Análise, comparação e verificação do comportamento de *malware* e *goodware* – Analysis, comparison and assessment of the behavior of *malware* and *goodware*.

Giovanni Bertão*, Prof. Paulo Lício de Geus.

Abstract

Most software installed in current user's machines were obtained from public software repositories, which are responsible for million downloads everyday. Whereas practical, the reliance on this type of service might render users susceptible to downloading infected software to their machines. In this work, we investigated the prevalence of malicious software on public software repositories, thus assessing Internet users security. We discovered that up to 33.5% of software available in multiple repositories were flagged as malicious by at least one antivirus.

Key words:

Malware Analysis, Program Tracing, Computer Security.

Introduction

One of the most threatening activities to current computer systems and their users are attacks deployed by malicious software known as *malware*. They subvert the legitimate operation of a system, thus affecting its integrity, confidentiality, and availability. *Malware* attacks might have multiple goals, such as information evasion, credential theft, identity forging, illegal content storage, attacks to third parties, etc. A popular way of attacking systems and their users is to embed a *malware* into legitimate software (also known as *goodware*), a process named trojanization. This way, the user is deceived into benefiting from the resources provided by the legitimate program while malicious activities are performed by the *malware* in the background. Considering this scenario, this paper investigates the presence of trojanized applications in popular online software repositories used by most Internet users, thus assessing Internet users security.

Methodology

To evaluate the public software repositories, we periodically collected the most downloaded software available in popular software repositories (Softpedia, Source Forge, CNet, File Hippo and File Horse) and submitted them to a series of analysis procedures. More specifically, the applications were automatically downloaded by a crawler and traced in a sandboxed environment¹ that automatically installs the software via an interface clicker implemented using an Autoit script. All analysis data was consolidated in a database that was further queried to estimate the risks that users are exposed, following guidelines already adopted by previous work².

Results and Discussion

During a six-month period, we analyzed a total of 46018 downloads, corresponding to 2935 distinct applications (based on their MD5 hash). Some applications were downloaded multiple times due to their presence in multiple repositories. We discovered that the software repositories present their own dynamic, changing, on average, 6% of the distributed binaries during each week. We also discovered that, on average, each application was updated 9 times during the observation period. The

application that was most updated, presented 33 binary changes. This high rate of updates opens a significant window for application trojanization. In practice, we discovered that one in each 3 applications were labeled as a *malware* by at least one antivirus solution, thus indicating a potential binary compromise. More specifically, we discovered that 36% of the malicious samples were labeled as a "Trojan.Agent" and 19% were labeled as "adWare". Table 1 details the antivirus detection per software repository.

Table 1. Percentage of Trojan per software repositories.

Repository	Trojans (%)
FileHorse	46.2%
Cnet	28.5%
FileHippo	16.5%
Source Forge	6.00%
Softpedia	2.50%

Conclusions

Our results show that up to 33.5% of files available in public software repositories were flagged as *malware* by at least one antivirus, thus indicating that users might be at risk if installing such applications. Therefore, we recommend users to not blindly rely on public software repositories despite their popularity and always verify the reputation of the downloaded files.

Acknowledgement

This work is supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Process 119077/2018-3.

¹ Botacin, M; Grégio, A e de Geus, P. "The other guys: automated analysis of marginalized malware," Journal of Computer Virology and Hacking Techniques, pp. 1–12, 2017. [Online]. Available: <http://dx.doi.org/10.1007/s11416-017-0292-8>.

² Bertão, G; Botacin, M; Grégio, A e de Geus, P. "Large scale studies: Malware needles in a haystack," in Anais Estendidos do XVIII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais. Porto Alegre, RS, Brasil: SBC, 2018, pp. 203–212. [Online]. Available: http://portaldeconteudo.sbc.org.br/index.php/sbseg_estendido/article/view/4159