system vulnerabilities, USB drives, or malicious web sites. Once a victim's computer is infected, the botnet software allows an attacker (also known as *bot master*) to take control and carry out malicious activities such as e-mail spamming or distributed denial-of-service (DDoS) attacks. Upon infection, and to remain undetected, bots typically update themselves, disable antivirus applications, block DNS lookups to certain domains, and download and run other types of malware.

Since the appearance of the first Internet Relay Chat (IRC) [2] based bot in 1993 [3], botnets have become a dangerous threat difficult to detect and dismantle. A novel malware used to create large botnets of small devices called Mirai [4] is believed to be behind a recent massive DDoS attack that disrupted access to tens of services in October 2016 [5]. This attack, which employed around 100,000 Internet of Things (IoT) devices, could be the largest DDoS attack in history, with an estimated throughput of 1.2 Tbps. Apart from the economic losses that downtime causes in industry [6], botnets can cause numerous issues in other sectors, such as defense [7] or public administrations [8].

The damaging potential of botnets has led to significant efforts to detect and prevent them [9]. However, key challenges remain in the area, such as [9] the continuous evolution in botnet creation, maintenance and communication mechanisms; the fact that botnet traffic is very similar to regular traffic in many cases; and the high computational resources required to analyze large amounts of information

Email addresses: javier.alvarez@bsc.es (Javier Álvarez Cid-Fuentes), claudia.szabo@adelaide.edu.au (Claudia Szabo), katrina.falkner@adelaide.edu.au (Katrina Falkner)