Alexandria Engineering Journal (2016) xxx, xxx–xxx

HOSTED BY

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

ORIGINAL ARTICLE

Detection of randomized bot command and control traffic on an end-point host

B. Soniya *, **M. Wilscy**

Dept. of Computer Science, University of Kerala, Trivandrum, India

Received 22 November 2015; accepted 4 April 2016

KEYWORDS

Botnet detection;
Randomized traffic;
Traffic modeling;
Anomaly based

Abstract Bots are malicious software entities that unobtrusively infect machines and silently engage in activities ranging from data stealing to cyber warfare. Most recent bot detection methods rely on regularity of bot command and control (C&C) traffic for bot detection but state-of-the-art bots randomize traffic properties to evade regularity based detection techniques. We propose a bot detection system that aims to detect randomized bot C&C traffic and also aim at early bot detection. To this end, separate strategies are devised for bot detection: (i) over a user session and (ii) time periods larger than a user session. Normal HTTP traffic and bot control traffic are modeled over a user session and a Multi-Layer Perceptron Classifier is trained on the two models and later used to classify unlabeled destinations as benign or malicious. For traffic spanning time intervals larger than a user session, *temporal persistence*, is used to differentiate between traffic to benign and malicious destinations. Testing with multiple datasets yielded good results.
© 2016 Faculty of Engineering, Alexandria University. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

A botnet is a network of compromised machines controlled by a botmaster through a command and control (C&C) channel [1]. Botnets are highly malevolent entities used by cyber criminals in launching various attacks ranging from data stealing, spamming, phishing and DoS attacks to cyber warfare against nations and detecting and mitigating their effects demand critical attention.

The major focus of research in botnet detection falls under the broad category of anomaly based intrusion detection systems [2] which can further be classified into host-based and network based. Host-based bot detection systems [3–7] are located on the host itself and inspect evidence collected from that host such as system calls executed and their sequence, files and registry entries modified, processes that are active, user input and interaction and malware signatures, to decide whether the machine is bot infected or not. But bots have evolved over the years and employ several evasive techniques to avoid detection including use of packers [8], polymorphism and other code obfuscation methods [9] and rootkit techniques [10,11]. A bot with rootkit component is used to hide its presence on the host by suppressing all evidence exhibited by it and thus evade host-based detection. Network-based botnet detection systems [1,12–17] are located at the network edge and look at packet traffic arriving at the network edge to identify possible bot generated communication patterns. Bot C&C is the weakest link in a botnet and sev-

* Corresponding author.
E-mail addresses: soniya@ieee.org (B. Soniya), wilsyphilipose@hotmail.com (M. Wilscy).
Peer review under responsibility of Faculty of Engineering, Alexandria University.

http://dx.doi.org/10.1016/j.aej.2016.04.004
1110-0168 © 2016 Faculty of Engineering, Alexandria University. Production and hosting by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article in press as: B. Soniya, M. Wilscy, Detection of randomized bot command and control traffic on an end-point host, Alexandria Eng. J. (2016), http://dx.doi.org/10.1016/j.aej.2016.04.004

eral recent bot detection approaches [18–20] exploit the regularity of communication patterns of bot C&C channel for bot detection. Newer bots such as Stuxnet randomize the communication pattern between bots and their masters thereby evading regularity based detection techniques and to our knowledge detection of randomized bot C&C traffic has not been addressed earlier. Another requirement for botnet detection that has largely been unaddressed by existing detection strategies is early detection of bots.

We propose a bot detection system which uses traffic analysis of an end-point host to identify bot command and control (C&C) communications even when the communication patterns are randomized to evade detection and also aims at early detection of bots. Towards this end, the bot detection system proposed is made up of two parts – a Multi Layer Perceptron (MLP) Classifier designed to separate user generated HTTP traffic and bot command and control traffic over a user session and a Temporal Persistence (TP) Classifier that makes use of temporal persistence, the property of bot control traffic to repeatedly contact its bot master over time to detect bot C&C traffic. The duration of a user session is found to be around 15 min [21] and normal user-generated HTTP traffic and bot control traffic are modeled over a user session and used to train a Multi Layer Perceptron (MLP) Classifier which is later used to separate bot and benign traffic within the user session, thereby enabling early detection. Bot and benign traffic are separately modeled and used to train the MLP Classifier and randomizing bot C&C traffic would move it away in feature space from the trained bot model but it would still be sufficiently further away from the normal user-generated HTTP traffic model to be classified as benign. Hence randomization of traffic features does not affect the classification accuracy of the MLP Classifier much as is seen in Section 4.5. Bot traffic could span time intervals larger than a user session and such traffic is taken care of by the TP Classifier. The TP Classifier monitors temporal persistence of traffic over multiple timescales varying from 150 min, 10 h, 20 h and 40 h. Further, the TP Classifier considers only the repeated nature of traffic to a destination and does not consider the exact time of bot communication or the exact time intervals between bot communications and hence can detect traffic to bot servers even when the traffic is randomized to evade detection.

The proposed bot detection system is trained and tested using bot traffic generated using Zeus and BlackEnergy bots run on DETER testbed [21] and normal HTTP traffic generated using a clean Windows XP machine. Testing is also done with *novel* traffic – traffic on which the system has not been trained – from Banbra, Bifrose, Dedler, Sasfis, Ramnit and Pushdo bots [22] and the overall bot detection rate of 97.7% and overall detection rate of bot destinations being 84.3%. It may be mentioned that more than 50% of bot samples in the novel dataset generated randomized traffic and the detection rate of the system is good. To estimate the false alerts generated by the system, we used an attack free subset of DARPA dataset [23] and a subset of LBNL dataset [37] and the False Positive Rate of the system are 6% and 2.3% respectively. In 30% of the test cases, bot destinations were identified within 15 min of data generation which is important since early identification of bots helps defenders to devise strategies in protecting against and mitigating further malicious behavior.

2. Related work

Botnet detection is an active area of research and the publications in this area are broadly classified into host-based [3–7] and network-based approaches [1,12–17]. Host-based systems, as already mentioned in Section 1, analyze evidence collected from hosts to detect bots. With bots incorporating rootkit behavior, host-level evidence collected is not reliable. Network-based botnet detection is a complementary approach and monitors network traffic for botnet detection. Based on the protocol used by bots, network-based botnet detection approaches are classified into IRC-based [13,24], P2P-based [25,26], SMTP-based [27,28], DNS-based [16,29] and so on. Not much work exists in detection of HTTP-based bots although generic botnet detection systems such as BotSniffer [30], BotMiner [1] and TAMD [14] are effective detectors. They look for similar communication patterns from multiple hosts to identify bots and hence cannot identify a single bot-infected host. BotHunter [31] models the bot-infection life cycle and identifies the modeled behavior on a host. But bot-infection life cycles evolve with time and detection methods based on the model fail to detect more recent bots. Botzilla [32] also identifies bot-infected hosts from network traffic, but is signature-based. Signature-based systems are effective in identifying known bots but fail to detect newer variants of existing bots as well as new bots.

Giroire et al. [33] use temporal persistence, a measure of repeatedly contacting a destination over time as the distinguishing characteristic to separate benign and malicious destinations. Persistence is a good measure for detecting bots, but the work in [33] is meant for hosts on an enterprise network which are well behaved and the destinations contacted by it are restricted to a limited set so that a static white-list would suffice. Traffic analysis of a host is done by several works [18–20] and utilizes regularity of bot communications to identify bot infection. Botfinder [19] and CoCoSpot [18] generate models of bot C&C behavior by clustering similar bot C&C channels and developing fingerprints for C&C channels. AsSadhan et al. [20] use signal processing techniques to identify periodic traffic from the host. These bot detection methods show a significant degradation in performance when bots use randomization techniques to evade detection. In the following section we present RCC Detector, a bot detection system that proposes to overcome this limitation.

3. Randomized bot command and control channel detector (RCC Detector)

RCC Detector, identifies bot command and control (C&C) communications from an end-point host through traffic analysis of the host. RCC Detector aims (i) at early detection of bots and (ii) to detect bot control traffic even when it is randomized to evade detection. With this aim in mind, RCC Detector is designed to consist of two classifiers: (i) an MLP Classifier that is trained to differentiate between normal user-generated HTTP traffic and bot control traffic over a user session and (ii) the TP Classifier that utilizes temporal persistence to identify bot control traffic for time periods larger than a user session. As already mentioned in Section 1, a user session is

Please cite this article in press as: B. Soniya, M. Wilscy, Detection of randomized bot command and control traffic on an end-point host, Alexandria Eng. J. (2016), http://dx.doi.org/10.1016/j.aej.2016.04.004

Download English Version:

https://daneshyari.com/en/article/7211337

Download Persian Version:

https://daneshyari.com/article/7211337

Daneshyari.com