Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Detecting fake anti-virus software distribution webpages

Dae Wook Kim*, Peiying Yan, Junjie Zhang

Department of Computer Science and Engineering, Wright State University, Dayton, OH, USA

ARTICLE INFO

Article history:
Received 8 August 2014
Received in revised form
2 November 2014
Accepted 19 November 2014
Available online 29 November 2014

Keywords:
Network security
Fake anti-virus software
Intrusion detection
Web document analysis
Statistical classification

ABSTRACT

Attackers are continually seeking novel methods to distribute malware. Among various approaches, fake Anti-Virus (AV) attacks represent an active trend for malware distribution. In a fake AV attack, attackers disguise malware as legitimate anti-virus software and convince users to install it. As web browsers become the most popular applications for users to access online resources, webpages have become the dominating means to launch fake AV attacks. In this paper, we presented an automated and effective detection system, namely *DART*, to identify fake AV webpages in the Internet. We proposed a collection of novel features to characterize an unknown webpage and then integrate them using statistical classifiers. These features focus on profiling a fake AV webpage from three aspects that are fundamentally important for its success, thereby resulting in the high detection accuracy and implying resistance against evasion attempts. We have performed extensive evaluation based on real fake AV webpages that are collected from the Internet. Experimental results have demonstrated that *DART* can accomplish a high detection rate of 90.4% at an extremely low false positive rate of 0.2%.

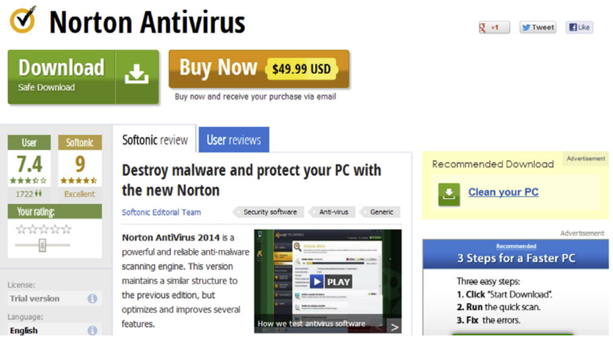
1. Introduction

Tremendous cyber-security concerns have led to computer systems with enhanced security features. As a result, it becomes increasingly difficult for attackers to directly compromise end users' systems by exploiting software vulnerabilities. As an alternative strategy, social engineering attacks, which take advantage of humans' psychological vulnerabilities, rapidly gain their popularity. Among various types of social engineering attacks such spamming and phishing, fake Anti-Virus (AV) attacks have become one of the most significant threats.

A fake AV attack is to disguise malware as anti-virus software and lure end users into installing it. As web browsers become the dominating network applications, webpages have become a major way for attackers to launch fake AV attacks (Fossi et al., 2009). Specifically, upon being visited by a user, a fake AV webpage usually claims that it offers a customized anti-virus product and encourages the user to install it, where the offered anti-virus product is actually a malicious executable. In order to stimulate users' action, a fake AV webpage could claim that it has identified (new) infections on end users' systems after a thorough scanning. Both the appearance of the anti-virus product and scanning process impersonate those of popular security vendors such as Microsoft Security Essentials and Symantec Norton. Fig. 1 presents an example of a typical fake AV webpage, which impersonates a safe webpage to download a free Symantec Norton Antivirus 2014. A user who approves the installation of the fake anti-virus

* *Corresponding author.*
E-mail addresses: kim.107@wright.edu (D.W. Kim), yan.20@wright.edu (P. Yan), junjie.zhang@wright.edu (J. Zhang).

http://dx.doi.org/10.1016/j.cose.2014.11.008
0167-4048/© 2014 Elsevier Ltd. All rights reserved.



Destroy malware and protect your PC with the new Norton

Norton AntiVirus 2014 is a powerful and reliable anti-malware scanning engine. This version maintains a similar structure to the previous edition, but optimizes and improves several features.

Three easy steps:
1. Click "Start Download".
2. Run the quick scan.
3. Fix the errors.

Fig. 1 – An example screenshot of a fake AV webpage.

software will immediately render his/her system infected. As a result of users' increasing awareness of malware threats and growing familiarity of anti-virus software, fake AV attacks are surprisingly successful. In fact, a recent study from Google has indicated that fake AV attacks are responsible for 50% of all malware delivered via Internet advertising and their prevalence keeps growing (Rajab et al., 2010).

Detecting fake AV webpages is therefore of great importance. However, it is a challenging task due to several typical characteristics for fake AV webpages. First, fake AV webpages require users' interaction to install malware and hence do not require any malicious contents to perform automatic exploitation without users' consent (e.g., shellcode). In addition, a fake AV webpage does not need to impersonate a *whole* authentic anti-virus webpage. Instead, it often uses a few representative elements from the impersonated webpage, such as the names and icons of anti-virus products, while the remaining webpage components could exhibit arbitrary semantics. Finally, in order to trick users into believing that the masqueraded products offer up-to-date solutions to emerging threats, attackers may frequently update the information for both products and threats (e.g., the product versions and threat names). These characteristics easily distinguish fake AV attacks from other prevalent web-based attacks such as drive-by downloads and phishing, and consequently impede the direct application of the detection methods for these attacks to detect fake AV attacks. For example, drive-by download detection methods that require the observation of webpage content for exploitation (Cova et al., 2010; Curtsinger et al., 2011) or automatic binary downloading (Lu et al., 2010) will fail to detect fake AV webpages due to the absence of malicious content for exploitation. Phishing-webpage detection methods that rely on the similarity between a potential malicious webpage and its impersonated authentic webpage (Rosiello et al., 2007) can be easily circumvented.

In order to effectively detect fake-AV detection and overcome the aforementioned challenges, we have designed a novel system, namely *DART*. *DART* employs a collection of features to profile a webpage and further discriminate between fake AV webpages and benign webpages by integrating these features using a statistical classifier. These features aim to characterize a webpage by answering three questions that are critical for the success of a fake AV webpage. First, how does a fake AV webpage increase the opportunity to be visited by an end user? What tricks does it use to convince an end user to install the executable? Where is a fake webpage actually located? Specifically, *DART* automatically extracts features that profile search engine optimization techniques used by fake AV webpages, various identities to impersonate authentic security webpages, and the network infrastructures. Since malware threats are constantly evolving, identities related to authentic anti-virus software, such as names of the product and malware, may actively change. In order to solve this challenge, *DART* can automatically discover diverse and trendy security keywords by performing semantic analysis based on messages from popular social networks such as Twitter. We have performed extensive evaluation based on data collected from the real-world network. Our experimental results have demonstrated that *DART* can achieve a high detection rate (90.4%) with a very low false positive rate (0.2%).

The rest of the paper is organized as follows. Section 2 introduces the related work. Section 3 describes how data was collected. We present the system design in Section 4 and the evaluation results in Section 5. A discussion of *DART* is provided in Section 6 and Section 7 concludes.

2. Related work

Extensive measurement efforts (Rajab et al., 2010; Stone-Gross et al., 2013) have been invested to study the severity of fake AV attacks and gained in-depth understanding of their infrastructures and operations. For example, Rajab et al. (2010) performed large-scale study based on malicious webpages collected by Google and concluded that fake AV accounted for 15% of all malware detected on the web. Stone-Gross et al. (2013) managed to acquire back-end servers for several fake AV campaigns and revealed sophisticated techniques adopted by attackers to accomplish robustness and agility of their networking infrastructures.

A few methods (Dietrich et al., 2013; Seifert et al., 2013), which target at detecting fake AV attacks, have been proposed. Dietrich et al. (2013) designed a method to detect fake AV attacks by dynamically analyzing suspicious binaries. Specifically, binaries are executed and their user interfaces will be collected. The method will aggregate binaries with similar user interfaces into clusters. If several binary instances in a cluster are known to be a fake AV, their maliciousness can be propagated to other binaries in the same cluster. Seifert et al. (2013) proposed a method to detect fake AV webpages, which focuses on detecting webpages that show an animation of an anti-virus scan. This method first extracts a collection of features from the snapshot of a webpage, which aim at characterizing visual elements related to anti-virus scanning. It further employs a statistical classifier to discriminate between fake AV webpages and legitimate ones based on the derived features. These methods have demonstrated promising detection performance. However, they suffer from several limitations, which may impede their practical deployment and effectiveness. Specifically, the first method (Dietrich et al., 2013) mandates the execution of a binary, making it extremely challenging to deploy it in end users' hosts. In addition, the execution of binaries implies considerable computation resources (e.g., mounting and