A Clinical Study of Risk Factors Related to Malware Infections

Fanny Lalonde Lévesque
École Polytechnique de Montréal
Montréal, Canada

Jude Nsiempba
École Polytechnique de Montréal
Montréal, Canada

José M. Fernandez
École Polytechnique de Montréal
Montréal, Canada

Sonia Chiasson
Carleton University
Ottawa, Canada

Anil Somayaji
Carleton University
Ottawa, Canada

ABSTRACT

The success of malicious software (malware) depends upon both technical and human factors. The most security conscious users are vulnerable to zero-day exploits; the best security mechanisms can be circumvented by poor user choices. While there has been significant research addressing the technical aspects of malware attack and defense, there has been much less research reporting on how human behavior interacts with both malware and current malware defenses.

In this paper we describe a proof-of-concept field study designed to examine the interactions between users, anti-virus (anti-malware) software, and malware as they occur on deployed systems. The 4-month study, conducted in a fashion similar to the clinical trials used to evaluate medical interventions, involved 50 subjects whose laptops were instrumented to monitor possible infections and gather data on user behavior. Although the population size was limited, this initial study produced some intriguing, non-intuitive insights into the efficacy of current defenses, particularly with regards to the technical sophistication of end users. We assert that this work shows the feasibility and utility of testing security software through long-term field studies with greater ecological validity than can be achieved through other means.

Categories and Subject Descriptors

K.6.5 [**Management of computing and information systems**]: Security and Protection—*Invasive software*; K.4.2 [**Computers and Society**]: Social Issues—*Abuse and crime involving computers*

CCS'13, November 4–8, 2013, Berlin, Germany.

http://dx.doi.org/10.1145/2508859.2516747.

Keywords

anti-virus evaluation; malware infection; field study; user behavior; clinical trial; risk factors

1. INTRODUCTION

The growth of malicious activity on the Internet is unabated. Beyond sheer numbers of different samples, what has made the problem of malware detection and prevention more complex is the variety of threats and attack vectors that modern malware authors deploy to infect their victim's computers. In the early 2000's massive computer worms infected large swaths of the Internet in a few hours through open network ports with almost no human intervention. Today, however, we have a situation closer to that of an earlier era where threats were propagated through diskettes, email attachments, and innocuous-looking Trojans. Malware infections now occur largely because users are enticed to take an action that leads to their computers being infected. Today, this could be opening an email attachment (still a popular classic), visiting a malicious Web site, or even willingly installing a piece of software whose true intention they ignore (move over cuteware, here comes the codec).

Meanwhile, anti-virus (AV) products have had to evolve as well. The signature-based file-scanning engines that used to be the core technology of AV products have been complemented by multiple layers of protection, including identification of hazardous URLs, reputation-based software classification, system behaviour monitoring, etc. Computers are no longer stand-alone machines that need to be protected as such, and what used to be a security problem —their connectedness— is increasingly being leveraged by AV vendors to better protect their customers. Periodic signature file updates are being replaced by on-demand resource lookups on databases in cloud infrastructures, who are in turn fed by the continuous reporting of millions of AV client installations. AV products have thus evolved into complex pieces of "anti-malware" software, or rather complex software systems involving several semi-independent components with which, in some cases, the user must interact. While many AV vendors try to make the installation and operation of their product as transparent to the user as possible (or rather as "user-proof" as possible), the truth is that how well the AV operates depends on its user. This depen-

dence is not only due to user configuration of the many features of AV products, but also to other user-driven factors such as how often the machine is connected to the Internet, how often its software and signatures are updated, but most importantly, how the humans in front of the machine are interacting with the computer and the Internet when confronted with situations where their actions could lead to infection.

In other words, the operating environment of both malware and AV products not only includes the machine it is trying to penetrate/protect, but also the network that connects it to the rest of the world and the *user* that sits in front of it. Indeed, the human is part of the operating environment of the machine, including the software that attempts to execute on it or protect it. It thus seems natural to adopt a *Homo in machina* (human in the machine) approach to evaluate not only the performance of AV products but also the susceptibility of users to getting their machines infected.

This change of paradigm is fundamental if we want to better understand what role user actions really play in the process of infection. In particular, it becomes paramount to understand how user characteristics, such as demographics, perception of threat, computer literacy, and user actions may affect the risk of infection by malware. For instance, while we might hypothesise that users who spend a lot of time browsing less reputable Web sites are more prone to malware infections based on the fact that they are often used by criminals to spread malware, it is important to be able to quantitatively confirm such hypotheses. This question goes beyond the performance evaluation of the AV alone, but of the AV *with* the human or of the human by itself.

This philosophy of *Homo in machina* is also in sharp contrast with current AV evaluation methods, largely based on automated tests in controlled environments, where machines installed with an AV product are subject to various infecting stimuli. In order to better assess the effectiveness of the multi-layered protection of modern AV, traditional file-scanning tests (also called "static" or "on-demand" tests in the AV industry) have largely fallen out of favour and are being replaced with tests involving known bad URLs (also called "dynamic" or "real-world"). While this latter type of tests does evaluate the performance of AV products as a whole (and not that of individual features), the truth is that they are not "real world" in that the effect of the main contributing factor in infections, i.e. the human factor, is not being measured. In addition, the results of these tests are often biased because the stimuli chosen by the tester could very well not be representative of what is typically experienced by the average user; this is often refereed to as the *sample selection problem*. Furthermore, such test do not measure the effectiveness of AV products at communicating risk effectively to the user and how they can affect user behaviour.

What we need then is to test using a methodology that can evaluate the interactions between humans, malware, and AV products (and other security software) with much greater ecological validity. Any testing methodology will have some impact on the phenomena being tested; to minimise this impact, testing needs to be done in an environment as close as possible to that of normal usage. One potential way to achieve such ecological validity is through conducting clinical trials of software as we proposed in 2009 [24]. With clinical trials, security software is installed and monitored on systems in regular use by regular users. Data is then gathered on the performance of the security software in protecting the system and on how the user interacted with the system during this time period. By correlating user behaviour, application use, and security software activity, we can gain insights into the interactions between all three in an ecologically valid context.

Thus, in order to address both the question of how to better evaluate AV and how to understand the influence of the human factor in infections, we decided to conduct such a clinical trial of anti-virus. In this paper we report on the first study of this kind, conducted at the École Polytechnique de Montréal from October 2011 to April 2012, involving 50 participants that used their own computers in everyday life during a 4-month period. The data collected during the study attempted to take into consideration many of the reasonable factors that could influence infection such as user profiling, user behaviour, host configuration and environment. In addition, the study collected data that would allow us to evaluate the performance of the AV and in some cases determine the causes of infections. In this paper, however, while we do present results on AV performance, we concentrate on finding and understanding the correlations between these factors and occurrences of infection, in order to determine which ones could be identified as risk factors. Detailed performance analysis of the various AV protection mechanisms through the determination of the causal mechanisms of infection is out of scope of this paper and will be the object of future research based on further analysis of the data produced by this experiment.

We should note that other methodologies, such as bench tests, cognitive walkthroughs, and lab-based user testing can give finer-grained insight into user and system behaviour than is possible in clinical trials, and can do so at lower cost. As we show here, though, clinical trials can gain insight into how systems perform and how users interact with them in practise, something that cannot be addressed with these other methodologies.

The remainder of the paper is organised as follows. Section 2 presents the related work. Section 3 describes the study and provides a summary of its methodology, a detailed description of which is also given in a previous methodology publication [14]. In Section 4, we describe and discuss in detail the results of the study in terms of threat detections by AV, missed detections, and identification of potential risk factors related to user characteristics and behaviour. While the study itself and a few of the preliminary results, especially on AV performance, were already presented at a industry conference shortly after the conclusion of the study [15], Section 4 contains a more in-depth and complete statistical analysis and interpretation of the study results, with an emphasis on risk factors. We discuss limitations, applications, and future work in Section 5 and conclude in Section 6.

2. RELATED WORK

Numerous studies evaluating the performance of AV products and the influence of user interactions factors on IT security have been conducted in recent years. There are currently several methods for evaluating anti-malware products used in the AV industry [4, 7, 8], but they do not reflect the performance of products in real life. Typical evaluation methods conducted by commercial testing labs (e.g. [2, 19]) are based on scanning collected or synthesised malware

samples along with legitimate programmes. 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 not validated [9, 12]. Even with a well-maintained malware collection, testing against such data sets has become unreliable due to the increasingly dynamic nature of malware. To partially address this issue, Vrabec and Harley [27] proposed emulating user interaction with the system and creating user-specific testing scenarios. Another alternate method for evaluating the performance of desktop AV is through the use of on-demand detection tools, that in addition to detecting installed threats can detect whether an AV was installed and which one. For example, the security company SurfRight made public a report [25] describing a 55-day study conducted at the end of 2009 involving more than 100,000 machines that used their product. This report included statistics of detections missed by the installed desktop AV products. Finally, infection statistics based on self-reporting of (perceived) security incidents by users, can be obtained from user surveys and be used to estimate AV performance, such as in the report published by Eurostat —the statistical office of the European Union— based on the 12-month reporting period in 2010 [6]. Unfortunately, self-reported rates of infection are probably quite inaccurate. Finally, the experiment proposed in [24] eliminates many of these problems and potential inaccuracies by adapting the concept of clinical trials to the computer security domain. The initial proposal was to evaluate security products using methods and controls similar to those used in clinical trials of medical products.

Other field studies in computer security have been conducted following an ethnographic approach. This type of approach explores the impacts of the manners, the customs, and the social, physical, and fiscal environment of users when they are facing computer security decisions. It primarily uses qualitative methods such as surveys and observations to understand how and why participants interact with computer systems. For example, Botta *et al.* [3] conducted an ethnographic study of security professionals. Rode [20] used this approach to examine parental behaviour in protecting children's on-line safety. The study showed that there are a host of security threats of which the children are not aware and provides an overview of parental rules and strategies for keeping children safe. Wash [28] used interviews to understand users' mental models of security. He identified eight 'folk models' of security threats that are used by home computer users to decide what security software to use, and which expert security advice to follow. De Luca *et al.* conducted a field observation of ATM users to evaluate PIN usage [5]. This study shows that contextual factors on security such as distractions, physical hindrance, trust relationships, and memorability have a big influence on PIN-based ATM use. Lastly, Sheng *et al.* focus on susceptibility of users against phishing attacks [21]. They conclude from the results of an on-line study with 1001 users that prior exposure to phishing education is associated with less susceptibility to phishing, suggesting that phishing education may be an effective tool. They also found that age is a contributing risk factor. Indeed, young people aged 18 to 25 are also more susceptible to phishing. This last assertion is confirmed by a study conducted by Milne *et al.* [16] that concluded, through a survey of 449 on-line buyers (also students), that age affected the behaviour of subjects when faced with a potential computer threat. Ngo and Paternoster [17] used the results of a survey based on the self-assessment of 295 students, to deduct that age is a significant risk factor of infection, with older respondents being less likely to get infected.

Another methodological approach is that based on the concepts and methods of Epidemiology. This approach, similarly as the ethnographic approach, uses statistical methods. However, it refers to the threat and how it spreads, while ethnography refers to environmental and demographic factors that influence the user when faced with this threat. By analogy to the field of health, the concepts that underlie this approach are: infection, detection, disinfection, quarantine, epidemics, environment control, etc. [22]. Its methods are borrowed from the biological sciences in order to determine the likely causes and risk factors for infection, understanding the spread of malware and, where appropriate, the methods to remedy it. This approach has been used in many studies. We summarise three of them here. First of all, Carlinet *et al.* used it to analyse the behaviour of ADSL customers and identify customer characteristics that are risk factors for malware infection. The study showed that using the Windows operating system and heavy usage of Web applications and streaming are major risk factors of malware infection. Unfortunately, the study does not say anything about the type of Web sites as a risk factor. Secondly, Kephart and White [10] attempted to adapt the epidemiological approach to determine the conditions under which virus epidemics are likely to occur, and in cases where they do, they explore the dynamics of the expected number of infected individuals as a function of time. They concluded that there is a threshold of rate of infection from which we can speak of an epidemic. Thirdly, Kondakci [11] used a stochastic model to analyse the epidemic states of infected computers and determine the state probabilities of susceptible populations. He shows that there is more than one state transition. A healthy but susceptible computer can become infected by a virus, can then become a transmitter, and can also return to a healthy state.

Regarding the impact of a user's domain of expertise on risk of infection, Solic and Ilakovac [23] conducted studies with two groups of participants: one group consisting of 19 doctors and another of 20 professors from an engineering school, who anonymously answered a questionnaire on their behaviour when faced with threats to their computer and the consequences that ensue. This study concluded that the domain of expertise does not have a major impact on user behaviour when it comes to threats and security risk.

On the other hand, the level of expertise or non-expertise in computer security does influence users' perception of security risk, as demonstrated by Ashgarpour *et al.* [1]. Two experiments were conducted: a quantitative study used to approximate the mental models of computer users with regards to common security risks, and qualitative interviews with experts and non-experts. The results of these two experiments suggest that the mental models of people concerned with security risks are strongly correlated to their computer security level of expertise.

In summary, we note that all studies reviewed cover separately one or a few of the factors described above. However, none of these studies cover all of these factors or study the relationship between them. Thus, one of our main contribu-

tion is to use the clinical trial method to analyse the impacts of all these individual factors on infection by malware to gain a better understanding of the causes of these infections.

3. STUDY DESCRIPTION

This first study of its kind was conducted as a 4-month proof-of-concept study involving only 50 participants in order to prove its feasibility as an alternative experimental performance evaluation approach in computer security. The details of the methodology have been published elsewhere [14], but we nonetheless provide a brief summary here. The study monitored real-world computer usage through diagnostics and logging tools, monthly interviews and questionnaires, and in-depth investigation of any potential infections. The study had the following goals:

1. Develop an effective methodology to evaluate anti-virus products in real-world environment;

2. Determine how malware infects computer systems and identify source of malware infections;

3. Determine how phenomena such as the configuration of the system, the environment in which the system is used, and user behaviour affect the probability of infection of a system;

The 50 participants were recruited through posters and newspaper advertisements on the Université de Montréal campus (where the École Polytechnique is located). A short on-line questionnaire was used to collect initial demographic information. Using these profiles, we categorised interested volunteers based on their gender, age group, status and field of work/study. We randomly chose a sample from each category in order to have a diverse and representative sample of users that included students and employees from various fields.

3.1 Ethical and Privacy Considerations

Since the study involved human subjects, the entire project had to undergo strict review by the *Comité d'évaluation des risques informatiques* (CÉRI) and the Ethics Review Board of the university. The Board imposed certain restrictions on the study such as limits to the type of (potentially) personal information kept, the length of time data could be kept, the purpose of the research, and adequate remuneration for participation in the study.

All raw data and statistics generated during the experiment were anonymised, as they were only identifiable through a unique number attributed to the laptop. Only the project leader knew what subject corresponded to which number, and this only for administrative purposes. This personal information was destroyed three months after the end of the study. It is therefore not possible to associate collected information with the identity of the subject.

All data collected was kept in a locked cabinet in a high-security zone, which was protected with three-factor authentication (biometrics, PIN and ID card). This work zone is completely isolated from the Internet and the university network, and only personnel authorised within the context of this project had access to the data. The security policy of the laboratory was also applied to the deletion of all personal data related to the experiment. This policy applies to all information whether on paper or electronic media, and conforms with Government of Canada standards.

The use of collected data during this experiment was bound to the stated research objectives of the project. Nonetheless, in circumstance where the law imposes it, such as the inadvertent discovery of information leading a reasonable person to believe that a (serious) crime has been committed or is about to be committed, we would have had to report this information to the appropriate authorities (law enforcement agencies, etc.).

Furthermore, given that the experiment required the handling of malware files, special precautions were taken in order to protect the university's IT infrastructure. For example, all files identified as potentially malicious were encrypted before being stored in the high security zone of the lab.

3.2 Equipment

The laptops that were provided to the subjects all had identical configurations, with the following software installed: Windows 7 Home Premium; Trend Micro's Titanium Maximum Security (Trend Micro's premium AV product for home users); monitoring and diagnostic tools including HijackThis, ProcessExplorer, Autoruns, SpyBHORemover, SpyDLLRemover, tshark, WinPrefetchView, WhatChanged; and custom Perl scripts developed for this experiment.

These scripts automated the execution of the tools and compiled statistical data about system configuration, the environments in which the system is used, and the manner in which the system is used. The data compiled by our scripts included: the list of applications installed and the list of applications for which updates are available; the number and the type of web sites visited; the number and the type of files downloaded; the list of browser plug-ins installed; the number of different hosts to which the laptop communicates per day; the list of the different locations from which the laptop establishes connection to the Internet; the average number of hours per day the laptop is connected to the Internet; and the average number of hours per day the laptop is on.

The AV product was centrally managed on our own server, in a manner similar as is usually done for corporate installations to centralise distribution of signature file updates. All the AV clients installed on the laptops were thus sending relevant information to our server about any malware detected or suspected infections as they occured.

Before deployment, we benchmarked the laptops by running tools and recording the output. The recorded 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; a list of browser helper objects (BHO); a list of the files loaded during the booting process; and a list of the pre-fetch files.

In order to avoid biases in user behaviour and at the same time limit the liability of the university, the laptops were sold to the participants at an advantageous, below retail-market price, with laptops staying in their possession at the end of the study.

3.3 Experimental Protocol

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.

To encourage the participants to remain in the study until its end, we paid them to attend the monthly in-person sessions. If participants completed all required sessions, the entire cost of the laptop would be reimbursed, along with an additional compensation. 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 not format the hard drive, not replace the operating system, not create a disk partition, not install any other AV product on the laptop, and not delete our software and tools.

Each month, participants booked an appointment via an on-line calendar system hosted on our server. During these monthly sessions, participants completed an on-line questionnaire about their computer usage and experience. The questionnaire was intended to assess the participant's experience with the AV product and gain insights about how the laptop was used. Meanwhile, the experimenter collected the local data compiled by the automated scripts. Diagnostics tools were also executed on the laptop to determine if an infection was suspected. If the AV product detected any malware over the course of the month, or if our diagnostics tools indicated that the laptop may be infected, we requested additional written consent from the participant to collect specific data, such as the browser history, the tshark log files (i.e. network traffic data), and the suspected file(s), in order to help us identify the means and the source of the infection.

In the last visit, participants completed an on-line exit survey about their experience during the study. The aim of this final survey was to identify activities or mindsets that may have unduly influenced the experimental results. We requested that participants keep the experiment data stored on their laptops for an additional three months, so that if we discovered that further analysis was necessary, we could contact them and seek their permission to collect and analyse additional relevant data. Finally, we provided them with a procedure for deleting the diagnostic tools and the scripts, as well as the experiment data stored on their laptop.

4. RESULTS AND DISCUSSION

Our analysis focuses on several aspects, first examining the number and type of detections found during the study, and secondly by exploring how user characteristics and behavioural patterns may have affected the likelihood of getting infected.

4.1 Threats Detected by AV

During the 4-month study, 380 files were detected on 19 different user machines by the AV product being evaluated. However, some of these files were detected twice or more on the same user machine. Removing these repetitions, we obtain a total of 95 detections.

In terms of overall virulence, this indicates that over a period of 4 months, 38% of our population was exposed to malware. In truth, this figure far exceeded the expectations of the members of the research team, largely based on their own experience (e.g. number of AV warnings over an equivalent period of time). More importantly, however, these results would indicate that, if they are representative of the whole user population, almost 1 out of 2 newly installed machine would be infected within 4 months if they had not had an AV installed! This figure might seem at first alarming and surprising, but in fact the Eurostat report mentioned earlier [6] indicates that over a period of 12-months in 2010, 31% of users reported a virus infection on their home computers, while 84% of these users reported having some kind of security software installed (AV, anti-spam, firewall, etc.). Thus, if these figures are to be trusted a theoretical 38% exposure rate should not be surprising.

In terms of the evolution of the number of infections over time, we can see that the level of detections is very similar for each month, contradicting the hypothesis that users are most at-risk when they first start using their machines. This is shown in Figure 1 where the distribution of the detections without repetition for each month is depicted.

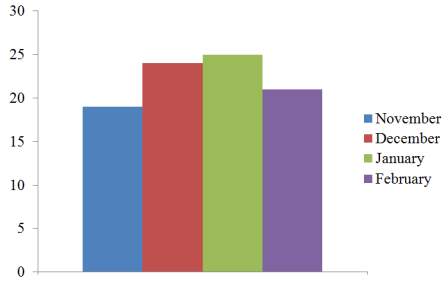

Figure 1: Unique malware detections by month

Finally, in terms of type of malware each of these detections was classified based on the information provided by the AV product. Figure 2 shows the distribution of malware detections by type. As we can see, almost all detections were classified as trojans, while viruses and adware have a relatively weak representation.

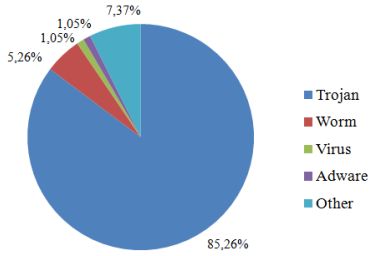

Figure 2: Malware detections by type

These figures are somewhat similar to those reported for overall infections by other AV vendors. For example, the first 2012 quarterly report from Panda Security [18], indicates that trojans account for most detections with a ratio of 63.30%, while worm, virus, adware and other have respectively ratios of 8.39%, 7.90%, 7.81% and 9.60%. Nonetheless, the differences with our results could be partially attributed to differences in the classification methods. For example, a file can be classified as a trojan by the AV product being evaluated and as a virus by another product. Furthermore, statistical error could be significant since our results are only based on a collection of 95 malware samples, while those of Panda Security are probably based on thousands of different samples.

While a detailed analysis of the causes and means by which these threats ended up on the computers still remains to be done in future work, we already know that 17 of these malware propagated through portable storage devices.

4.2 Missed Detections

Our experimental protocol [14, 13] describes in detail the monthly procedure for identifying and classifying suspicious files that were not detected by the AV. This process of identification and classification is based on user reporting of suspicious machine behaviour, the analysis of logs from the monitoring tools, the results of automated queries to on-line sources with respect to processes found on the machine, file and start-up programme databases (obtained automatically by scripts that we wrote), and any other relevant piece of information that the technician conducting the review might deem relevant.

Suspicious files found on the computer were classified into four categories: dangerous, suspicious, safe and unrated. All files marked as dangerous, suspicious and unrated were subjected to a more in-depth analysis. When we suspected that a file might be dangerous, additional data were collected with the consent of the user, including the actual browsing history, the suspicious file, and other related files present on the computer.

Our analysis identified 20 possible infections on 12 different machines. The most useful detection tool was HijackThis, which was involved in identifying 18 of the suspected infections. SpyBHORemover helped us find one additional infection. The last suspected infection was reported by the user, who called the project manager using the provided contact number when he suspected that his machine had been infected. All suspicious files were captured during the monthly visits, except for the user-reported suspected infection. While the logs show the location and filename, the file could not be retrieved as it seems that the suspected malware uninstalled itself between the time the user called in and the following lab visit. All captured files (19 out of 20) were later scanned with the evaluated AV product to see if they would be detected *a posteriori*. Even several months after the end of the experiment, none were detected by the AV product or identified as a potential threat. We scanned the captured files *a posteriori* with the VirusTotal service to compare the results obtained by several AV products and to compare these later results with those obtained a few months earlier. Additionally, we searched the Internet to find as much detail as we could for each of these 20 detections. As a result of this analysis, we classified two of the samples as clean, seven as unwanted software, nine as adware, one as definite malware, and one as suspected (but unconfirmed) malware.

The detected adware samples were either BHO or toolbars. In all cases, they were unknowingly installed by the users. Their effects included changing the web browser home page, redirecting web searches, or displaying advertisements. Further analysis will be required to determine if these adware are indeed malicious, in that they show additional behaviour that might have further consequences for the user than those described (e.g. theft of personal/private information). While we have not yet analysed in detail the two suspected malware samples, we have confirmed that one of them is rogueware. As previously mentioned, the corresponding user contacted us to inform us that his laptop was probably infected. It turned out that the laptop was infected with the fake "AV Security Scanner". Windows were regularly appearing to inform the user that harmful software was on his computer and every application started was killed except for web browsers. In order to get rid of these infections, the user was invited to register and provide his contact and payment information. At that moment, the user suspected that he may be infected and contacted us. As explained before, since the files disappeared from the computer before it was brought in for inspection, it was not possible for us to verify if the AV product detected this threat *a posteriori*.

Overall, 18 threats have been detected on 10 machines, which represents 20% of the users. One first point of comparison is the above-mentioned Eurostat report. Unfortunately, that report does not provide separate infection statistics for the population with and without AV installed. However, if we assume a theoretical comparison population with the same ratio of with/without AV (84% and 16%, respectively) and the same infection rates as we observed or inferred (20% and 38%), this would give a combined infection rate of 23%, in comparison with the self-reported 31% combined infection ratio of the Eurostat report. It has been said that the Eurostat report might have been an underestimation due to users only being able to notice a fraction of the actual infections, but it can be equally argued that they might be exaggerated due to users being paranoid and attributing performance problems to "viruses". Closer examination of the Eurostat results indicate a strong variance of reported infection percentage by EU countries, e.g. low twenties for Germany, Netherlands, Finland, and 40% and higher for many Eastern European countries. Thus, it would appear the lower infection ratio observed for our Canadian users might be related to geographical factors (whether location or cultural).

Another point of comparison is the SurfRight report [25]. Over a period of 55 days, 107,435 users used the Scan Cloud product, 73% of which were found to have an up-to-date AV product installed. Of those, Scan Cloud found that 32% were infected, while 46% of unprotected machines were infected. In comparison with our 20% and 38% ratio, it would appear that our sample population was less at risk than those using SurfRight's Scan Cloud. One possible explanation is simply that one of the motivations for using such a product is that the user already suspects that his machine is infected, probably a good indicator that it already is.

In all cases, it is important to point that straight comparison of these numbers is not significant given the fact that the definition and classification methods for threats in these studies are quite different. In our case we depend on a classification given by the Trend Micro AV product and our own investigations, similarly as for the SurfRight date, while the Eurostat results totally depend on user's self-assessment.

4.3 User Profiling and Behaviour

We examined whether an increase in certain types of user behaviour leads to a higher probability of the users' system being infected with malware. We also investigated whether user demographic factors and characteristics had any bearing on incidences of infection.

4.3.1 Characteristics and demographic factors

We examined whether gender, age, employment/student status, and work/study domain, and computer expertise had

any relationship with likelihood of getting infected. To test the impact of these characteristic and demographic factors, we divided the users in two groups. The first group contains *at-risk* users, which are those who received at least one detection, and the second group contains *low-risk* users who received no detections during the experiment. Table 1 shows the users distribution between the total population and the at-risk group based on user characteristics and demographic factors.

Table 1: Proportion of users for each factor

Factor		Total population	At-risk population
Gender	Male	60%	61%
	Female	40%	39%
Age	18-24	38%	35%
	25-35	38%	48%
	36+	24%	17%
Status	Student	64%	70%
	Worker	30%	26%
	Unemployed	6%	4%
Field	Computer Science	26%	22%
	Natural Sciences	52%	48%
	Arts/Humanities	22%	30%
Computer Expertise	High	18%	31%
	Low	82%	69%

The results in Table 1 suggest that age and computer expertise could potentially be risk factors for increased incidence of infection. We therefore conducted statistical analysis to determine whether this was the case.

Prior to our analysis, we performed a residual analysis in order to find the presence of outliers among the data with respect to the number of detections. The normal probability plot of raw residuals, as shown in Figure 3, allowed us to identify the presence of one outlier (who had 28 unique detections), reducing the sample to 49 users.

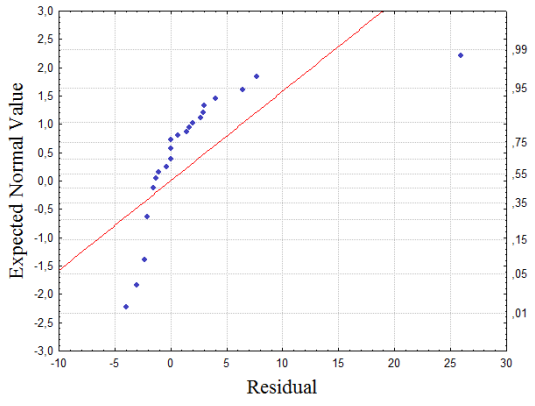

Figure 3: Normal probability plot of raw residuals

We conducted a logistic regression analysis to assess if particular characteristic and demographic factors may increase the infection risk. The dependent variable was the level of risk (at-risk vs. low-risk groups). It was indicated by either 1 or 0, depending on whether the user was exposed to any threats during the experiment. Results of the logistic regression analysis are summarised in Table 2. Items in bold are statistically significant.

We also used one-way ANOVA tests to assess whether the number of unique detections was affected by the characteristics and demographic factors. Results of the ANOVAs are presented in Table 3. In all cases, we consider $p < 0.05$ as statistically significant.

Table 2: Logistic regression analysis comparing risk of exposure

Factor	χ^2	p-value
Gender	0.098	0.75
Age	2.220	0.34
Status	0.465	0.80
Field	0.307	0.86
Computer expertise	**0.299**	**0.042**

Table 3: One-Way ANOVA comparing the number of unique detections

Factor	SS	MS	F	p
Gender	10.36	10.36	1.69	0.20
Age	19.25	9.62	1.58	0.22
Status	5.10	2.55	0.40	0.67
Field	21.56	10.78	1.79	0.18
Computer Expertise	4.41	4.41	0.70	0.41

Gender.

The total population included 30 males and 20 females which gives a proportion of 60% and 40% respectively. Table 1 shows that the gender distribution among the 23 at-risk users is 61% for the males and 39% for the females. The logistic regression analysis revealed no statistically significant differences between males and females (Table 2) with respect to exposure to threats.

We next examined the number of detections to determine whether there were any differences between the two groups. Figure 4 suggests that there is a minor difference in the average number of unique detections between males and females. However, statistical analysis reveals no statistically significant difference between males and females for number of unique detections (Table 3). While previous studies [15, 16] have addressed gender difference in computer security, to the best of our knowledge, no empirical experiment has evaluated gender differences in risk of infection. Our statistical tests suggest that gender is not a risk factor.

Age.

We divided our users into three age groups as evenly as possible (although we note that the older age group has fewer users due to our sample). Table 1 shows that the proportion of 18 to 24-year-olds who are at-risk is almost the same as for the total population. For those 25 to 35, the proportion in the at-risk group (48%) is higher than for the total population (38%). And for the group 36+ age group, we observe a slight decrease of 7% in the proportion between the total population and the at-risk group. Our logistic regression re-

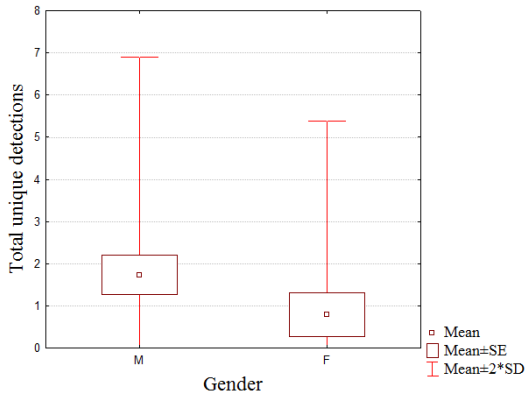

Figure 4: Box plot of unique detections by gender

sults (Table 2) show no statistically significant difference for the impact of the age on exposure to threats.

We further investigated whether the number of detections was impacted by age of the participant. Figure 5 mildly suggests differences between age groups, however these results were not statistically significant (Table 5).

Unlike the previous studies mentioned in Section 2 [16, 21, 17] that concluded that younger users are more at-risk, our statistical results suggest that age is not a significant risk factor. This discrepancy can be explained first because the experimental methods are quite different: all of them involve surveys of users where susceptibility levels are evaluated either through a theoretical model or from user self-declarations of previous incidents, and not from actual observation. Second, these results are not (all) specific to malware infections. Finally, the granularity of the age data recorded is different (coarse in our case) so it is hard to compare precisely these discrepancies, especially since the age distributions are quite different.

In any case, what is clear is that none of these studies, including ours, can be used to make categorical statements about risk and age, as all of these studies were based on biased samples of the population, i.e. students and other persons recruited on campus. Large scale studies with more uniform population sampling will be required to settle the issue of age as a risk factor for malware infections.

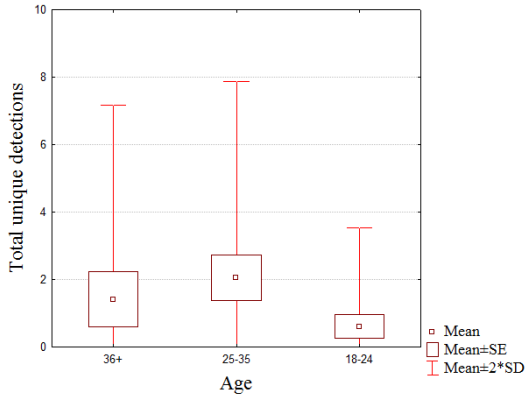

Figure 5: Box plot of unique detections by age

Status.

Participants were classified in three categories: students, workers, or unemployed. Table 1 indicates that the proportion between the total population and the at-risk group is quite similar for each category. Logistic regression confirms that no statistically significant differences exist.

Similarly, Figure 6 illustrates the relationship between these categories and the number of detections. Results of the ANOVA show no statistically significant differences (Table 3), further supporting that student/employment status is not a risk factor.

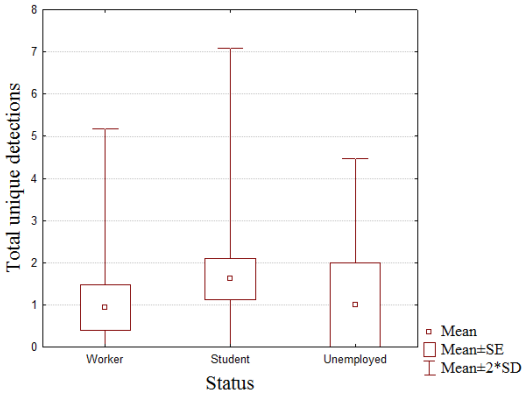

Figure 6: Box plot of unique detections by student/employment status

Work/study domain.

We recruited participants based on their domain of work or study in order to have a representative sample. As shown in Table 1, 26.5% of participants self-identified as being in computer science/information technology fields, 47% in natural sciences and 26.5% in arts and humanities. Although the table suggests that those in the Arts/Humanities might be slightly more at-risk, results of the Logistic Regression (Table 2) show no statistically significant differences between the three groups.

Figure 7 presents the average number of unique detections per domain. Interestingly, it appears that participants in the Arts/Humanities have fewer detections. However, these differences were not statistically significant (Table 3). Our statistical tests provide no support indicating that domain of work or study is a risk factor.

Computer expertise.

We assessed computer expertise by asking users about their proficiency with certain technical tasks. Users were considered to have a high level of computer expertise if they had previously completed all of the following tasks: configured a home network, created a web page, and installed or re-installed an operating system on a computer. Overall, 18% of users were classified as computer experts for the purposes of our analysis. As observed in Table 1, those with high expertise were nearly twice as likely to be in the high-risk group when compared to the total population. This may indicate that a high level of expertise increases the risk of exposure to threats. This was confirmed by the logistic

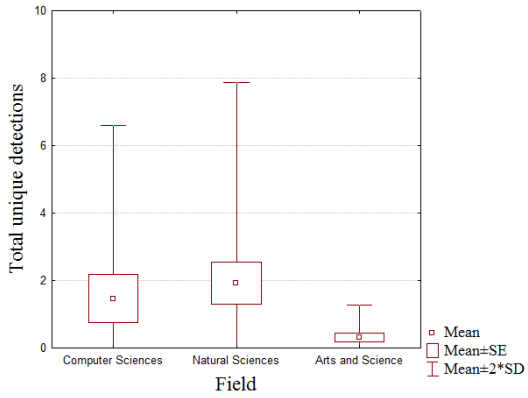

Figure 7: Box plot of unique detections by domain expertise

regression which shows that those with high computer expertise were statistically more likely to be exposed to threats.

As illustrated in Figure 8, we further compared the number of detections between the high and low computer expertise groups. Higher variance is apparent for the high-expertise group, but the differences are not statistically significant (Table 3). Our results suggest that while computer experts are more likely to have at least one exposure to threats, overall they do not see a higher number of detections.

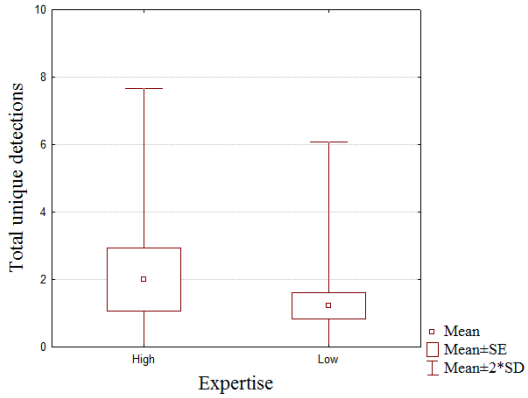

Figure 8: Box plot of unique detections by computer expertise

Summary of user characteristics.

In summary, we found little evidence linking user characteristics to increased risk of threat exposure. Gender, age, student/employment status, and domain of expertise showed no statistically significant differences. However, we did find partial support linking high computer expertise to increased risk of exposure.

4.3.2 *Behavioural factors*

To assess if specific user behaviours led to a higher risk of infection, we focused our analysis on 4 variables: most used browser, total number of applications installed, total number of websites visited, and categories of websites visited.

Using a similar approach to that described in section 4.3.1, we first conducted a logistic regression analysis where users were divided in two groups: at-risk and low-risk users. Secondly, we performed a general regression analysis to assess if there was a relationship between the behavioural factors analysed and the number of unique detections. General regression analysis was chosen instead of ANOVA because some of the independent variables represent continuous values (e.g. the number of gambling websites visited) rather than categorical data. Table 4 summarises the statistical results. Items in bold are statistically significant.

Browser.

Each month, users were asked which web browser they used most. We compiled these results in order to obtain for each user the most frequently used browser for the entire duration of the study. Table 5 summarises the proportion of users who utilised each web browser for the total population and the at-risk population. Even though a small increase is observed between the two populations for Chrome, our logistic regression analysis revealed no statistically significant differences between the two groups with respect to the type of browser used.

Table 5: Most frequently used browser

Browser	Total population	At-risk population
Internet Explorer	30%	17.4%
Firefox	30%	26.1%
Chrome	40%	56.5%

We also studied whether the use of a specific browser may influence the number of unique detections per user. To this end, the results obtained from our general regression analysis in Table 4 confirm our previous findings. The usage of a specific web browser is not related to the risk of infection.

Applications installed.

We monitored the number of applications installed by each user. To assess the potential effect of number of applications on the risk of infection, we performed a logistic regression, as shown in Table 4, which suggests that installing many applications may increase the exposure to malware and thus the probability of being infected.

This finding is also supported by our general regression analysis which shows that number of applications installed significantly impacts the number of detections (Table 4). The more a user installs applications, the more he is likely to get infections, as illustrated in Figure 9. Our analysis supports the idea that number of applications is a risk factor for infection.

Websites visited.

The number of websites visited was also recorded for the entire duration of the study to evaluate the impact on the risk of infection. Figure 10 shows that at-risk users visited many more websites than low-risk users, suggesting that visiting many websites could potentially increase exposure to threats. Logistic regression analysis confirms this finding (Table 4), showing that the number of websites significantly impacts the risk of exposure.

Table 4: Statistical analysis results of user behaviour factors

	Logistic Regression		General Regression	
Factor	χ^2	p-value	t-value	p-value
Type of browser	2.563	0.278	0.75	0.50
Number of applications installed	**4.709**	**0.030**	**2.29**	**0.03**
Total number of websites visited	**6.247**	**0.012**	1.71	0.09
Number of streaming media/MP3 sites visited	**11.999**	**0.001**	**3.372**	**0.001**
Number of peer-to-peer sites visited	**6.864**	**0.009**	1.942	0.064
Number of Internet Infrastructure sites visited	**7.469**	**0.006**	**5.466**	**0.000**
Number of Software download sites visited	**14.326**	**0.000**	**4.012**	**0.000**
Number of Sports sites visited	**4.194**	**0.041**	**2.601**	**0.013**
Number of Social Networking sites visited	**6.260**	**0.012**	1.965	0.056
Number of Computers/Internet sites visited	**7.357**	**0.007**	**2.292**	**0.026**
Number of Gambling sites visited	**4.998**	**0.025**	**3.601**	**0.001**
Number of Pornography sites visited	2.930	0.087	**6.425**	**0.000**
Number of Illegal/Questionable sites visited	0.022	0.881	**2.697**	**0.025**
Number of Translator/Cached sites visited	1.689	0.194	**5.799**	**0.000**

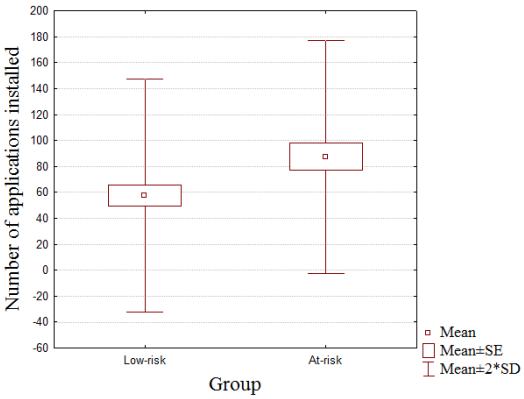

Figure 9: Box plot of applications installed by group

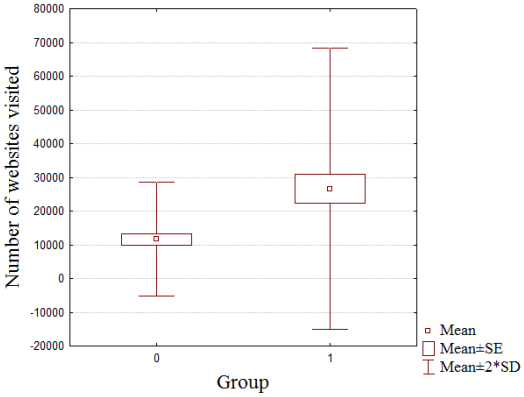

Figure 10: Box plot of websites visited by group

However, our general regression analysis did not confirm that the number of websites visited is a significant risk factor. The associated p-value (0.09) is under 0.1, which could suggest that the effect may be a potential risk factor that would need to be validated in a larger-scale study.

Type of websites visited.

We further wanted to analyse if particular types of websites were more prone to causing malware exposure. To this end, we classified each website visited using the Site Safety Center of Trend Micro [26]. We then performed a logistic regression using the number of websites visited for each category to determine if some categories were riskier. As shown in Table 4, our results allowed us to identify eight risky categories: streaming media/MP3, peer-to-peer, Internet infrastructure, software downloads, sports, social networking, and computers/Internet and gambling.

The general regression analysis showed quite similar results (Table 4) with nine categories showing that more frequent visits led to higher number of detections. Statistically significant results were found for the following categories: streaming media/MP3, infrastructure sites, software download, sports, computers/Internet, gambling, pornography, illegal/questionable and translator/cached. Our results indicate that type of website visited is a risk factor for infection.

Summary of user behaviour.

We have been able to identify three different risk factors related to user behaviour. We found that users who install many applications are more prone to install infected applications, thus increasing his risk of being infected. Also, visiting many websites could be considered as a risk factor as some pages have malicious code that is automatically executed. Finally, we have also confirmed that certain categories of websites place users at a greater risk of getting infected.

5. DISCUSSION

The results we have obtained and discussed here are subject to certain limitations. For one, the AV performance evaluation is limited to only 113 detected threats, a very small number compared to the numerous threats in the wild, especially considering that some of these may be false positives. In addition, the false negative number might also be underestimated due to the fact that we cannot guarantee that our protocol caught all malware missed by the AV. In other words, we do not have absolute ground truth.

One obvious limitations of the study concerns potential biases in selection of the population. First of all, subjects were located in the Greater Montréal area. Second, their

age distribution is different from that of Canadian Internet users. However, unlike some of the studies discussed, ours is not restricted to undergraduate students in the 18-24 group; in our study, they represent only 38% of our subjects.

Another potential source of bias might have been the fact the users knew that they were part of a computer security experiment. This knowledge might have caused them to alter their usage of their laptop. We asked that question in the exit survey and 43 users claimed that they did not modify their behaviour. Of the other 7 users, 2 admitted having modified their behaviour to fulfil experiment constraints (no OS reinstallation, creation of partitions, etc.), 2 others admitted voluntarily not performing potentially embarrassing activities on the computer, 1 mentioned refraining from visiting secure Internet banking sites, 1 admitted forcing himself to use the computer more frequently, and the last one explained that he controlled access to his computer in order to ensure being its only user. All in all, and considering the usage statistics we gathered show normal to high levels of computer and web activity, and given the fact that the laptops were sold to and were to be kept by the subjects, we are confident this potential experimental bias did not significantly affect our results.

Finally, even though we were able to determine several factors correlated with risk of infection, these factors in themselves are not sufficient to explain the causal link leading to infection. To this effect, a more detailed analysis of the collected data is required in order to determine the sources and means of infection for each of the 113 detected threats. Only then will we be able to determine which of these of these factors are causes of infection, and which are consequences of other causal common factors that were not considered in this study, e.g. risk averseness or risk propensity of users, etc.

In terms of applications, we believe the single product field study methodology described in this paper is of potential utility in at least two contexts. One is that it should be suitable for AV vendors seeking to understand how their products perform in real-world usage and might help identify which aspects of the product (user interface, detection, remediation, etc.) could be further improved. The other is to help understand what characteristics of user behaviour lead to higher risks of infection. These characteristics could be used to improve the content and targeting of user education and training; they could also be used by insurance companies to assess relative risk in IT insurance policies.

Today most AV tests are lab-based and designed to identify which AV products perform better than others, whether the purpose is to allow users to make a more educated choice of product or to help AV vendors determine R&D and marketing strategies. Given the advantages in term or ecological validity and availability of user data of field tests such as the one described here, a natural question is whether it is feasible to conduct a comparative field study of AV products to complement the results of lab-based comparative testing.

One of the major issues in lab-based comparative AV testing is ensuring that all AV products are evaluated under exactly the same conditions. Not only should they be tested in the same environment, but they should also be exposed to the same threats at the same time. While it is relatively easy to guarantee consistent conditions when tests are performed within a controlled environment, such consistency cannot be guaranteed in field studies. The exposition of the product to threats cannot be controlled as it is user driven. On the other hand, field studies are inherently unbiased in that threats applied to each AV are independently "chosen" by users who have no vested interest in the test results. Furthermore, the law of large numbers guarantees that for sufficiently large populations, each product will be exposed to a large enough sampling of threats to make the results statistically significant. In order words, and to eliminate bias in user-driven threat selection, comparative field studies should be conducted with a large enough population and over a large enough period of time to guarantee a statistically significant sample of user and malware behaviour. Based on our experiences documented here, we postulate that such a study should include at least 200 participants per AV tested and last at least four months to accomplish that aim.

Thus, if we are to compare multiple AV products, we can easily be looking at a study with a thousand or more subjects. Increasing to this scale will likely require increased use of automation to gather user feedback. Such automation, however, is potentially beneficial as well because it allows for feedback to be obtained in context, i.e. just after a user has interacted with the AV software. Care must be taken, though, to make sure the automated questions do not themselves overly influence the results.

It would also be nice to gather more detailed information about user behaviour and computer state. Any such efforts will require very careful ethical review. We note that medical clinical trials involve people revealing intimate life details and subjecting themselves sometimes to interventions that can potentially kill them. As such, it should be possible to create protocols that provide sufficient care for end users and their computers while providing greater insight into malware attacks and defences as they happen in the field.

6. CONCLUSION

We have presented the results from the first field study of anti-virus software performed with real users in non-laboratory conditions. While the population studied was small compared to medical clinical trials, it is comparable to that of other usability studies and was sufficient to obtain some interesting results with respect to infection risk factors.

In terms of AV performance, our results indicate that 38% of users got exposed to a threat caught by the AV, indicating that at least 38% of the population would have got infected had they had no AV installed. In addition, 20% of our users were found to have been infected by some form of undesirable software that was not detected by the AV. These figures are indeed alarming, but they are comparable with figures reported by other studies conducted by other methods (user self-reporting and on-line scanning of machines).

In terms of risk factors, our results show that user behaviour is indeed significant, thus confirming previous work in the area. However, our results show that user characteristics such as age or gender are not significant risk factors, contradicting related research. We observed some surprising patterns in web browsing risk, with seemingly innocuous categories of sites such as sports and Internet infrastructure being associated with a higher rate of infection while more "shady" sites such those containing pornography and illegal/questionable content were less so. And somewhat non-intuitively, we found that computer expertise is a weak factor *increasing* the risk of infection.

Beyond the contribution of these results, this work demonstrates that field studies are a viable alternative to lab-based

AV evaluation. The methodological issues are comparable to that of other computer science user studies, particularly those in usable security. While studies comparing multiple AV or other security products will require larger populations to get statistically significant results, increasing use of automation should allow such studies to be performed at relatively modest cost. We thus believe the work presented here illustrates the merits of future larger scale clinical trials of AV and other security software.

7. ACKNOWLEDGMENTS

This project was funded by Trend Micro and Canada's Natural Sciences and Engineering Research Council (NSERC), through the Inter-networked Systems Security Network (ISSNet) Strategic Research Network, the Discovery Grant programme, and a Canada Research Chair (fourth author).

8. REFERENCES

[1] F. Asgharpour, D. Liu, and L. J. Camp. Mental models of computer security risks. In *Workshop on the Economics of Information Security (WEIS)*, 2007.

[2] AV Comparatives. File detection test of malicious software. Technical report, AV Comparatives, 2013.

[3] D. Botta, R. Werlinger, A. Gagne, K. Beznosov, L. Iverson, S. Fels, and B. Fisher. Towards understanding IT security professionals and their tools. In *ACM Symposium on Usable Privacy and Security (SOUPS)*, pages 100–111, 2007.

[4] J. Canto, M. Dacier, E. Kirda, and C. Leita. Large scale malware collection: lessons learned. In *IEEE SRDS Workshop on Sharing Field Data and Experiment Measurements on Resilience of Distributed Computing Systems*, 2008.

[5] A. De Luca, M. Langheinrich, and H. Hussmann. Towards understanding ATM security: a field study of real world ATM use. In *ACM Symposium on Usable Privacy and Security (SOUPS)*, 2010.

[6] Eurostat. Nearly one third of internet users in the EU27 caught a computer virus. `http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/4-07022011-AP/EN/4-07022011-AP-EN.PDF`, February 2011.

[7] I. Gashi, V. Stankovic, C. Leita, and O. Thonnard. An experimental study of diversity with off-the-shelf antivirus engines. In *IEEE International Symposium on Network Computing and Applications (NCA)*, 2009.

[8] S. Gordon and R. Ford. Real world anti-virus product reviews and evaluations: the current state of affairs. In *National Information Systems Security Conference*, 1996.

[9] D. Harley and A. Lee. Who will test the testers. In *18th Virus Bulletin International Conference*, pages 199–207, 2008.

[10] J. Kephart and S. White. Directed-graph epidemiological models of computer viruses. In *IEEE Symposium on Security and Privacy*, 1991.

[11] S. Kondakci. Epidemic state analysis of computers under malware attacks. *Modelling Practice and Theory*, 16:571–584, 2008.

[12] P. Kosinar, J. Malcho, R. Marko, , and D. Harley. AV testing exposed. In *20th Virus Bulletin International Conference*, 2010.

[13] F. Lalonde-Levesque. Évaluation d'un produit de sécurité par essai clinique. Master's thesis, École Polytechnique de Montréal, August 2013.

[14] F. Lalonde-Levesque, C. Davis, J. Fernandez, S. Chiasson, and A. Somayaji. Methodology for a field study of anti-malware software. In *Workshop on Usable Security (USEC)*, pages 80–85. LNCS, 2012.

[15] F. Lalonde-Levesque, C. Davis, J. Fernandez, and A. Somayaji. Evaluating antivirus products with field studies. In *22th Virus Bulletin International Conference*, pages 87–94, September 2012.

[16] G. R. Milne, L. I. Labrecque, and C. Cromer. Toward an understanding of the online consumer's risky behavior and protection practices. *Journal of Consumer Affairs*, 43:449–473, 2009.

[17] F. T. Ngo and R. Paternoster. Cybercrime victimization: An examination of individual and situational level factors. *International Journal of Cyber Criminology*, 5(1):773–793, 2011.

[18] Panda Security Labs. Panda Labs quarterly report January - March 2012. `http://press.pandasecurity.com/wp-content/uploads/2012/05/Quarterly-Report-PandaLabs-January-March-2012.pdf`, 2012.

[19] PC Security Labs. Security solution review on Windows 8 platform. Technical report, PC Security Labs, 2013.

[20] J. A. Rode. Digital parenting: designing children's safety. In *British Human Computer Interaction Conference (British HCI)*, pages 244–251, 2009.

[21] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs. Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In *ACM Conference on Human Factors in Computing Systems (CHI)*, pages 373–382, 2010.

[22] R. Shyamasundar, H. Shah, and N. Kumar. Malware: from modelling to practical detection. *Distributed Computing and Internet Technology*, pages 21–39, 2010.

[23] K. Solic and V. Ilakovac. Security perception of a portable PC user (the difference between medical doctors and engineers): A pilot study. In *Medicinski Glasnik*, volume 6, pages 261–264, 2009.

[24] A. Somayaji, Y. Li, H. Inoue, J. Fernandez, and R. Ford. Evaluating security products with clinical trials. In *USENIX Workshop on Cyber Security Experimentation and Test (CSET)*, 2009.

[25] SurfRight. 32% of computers still infected, despite presence of antivirus program. `http://www.surfright.nl/en/home/press/32-percent-infected-despite-antivirus`, 2009.

[26] Trend Micro. Website classification. `http://solutionfile.trendmicro.com/solutionfile/Consumer/new-web-classification.html`, 2012.

[27] J. Vrabec and D. Harley. Real performance? In *EICAR Annual Conference*, 2010.

[28] R. Wash. Folk models of home computer security. In *ACM Symposium on Usable Privacy and Security (SOUPS)*, page 11, 2010.