Methodology for a Field Study of Anti-Malware Software

Fanny Lalonde Lévesque[1], Carlton R. Davis[1], José M. Fernandez[1]
Sonia Chiasson[2], Anil Somayaji[2]

[1] École Polytechnique de Montréal, Montréal Canada
{fanny.lalonde-levesque, carlton.davis, jose.fernandez}@polymtl.ca
[2] Carleton University, Ottawa Canada
{chiasson, soma}@scs.carleton.ca

Abstract. Anti-malware products are typically evaluated using structured, automated tests to allow for comparison with other products and for measuring improved efficiency against specific attacks. We propose that anti-malware testing would benefit from field studies assessing effectiveness in more ecologically valid settings. This paper presents our methodology for conducting a 4-month field study with 50 participants, including discussion of deployment and data collection, encouraging retention of participants, ethical concerns, and our experience to date.

Keywords: anti-malware testing, field study, user study

1 Introduction

How should the effectiveness of anti-malware software be assessed in practice? Current strategies typically involve automated testing against standard datasets, sometimes with automated user profiles to imitate user interaction with security messages [1]. Even with the more advanced tests that include user profiles, this assumes that users' behaviour and all of the variables affecting their computing environments can be predicted and reflected in these automated profiles.

We suggest that many infections are due to direct or indirect user actions that allow malware to infect a system. These actions (or inactions) may occur immediately prior to infection, weeks or months prior to infection, or may even occur over time so that a combination of actions lead to a vulnerable system state. These situations would not be accurately reflected in automated testing, nor would they be identifiable through traditional lab-based user testing.

One alternative is to conduct long-term field studies of anti-malware software with real users in more ecologically valid settings. By monitoring real usage over time, one can gain a better understanding of how anti-malware systems are used and how external factors influence their effectiveness. However, a large number of confounding variables exist which significantly complicates the data analysis. In our study, we wanted to provide a common and controlled "clean slate" to begin the experiment, but somehow allow users to take ownership over their

system and use it as they would normally, while we monitor the system for signs of infection.

In this paper, we present our approach to conducting a field trial of an anti-malware product. Section 2 summarizes related work in anti-malware testing and conducting field studies of security products. Section 3 describes our methodology for conducting the trial, including our approach to selling laptops and reimbursing participants for their purchase throughout the study. Section 4 discusses how we addressed ethical and privacy issues, while Section 5 highlights our experience with this ongoing study. The paper concludes with a discussion of our anticipated analysis and brief description of a larger scale follow-up study.

2 Background

Although there are currently several methods for evaluating anti-malware products [7], they do not reflect the performance of products in real life. Typical evaluation methods are based on scanning collected or synthesized malware along with legitimate programs. While such approaches can measure raw detector accuracy, they cannot take into account factors such as user interactions, evolving threats, and different environments. One major issue is that the sample collection is often too small, inappropriate, and unvalidated [8, 9]. Even with a well-maintained malware collection, testing against such data sets has become unreliable due to the increased dynamic nature of malware. To partially address this issue, Vrabec and Harley [13] proposed emulating user interaction with the system and creating user-specific testing scenarios.

In the broader security community, field studies of computer security are frequently advocated but are still relatively uncommon in the literature, likely due to the costs, time demands, and potential security and privacy risks to users. Recent field studies of security software have mostly involved evaluating the use of authentication mechanisms [3, 6, 4]. In 2009, Somayaji et al. [11] introduced the concept of computer security clinical trials. The conceptual proposal was to evaluate security products using methods and controls similar to those used in clinical trials of medical products, but no studies have been conducted thus far.

Ethnographic studies examine usage of security systems in the field, but use qualitative methods such as interviews, diaries, and observation to understand how and why participants interact with computer systems. Botta et al. [2] conducted an ethnographic study of security professionals, Rode [10] examined parental behaviour in protecting children's online safety, Wash [14] used interviews to understand users' mental models of security, and De Luca et al. conducted a field observation of ATM usage to evaluate PIN usage [5].

While some of the above studies mention anti-malware usage as a security measure taken by users, it is not the focus of these studies. To our knowledge, there are no published user studies focusing specifically on anti-malware usage.

3 Study Description

The goals of this study are to: (1) determine how phenomena such as the configuration of the system, the environment in which it is used and user behavior can affect the probability of infection of a system; (2) develop an effective methodology to evaluate anti-malware products in real-world environment; (3) determine how malware infects computer systems, and identify sources of malware infections. The study includes monitoring real-world computer usage through diagnostics and logging tools, monthly interviews and questionnaires, and in-depth investigation of any potential infections.

We are conducting a 4-month field study with 50 participants that were recruited through posters and newspaper advertisements on campus. A short online intake questionnaire was used to collect initial demographic information. Using these profiles, we categorized interested volunteers and randomly chose a sample from each category in order to have a diverse and representative sample of users that include students and employees from various fields.

3.1 Equipment

We supplied laptops with identical configuration to the participants. The following software was installed: Windows 7; the antivirus (AV) product to be evaluated; diagnostics tools, such as HijackThis, ProcessExplorer and Autoruns; and custom Perl scripts which we developed. We utilised the scripts to automate the execution of the tools as well as for compiling statistical data regarding the system configuration, the environments in which the system is used, and the manner in which the system is utilised. The AV product is centrally managed on our server. An AV client installed on the laptops sends relevant information to the server about any malware detected or suspected infections as they occur.

Before deployment, we benchmarked the laptops by running the diagnostics tools and recording the output. The information included: a hash of all files plus information about whether the files were signed; a list of auto-start programs; a list of processes; a list of registry keys; and a list of browser helper objects.

3.2 Procedure

The study consisted of 5 in-person sessions: an initial session where participants received their laptop and instructions, followed by monthly 1-2 hour sessions where we performed analysis to determine if the laptop was infected.

Participants initially purchased the laptops from us at a reduced rate; it was theirs to keep after the study. To encourage the participants to remain in the study, we paid them to attend the monthly in-person sessions. If participants complete all required sessions, the entire cost of the laptop would be reimbursed, along with an additional honorarium. We encouraged participants to configure their laptop as they desired and use it as they would normally use their own computer. The only restrictions applied during the experiment were that the

participants do not format the hard drive, do not replace the operating system, and not install any other AV product on the laptop.

Each month, participants booked an appointment via an online calendar system hosted on our website. During these monthly sessions, participants completed an online questionnaire about their computer usage and experience, while the experimenter collected the local data compiled by the automated scripts. The questionnaire was intended to assess the participant's experience with the AV product and gain insights about how the laptop was used.

The data compiled by our scripts included, but was not limited to, the list of applications installed, the average number of hours per day the laptop is connected to the Internet, and the number of web sites visited. Diagnostics tools were also executed on the laptop to determine if infection was suspected. If the AV product detected any malware over the course of the month, or if our diagnostics tools indicate that the laptop may be infected, we requested additional written consent from the participant to collect data that will help us identify the means and the source of the infection.

Before the last visit, participants completed an online survey about their experience during the study. The aim of this exit survey was to identify activities or mindset that may have unduly influenced the experimental results. We chose to administer the survey apart from the in-person session in case participants were more comfortable revealing such information while not in the presence of the experimenter. In the last session, we requested that participants keep the experiment data stored on their laptops for an additional three months, so that if we discover that further analysis is necessary, we can contact them and seek their permission to collect and analyse the relevant data. Nonetheless, we provided a procedure for deleting the diagnostic tools and the scripts, as well as the experiment data stored on their laptop. We also explained that residual data may still remain on the laptop even after the experiment data is deleted. If they wanted to completely remove all traces of the experiment from their laptop, we referred them to external resources for re-imaging the laptop.

4 Ethical and Privacy Considerations

This study was reviewed and approved by the Computer Risks Evaluation Board (CREB) and the Research Ethics Board (REB) of École Polytechnique de Montréal. Ethical and privacy guidelines were of particular concern because the experiment involved the collection of personal data over an extended period of time.

To preserve anonymity, each participant was assigned an identification number such that the identity of the participant was not linked to any data during analysis. No personally identifiable information, such as usernames and passwords were collected, content of personal documents stored on the computer were not examined, and no exact URLs were collected (only aggregate data about categories of web sites such as "social networking" and "gaming").

Because the study involved malware, necessary precautions were taken to protect the university's infrastructure as well as that of the users. For example, in the event that an infection could not be cleaned by the AV product, we relayed

the relevant details to the AV company. The company developed and provided a product update to detect and remove the infection. This update was applied to participants' laptops as part of regular automated software updates.

5 Experience to date

The study officially started in November 2011. The first step was to configure the laptops and meet all 50 participants individually to provide instructions, have them sign the consent form, and pay for their laptop. As noted earlier, the full cost of the laptop will be reimbursed, with an additional honorarium, provided that the participant attends all four monthly visits. Partial reimbursement will be provided if only some of the sessions are completed.

The study is ongoing. Participants have their laptops and the AV clients have been communicating with our AV server. Thus far (November 2011), a total of 18 malware incidences have been reported on eight of the laptops. Also, we know that at least one incidence is an actual infection; the participant informed us that a program on his laptop requested that he pay money to upgrade his AV software. The responsible program was confirmed to be a known malicious scareware that pretends to be legitimate security software [12]. All incidences will be explored when these participants return for their first monthly session.

The number of incidences in the first month is much higher than anticipated. This initial spike may be because participants were installing software and customizing their computer. We will be closely exploring and analysing the data collected during the first monthly sessions. It remains to be seen whether this rate of infection will persist throughout the rest of the study.

Most participants have expressed a high level of willingness to collaborate and some have even shown scientific interest in the study. Surprisingly, some participants asked us how they should act to get their laptop infected, to which we responded that they should use their laptop normally. In the event that participants need assistance, we provided them a telephone number and an email address that they could use to contact us. Other than the participant whose laptop has been infected with malware, only a few participants have contacted us via email to obtain support, and none of them has contacted us via telephone.

6 Conclusion and Future Work

This study is intended to demonstrate what we believe is a more effective way of evaluating anti-malware products: the main conjecture being that it is imperative that actual users be involved in the evaluation process, and that they use the products in realistic environment over an extended period of time. Our study is in progress, but results so far point to a rich data set that will provide evidence for how user behaviour and environments of use affects incidences of malware.

We will be analyzing data on a monthly basis to ensure that we are collecting appropriate data and to determine if any modifications are necessary. Overall, we intend to perform in-depth statistical analysis to determine whether there is a correlation between user behaviour and incidences of infection, as well as probing specific incidences to fully understand the causes of infection.

We will use our findings to inform a second, larger study examining specific variables and confirming the results from the first study. This second study will be designed to determine which factors (i.e. type of AV product or user behaviour) has the most impact on incidences of infection of a system. It will compare multiple AV products, more participants will be involved, and the study will take place over a longer period of time, likely over 6 to 12 months. We hope the results of this follow-on work will help inform the design of future consumer-level security products.

7 Acknowledgement

This project has been funded by the NSERC Internetworked Systems Security Network (ISSNet), MITACS and Trend Micro.

References

1. Anti-Malware Testing Standards Organization: AMTSO testability guidelines. Tech. rep., http://www.amtso.org/documents.html (May 2011)
2. Botta, D., Werlinger, R., Gagné, A., Beznosov, K., Iverson, L., Fels, S., Fisher, B.: Towards understanding it security professionals and their tools. In: ACM Symposium On Usable Privacy and Security (SOUPS). ACM (2007)
3. Brostoff, S., Sasse, M.: Are Passfaces more usable than passwords? A field trial investigation. In: British Human-Computer Interaction Conference (HCI) (2000)
4. Chiasson, S., Biddle, R., van Oorschot, P.C.: A second look at the usability of click-based graphical passwords. In: ACM Symposium on Usable Privacy and Security (SOUPS) (2007)
5. De Luca, A., Langheinrich, M., Hussmann, H.: Towards understanding ATM security: a field study of real world ATM use. In: ACM Symposium On Usable Privacy and Security (SOUPS) (2010)
6. Florencio, D., Herley, C.: A large-scale study of WWW password habits. In: ACM World Wide Web Conference (WWW) (2007)
7. Gordon, S., Ford, R.: Real world anti-virus product reviews and evaluations - the current state of affairs. In: 19th National Information Systems Security Conference (NISSC) (1996)
8. Harley, D., Lee, A.: Who will test the testers? In: 18th Virus Bulletin International Conference (2008)
9. Košinár, P., Malcho, J., Marko, R., Harley, D.: AV testing exposed. In: 20th Virus Bulletin International Conference (2010)
10. Rode, J.A.: Digital parenting: designing children's safety. In: British HCI Conference (BCS-HCI) (2009)
11. Somayaji, A., Li, Y., Inoue, H., Fernandez, J.M., Ford, R.: Evaluating security products with clinical trials. In: Workshop on Cyber Security Experimentation and Test (CSET) (2009)
12. Stone-Gross, B., Abman, R., Kemmerer, R.A., Kruegel, C.: The underground economy of fake antivirus software. In: Workshop on the Economics of Information Security (WEIS) (2011)
13. Vrabec, J., Harley, D.: Real performance? In: EICAR Annual Conference (2010)
14. Wash, R.: Folk models of home computer security. In: ACM Symposium On Usable Privacy and Security (SOUPS) (2010)