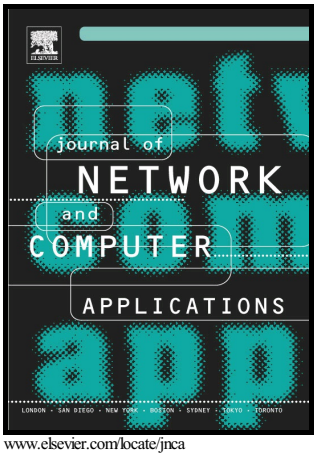
Survey of Approaches and Features for the Identification of HTTP-Based Botnet Traffic

Dilara Acarali, Muttukrishnan Rajarajan, Nikos Komninos, Ian Herwono



www.elsevier.com/locate/jnca

PII: S1084-8045(16)30236-3
DOI: http://dx.doi.org/10.1016/j.jnca.2016.10.007
Reference: YJNCA1734

To appear in: *Journal of Network and Computer Applications*

Received date: 24 March 2016
Revised date: 8 September 2016
Accepted date: 13 October 2016

Cite this article as: Dilara Acarali, Muttukrishnan Rajarajan, Nikos Komnino and Ian Herwono, Survey of Approaches and Features for the Identification o HTTP-Based Botnet Traffic, *Journal of Network and Computer Applications* http://dx.doi.org/10.1016/j.jnca.2016.10.007

Available online at www.sciencedirect.com

ScienceDirect

Journal Logo

00 (2016) 1–19

Survey of Approaches and Features for the Identification of HTTP-Based Botnet Traffic

Dilara Acarali[a], Muttukrishnan Rajarajan[a], Nikos Komninos[a], Ian Herwono[b]

[a]*School of Engineering and Mathematical Science, City University London, London, United Kingdom.*
[b]*Security Futures Practice, Research & Innovation, British Telecom, Ipswich IP5 3RE, United Kingdom.*

Abstract

Botnet use is on the rise, with a growing number of botmasters now switching to the HTTP-based C&C infrastructure. This offers them more stealth by allowing them to blend in with benign web traffic. Several works have been carried out aimed at characterising or detecting HTTP-based bots, many of which use network communication features as identifiers of botnet behaviour. In this paper, we present a survey of these approaches and the network features they use in order to highlight how botnet traffic is currently differentiated from normal traffic. We classify papers by traffic types, and provide a breakdown of features by protocol. In doing so, we hope to highlight the relationships between features at the application, transport and network layers.

Keywords: Bot, botnet traffic, network analysis, feature analysis, network-based detection

1. Introduction

The digital age has brought many benefits to society, with the Internet acting as an enabler for growth and development across practically all sectors of business and industry. Unsurprisingly, it has also become an attractive and fertile environment for criminal activity. Malware programs, which may take many forms, are now frequently used for financial theft, identity theft, espionage, and disruption of services. A particularly troublesome type of malware is a bot, an exploited system which acts as a remote tool for an attacker to control and use the resources of a target system. Typically, the attacker (called the botmaster) will do the same for multiple systems and then use them collectively, in what is known as a botnet, to launch attack campaigns.

Botnet classification is based on the type of communication protocol used. The main classes currently defined by the research community are P2P-based, IRC-based, and HTTP-based. P2P-based botnets use a decentralised architecture, in which each node can act as both a client and a server. In contrast, both IRC and HTTP-based botnets operate with a centralised architecture, where bots are required to connect to a command and control (C&C) server in order to upload data or receive commands. For this study, we have chosen to focus on HTTP-based botnets. The wide-spread use of web-based services has pushed the adoption of HTTP as an advantageous communication protocol thanks to the fact that it is usually permitted by firewalls, and it allows bot traffic to blend in with vast volumes of benign activity. The benefits of using HTTP rather than IRC are demonstrated and discussed by Farina et al. (2016) and Gu et al. (2008). Additionally, McAfee (2015) listed at least 5 HTTP-based botnets as top spammers in their 2014 Q4 Threat Report. This was also echoed by Symantec (2014), where HTTP-based botnets made up over half of their 2014 top 10 spamming botnets.

The difficulty of detecting botnet traffic (especially for HTTP-based botnets) amongst web activity is a current research challenge. Rostami et al. (2014) compared clean network traces to those collected from HTTP-based botnets. They focused on HTTP data in client-side PCAP files to highlight differences in clean and malicious traffic at the application layer. However, they do not consider how this traffic may manifest at other layers, nor how it may be measured. Both Haddadi & Zincir-Heywood (2014) and

Email addresses: dilara.acarali@city.ac.uk (Dilara Acarali), r.muttukrishnan@city.ac.uk (Muttukrishnan Rajarajan), nikos.komninos1@city.ac.uk (Nikos Komninos), ian.herwono@bt.com (Ian Herwono)