Malware in Pirated Software: Case Study of Malware Encounters in Personal Computers

S. Kumar, L. Madhavan, M. Nagappan and B. Sikdar
Department of Electrical and Computer Engineering
National University of Singapore
Singapore 129788

***Abstract*—Software piracy is a common occurrence, and a significant fraction of the personal computers have some pirated software installed. Cyber-criminals often use pirated software as a vector to spread malware by bundling malicious software with the pirated software. This paper presents the results of a case study that aims to quantify the incidence of malware in pirated software that come bundled with new personal computer purchases. The paper also evaluates the types of malware that are present in the samples in our case study, and the locations in the file system where these malware are detected. The results show that 63% of the samples procured for the case study showed presence of malware and the incidence of malware varies with the geographical location where the sample was procured. Our results also indicate that Trojans and Hacktools are the most prevalent families of malware in our samples.**

***Keywords*-malware; software piracy;**

I. Introduction

Computer malware is a global problem that causes significant economic losses and loss of productivity. Incidences of malware have been increasing consistently at a rate of 50% per year, and the number of unique malware samples has reached 433 million as of 2015 [1]. Malware typically result in theft of personal and sensitive data, spam and denial of service attacks, as well as ransomware. Malware development and exploitation is an industry with suppliers, markets, and service providers ("cybercrime as a service") [1]. Malware bundled with pirated software is one of the common ways by which computers are infected [2]. In many cases, new computing equipment is sold with pirated software that contain malware. Also, malware may infect computing equipment from pirated software downloaded through the Internet or bought from vendors. As a first step to understand the link between pirated software and malware, this paper quantifies the incidence of malware in pirated software that comes bundled with the purchase of new computing equipment.

It is an open secret that pirated software is discreetly (or even openly) sold by vendors of computing equipment and accessories in many countries, and is also easily accessible on the Internet (e.g. through peer-to-peer networks). Pirated software is also sold on fake websites that mislead the buyers about the authenticity of the software, or through classified advertisements. Pirated software is routinely used by cyber criminals to spread malware and users looking for free or significantly cheaper alternatives to legal software often fall prey to infections from malware bundled with pirated software. The most common malware included in such pirated software is usually spyware and keyloggers that target personal information such as bank account and credit card numbers, passwords, and address books. In addition, pirated software may also infect the host computer with viruses, worms, Trojan horses, rootkits, and unwanted Adware. May studies indicate that a significant fraction of computers in the world use pirated software, thus potentially posing serious security risks. For example, the Business Software Alliance (BSA) estimates that 42% of software installed on personal computers in 2011 was pirated [3]. A detailed study of the correlation between pirated software and malware can quantify the security risks and this paper seeks to answer this question.

Many studies in existing literature have investigated the incidence of malware in computers and tried to isolate the reasons behind their spread. A number of existing studies that primarily focus on the relationship between user behavior and malware in their computers have highlighted the impact of technical, social, economic and policy issues on the prevalence of malware. In [4] the authors studied the usage patterns of subjects to identify risk factors for being infected with malware and these include demographic factors (e.g. age and gender) and behavioral factors (e.g. applications used and browsing history). In [5] the authors evaluate the incidence of malicious activity in residential users based on monitoring their network traffic. Along similar lines, [6] evaluates the possibility of predicting the likelihood of users becoming victims of web attacks based on analyzing their web browsing behavior. Security logs from an enterprise were analyzed in [10] to characterize the likelihood of malware encounter among the enterprise personnel. An experimental study that evaluates the impact of human behavior in the spread of Internet based malware is presented in [7]. Demographic factors that affect the success of phishing attacks and the effectiveness of anti-phishing educational material is presented in [8]. In contrast to these studies which primarily look at human behavioral aspects of cyber-security, our work focuses on the link between pirated software and malware. In [9] the authors use Symantec anti-virus telemetry data to analyze international variation in the prevalence of malware and highlight pirated software from peer-to-peer and the Internet as a source of malware. However, it does not quantify the incidence of malware in pirated software.

This paper reports on the results of the malware analysis of 194 personal computers (the "samples" in our study) with pirated software bought in 11 countries. The sam-

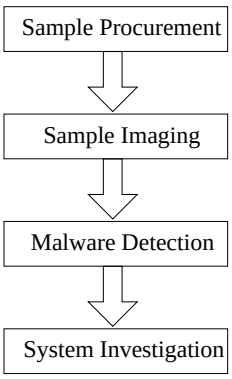

Figure 1. Methodology for malware detection and analysis.

ples were purchased through independent contractors that posed as regular customers of personal computers (both desktop and laptop). Each computer was examined by a set of five anti-virus software to detect the presence of malware. The results of the scans of the anti-virus engines were also examined to determine the type of malware present in these samples as well as the location of the malware in the file system. Our results show that a majority (63%) of the samples in the case study are infected with malware at the point of sale. In addition, our results show that Trojans and Hacktools were the most common families of malware encountered in our samples, and these were usually hidden among the application, system, and program files of the computer.

The rest of the paper is organized as follows. Section II presents the methodology for evaluating the incidence of malware in pirated software. Section III presents the measurement results. Finally, Section IV concludes the paper.

II. Methodology

This section presents the methodology used for obtaining our results. Figure 1 shows various steps involved in the methodology and these include: (i) sample procurement, (ii) imaging of sample hard disk, (iii) malware detection, and (iv) system investigation. The details of each step of the methodology is described below.

1. Sample procurement: This step consists of making purchases of personal computers (both desktops and laptops) to be analyzed for the presence of malware. The sample procurement step has two objectives: procurement of hardware with different specifications and from multiple sources, and procurement of samples from different countries to evaluate the impact of regional biases, if any. The samples purchased for this study consisted of 194 personal computers, including both desktops and laptops. These samples were procured from 11 countries in Asia, Europe, and North and South America. The vendors of these samples were chosen at random and include personal computer (PC) shops in information technology (IT) specific malls and shopping complexes, localities with shops that specialize in computing and IT products and services, as well as standalone shops in street markets. The samples were procured in late 2013 and early 2014.

We employed independent contractors to procure the samples. These contractors acted as "normal walk-in customers" such as students, professionals, home-makers, small business owners etc. when they visited the shops. Thus our focus was on targeting typical personal computer sales environments where walk-in customers interact with the sales staff before making a purchase. In such scenarios, sales staff often offer pirated software (e.g. operating system) as an incentive to increase sales. Globally or nationally recognizable stores as well as direct purchases from original equipment manufacturers were avoided since they usually do not install pirated software in the merchandise they sell. Also, we emphasize that our purchase team did not specifically request computers with pirated software. While the purchase team did discuss brand options, hardware (e.g. processor and memory) specifications, and pricing, in many cases this led to the sellers offering free software as an added incentive to make the sale.

2. Sample imaging: To analyze each sample computer for the presence of malware, we first create an image of the hard disk of each of the samples. The images were created using Microsoft Corporation's disk2vhd software tool and the imaging process consists of making a sector-by-sector copy of the contents of the hard disk. All partitions of the hard disks were selected when creating the images. The main motivation for using images for our analysis is that the contents of the hard disk can be analyzed without the risk of contamination or modification of the original sample. Thus, the scans for malware analysis for each sample are done on a copy of its image. The imaging tool creates Virtual Hard Disk (VHD) versions of the physical disks and these images are subsequently loaded in virtual machine (VM) environments for the system analysis part of the overall methodology. Consequently, any breakout during the analysis or any inadvertent modifications to the sample by the anti-virus engines are limited to the copy of the image.

3. Malware detection: To detect malware in the personal computers purchased for this study, the hard disk image of each sample was scanned with multiple anti-virus software. Each image was scanned by five anti-virus engines: Avira, AVG, Avast, Microsoft Security Essentials, and Kaspersky. For any given sample, a separate copy of the software image was used for each of the five anti-virus engines. A separate copy of the image for each engine ensured that any inadvertent changes to the image by one anti-virus engine does not affect the results of other engines. Also, it ensures that the image scanned by each anti-virus engine for a given sample is identical (and the same as the original).

Each scan by an anti-virus engine for any sample was based on the following rules:

1) The latest definitions and updates for the anti-virus engine were downloaded before each scan.
2) The anti-virus engine's settings were configured to scan all files and directories.
3) The options for automatically removing malware

Table I
NUMBER OF SAMPLES FROM EACH COUNTRY

USA	14
China	20
India	19
Indonesia	19
South Korea	15
Thailand	19
Brazil	19
Mexico	11
Russia	18
Ukraine	20
Turkey	20

was turned off. On the completion of each scan, the malware samples were copied and saved for further investigation.

4) The output of the scan such as the details of the malware identified and their locations was recorded.

Once each sample is scanned by the five anti-virus engines, the results were collated and the number of unique malware in each sample was counted.

4. System investigation: The system investigation stage of the investigation focuses on identifying changes to the system settings during installation or by malware. To conduct the system investigation, each sample image is first attached to a virtual machine in the Microsoft Hyper V virtualization platform. The sample is then investigated to look for the following signs of tampering:

1) The operating system version, product, and activation keys are noted. These keys may be checked against Microsofts database to determine if the operating system is genuine.
2) The version, license and product ID of Microsoft Office, if installed, is noted.
3) The default username and password is noted.
4) A list of all application software installed on the sample is created.

In many cases, there were various errors that were encountered with attaching the image of a sample to a virtual machine. In these situations, the contents of sample were directly used as the hard-disk of a personal computer or laptop. The computer was then powered up and the signs of tampering, as listed above, are noted.

III. PREVALENCE OF MALWARE IN PIRATED SOFTWARE

This section presents the results of our study. To evaluate the link between pirated software and malware, we scanned each of the 194 samples that we acquired by five anti-virus engines. The number of samples obtained from each country in our study is presented in Table I. In addition to the overall results from all samples, we also provide country specific results. The results of these scans and our inferences are as follows.

A. Malware Incidence

Figure 2 shows the incidence of malware in the samples obtained from each country. While the figure shows the incidence of malware on the samples categorized by country, the overall sample rate considering all 194 samples in our case study is 63%. While our results showed the presence of malware in samples from all countries, we note that the incidence of malware is relatively higher in Asian countries. Our results are consistent with other studies such as [11] and [12] which report on cyber-crime and malware rates in China and India, respectively. There are a number of reasons that contribute to the higher incidence of malware in these countries. These include lack of awareness about risks associated with pirated software, monetary reasons, and being specifically targeted by cyber-criminals by bundling malware with pirated software.

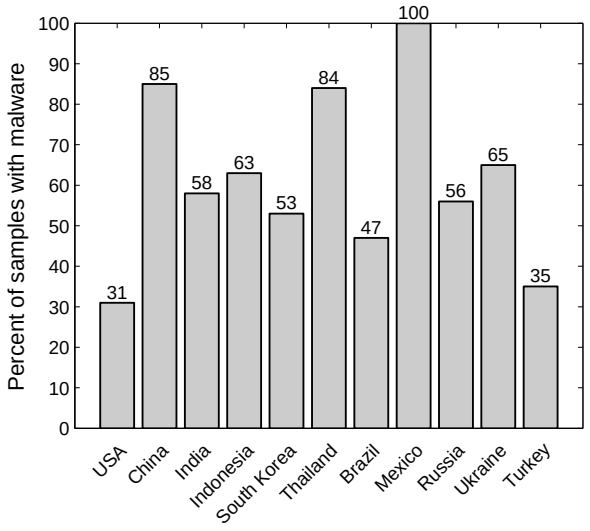

Figure 2. Incidence of malware per country. Overall rate: 63%.

B. Malware Families

Malware vary in the exploits they use, the damage they cause, and they mechanism they employ to spread. Most anti-virus engines classify malware into families such as worms, viruses, adware, trojans etc. Figure 3 shows the number of unique occurrences of malware in our samples, classified into they families they belong to. We note that Trojans and Hacktools are the most common families that were encountered. The frequent encounter of Hacktools suggests the involvement of organized cyber-criminals who use these malware to collect passwords, bank and credit card numbers. Also, the large incidence of Trojans is an attestation to their versatility at causing damage to the infected system and the ability to open back doors that give cyber-criminals remote access to the system.

C. Malware Location

This section presents the results related to the location of the malware in the samples' file system. The locations of the detected malware were obtained from the reports of the anti-virus engines. The locations of the malware are grouped into six categories and Figure 4 shows the relative occurrence of malware in

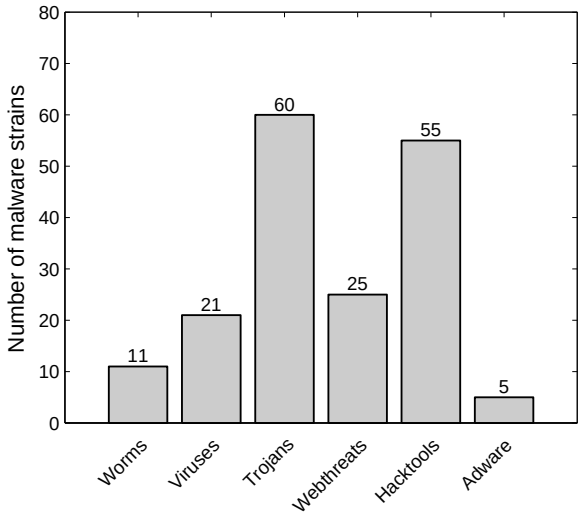

Figure 3. Number of malware strains encountered.

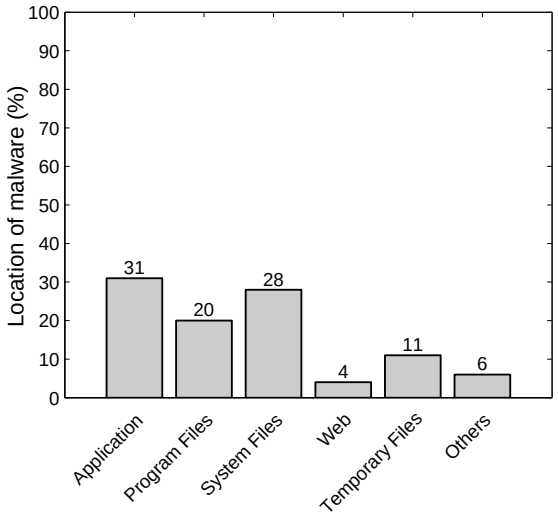

Figure 4. File system location where malware was found.

each of these categories. The results correspond to the observations of the malware from all 194 samples. Our results show that the predominant location of the malware was in files related to applications, followed by files related to the Windows operating system (in the directory `C:\Windows\System`). The Program Files correspond to the directory `C:\Program Files\`, the malware in the web related location corresponds to the directory `Temporary Internet Files`, the Temporary Files correspond to the directory `C:\Users\User Name\AppData\Local\Temp`, and Others correspond to malware found in all other locations. Some existing studies such as [13] report that web attacks are the most popular methods employed by malware authors to compromise personal computers. However, these require some amount of user interaction before the computers are compromised. On the other hand, the results reported in this paper are for new computers that have not been used, and this may explain some of the differences in our reported results. There is also some in the malware locations as a function of the geographical location (e.g. personal computers in India were more likely to have malware in applications related to games).

D. Limitations

Owing to the sample size and the nature of sample collection, the results and conclusions of this paper are subject to a number of caveats. Firstly, while the overall sample size is 194, the number of samples for none of the countries exceeded 20. Consequently, the results presented in this paper cannot be deemed to be representative of any country. However, the results do provide an insight into the nature and prevalence of malware in pirated software. Also, the overall sample size is large enough to provide inferences that are statistically significant. Secondly, it is well-known that no anti-virus engine can detect all malware and our results showed that the malware detection capability of each of the five anti-virus engines is different. Consequently, the malware detection rates reported in this paper should be interpreted as lower bounds on the overall rates. Thirdly, there are additional channels of procurement of personal computers that this study did not consider. For example, we did not procure any samples through the sale of second-hand (previously used) computers. Finally, there are unknown factors that affect similar studies and the generality of their results [10]. However, studies such as these are necessary and act as building blocks that help us develop a better understanding of the factors that affect the prevalence and spread of malware, and establish the link between software piracy and malware.

IV. Conclusion

This paper presented a case study of the presence of malware in pirated software. The case study examined 194 personal computers bought from 11 countries with five anti-virus engines. The results showed considerable geographical variation in the incidence of malware and an overall infection rate of 63%. Our results also indicated that Trojans and Hacktools were the most common families of malware that were encountered. The case study highlights the link between pirated software and malware and the possibility of cyber-criminals using distribution of pirated software to spread malware.

References

[1] McAfee Labs, *McAfee Labs Threats Report,* August 2015.

[2] M. Kammerstetter, C. Platzer, and G. Wondracek, "Vanity, cracks and malware: Insights into the anti-copy protection ecosystem," *Proc. Computer and Communication Security Conference,* pp. 809-820, October 2012.

[3] Business Software Alliance, *Shadow Market: 2011 BSA Global Software Piracy Study,* May 2012.

[4] F. Levesque, J. Nsiempba, J. Fernandez, S. Chiasson and A. Somayaji, "A clinical study of risk factors related to malware infections," *Proc. ACM CCS,* pp. 97-108, Berlin, Germany, November 2013.

[5] G. Maier, A. Feldmann, V. Paxson, R. Sommer and M. Vallentin, "An assessment of overt malicious activity manifest in residential networks, *Detection of Intrusions and Malware, and Vulnerability Assessment,* pp. 144-163, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[6] D. Canali, L. Bilge and D. Balzarotti, "On the effectiveness of risk prediction based on users browsing behavior," *Proc. ACM Symposium on Information, Computer and Communications Security,* pp. 171-182, Kyoto, Japan, June 2014.

[7] K. Onarlioglu, Y. Yilmaz, E. Kirda and D. Balzarotti, "Insights into user behavior in dealing with internet attacks," *Proc. Network and Distributed System Security Symposium,* San Diego, CA, February 2012.

[8] S. Sheng, M. Holbrook, P. Kumaraguru, L. Cranor and J. Downs. "Who falls for phish?: A demographic analysis of phishing susceptibility and effectiveness of interventions," *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems,* p. 373, Atlanta, GA, April 2010.

[9] G. Mezzour, K. Carley and L. Carley, "An empirical study of global malware encounters," *Proc. ACM Symposium and Bootcamp on the Science of Security,* New York, NY, 2015.

[10] T.-F. Yen, V. Heorhiadi, A. Oprea, M. Reiter and A. Juels, "An epidemiological study of malware encounters in a large enterprise," *Proc. ACM CCS,* pp. 1117-1130, 2014.

[11] N. Kshetri, "Cyber-victimization and cybersecurity in China," *Communications of the ACM,* pp. 35-37, vol. 56, no. 4, April 2013.

[12] S. Mishra, S. Dhir and M. Hooda, "A Study on Cyber Security, Its Issues and Cyber Crime Rates in India," *Proc. ICICSE,* pp. 249-253, Hyderabad, India, August 2015.

[13] Symantec Corporation, *Internet Security Threat Report (ISTR),* vol. 18, April 2013.