Chapter 82
Detection of Drive-by Download Attacks Using Machine Learning Approach

Monther Aldwairi
Jordan University of Science and Technology, Department of Network Engineering and Security, Irbid, Jordan

Musaab Hasan
Zayed University, College of Technological Innovation, Abu Dhabi, UAE

Zayed Balbahaith
Zayed University, College of Technological Innovation, Abu Dhabi, UAE

ABSTRACT

Drive-by download refers to attacks that automatically download malwares to user's computer without his knowledge or consent. This type of attack is accomplished by exploiting web browsers and plugins vulnerabilities. The damage may include data leakage leading to financial loss. Traditional antivirus and intrusion detection systems are not efficient against such attacks. Researchers proposed plenty of detection approaches mostly passive blacklisting. However, a few proposed dynamic classification techniques, which suffer from clear shortcomings. In this paper, we propose a novel approach to detect drive-by download infected web pages based on extracted features from their source code. We test 23 different machine learning classifiers using data set of 5435 webpages and based on the detection accuracy we selected the top five to build our detection model. The approach is expected to serve as a base for implementing and developing anti drive-by download programs. We develop a graphical user interface program to allow the end user to examine the URL before visiting the website. The Bagged Trees classifier exhibited the highest accuracy of 90.1% and reported 96.24% true positive and 26.07% false positive rate.

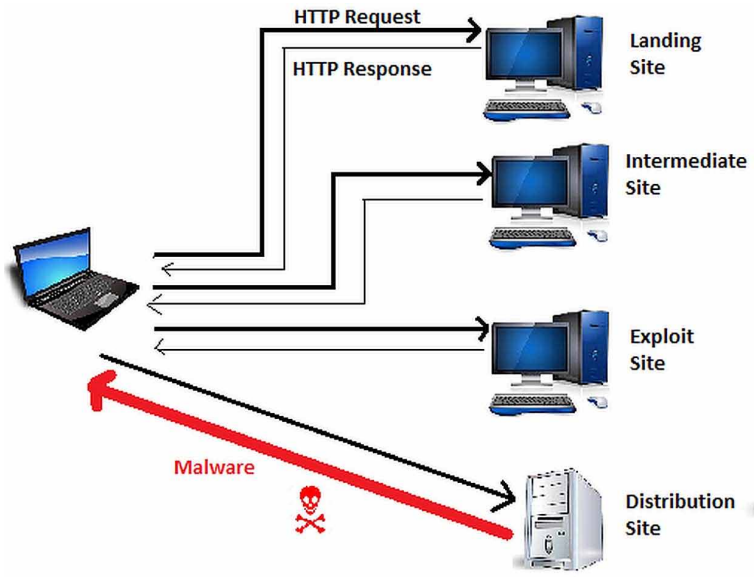
DOI: 10.4018/978-1-7998-2460-2.ch082

Copyright © 2020, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

INTRODUCTION

Everyday Internet users are a target by a large number of attackers who are constantly searching for vulnerabilities to perform various attacks with different motivations and intentions (Harley & Bureau, 2008). Narvaez, Endicott-Popovsky, Seifert, Aval and Frincke (2010) considered drive-by download attacks as one of the most important types of these attacks in which the attacker uses legitimate and illegitimate websites to spread malicious code. A file is downloaded to the user machine without trigger by exploiting a web browser vulnerability. The file usually contains a malicious code that runs on the target computer. This malware could be used to steal confidential data, create a backdoor or serve any imaginable malicious intent. Leit and Cova (2011) believe that drive-by downloads are involved in the spread of most of the recent malware infections.

Figure 1. Drive-by downloads attack flow



Matsunaka, Urakawa, and Kubota (2013) found that the user is simply subjected to this attack by clicking a link in a phishing email, malicious hyperlink, or unwanted popup window. Figure 1 shows one possible scenario to launch a drive-by download attack. First, a malicious website is setup, called the landing or mothership website. This website could be mimicking a legitimate website or actual legitimate website where malicious code is injected. Once the websites are injected with the attack code, they act the first point in a chain of redirections to multiple intermediate websites. The point of the redirections is to hide the actual exploit servers and mislead investigators. The users are finally redirected to the exploit website, which includes a more elaborate malicious code charged with searching for vulnerabilities and flaws based on the version of the user's web browser and operating system. Once vulnerability is located it will be exploited by the malware distribution website to download and install the desired malware directly to user's device without his knowledge. All drive-by attacks need not to follow the exact same flow but the main idea remains valid.

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/detection-of-drive-by-download-attacks-using-machine-learning-approach/252100

Related Content

Liquid Humanitarianism
Syed Ismyl Mahmood Rizvi (2019). *Media Models to Foster Collective Human Coherence in the PSYCHecology (pp. 220-236).*
www.irma-international.org/chapter/liquid-humanitarianism/229338

Graph-Based Semi-Supervised Learning With Big Data
Prithish Banerjee, Mark Vere Culp, Kenneth Jospeh Ryanand George Michailidis (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 214-244).*
www.irma-international.org/chapter/graph-based-semi-supervised-learning-with-big-data/252027

Unleashing Artificial Intelligence onto Big Data: A Review
Rupa Mahantyand Prabhat Kumar Mahanti (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 1682-1697).*
www.irma-international.org/chapter/unleashing-artificial-intelligence-onto-big-data-a-review/252106

Quantitative Semantic Analysis and Comprehension by Cognitive Machine Learning
Yingxu Wang, Mehrdad Valipourand Omar A. Zatarain (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 673-688).*
www.irma-international.org/chapter/quantitative-semantic-analysis-and-comprehension-by-cognitive-machine-learning/252051

A Taxonomy of Data Mining Problems
Nayem Rahman (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications (pp. 512-528).*
www.irma-international.org/chapter/a-taxonomy-of-data-mining-problems/252041