Technological and Human Factors of Malware Attacks: A Computer Security Clinical Trial Approach

FANNY LALONDE LÉVESQUE, École Polytechnique de Montréal
SONIA CHIASSON and ANIL SOMAYAJI, Carleton University
JOSÉ M. FERNANDEZ, École Polytechnique de Montréal

The success (or failure) of malware attacks depends upon both technological and human factors. The most security-conscious users are susceptible to unknown vulnerabilities, and even the best security mechanisms can be circumvented as a result of user actions. Although there has been significant research on the technical aspects of malware attacks and defence, there has been much less research on how users interact with both malware and current malware defences.

This article describes a field study designed to examine the interactions between users, antivirus (AV) software, and malware as they occur on deployed systems. In a fashion similar to medical studies that evaluate the efficacy of a particular treatment, our experiment aimed to assess the performance of AV software and the human risk factors of malware attacks. The 4-month study involved 50 home users who agreed to use laptops that were instrumented to monitor for possible malware attacks and gather data on user behaviour. This study provided some very interesting, non-intuitive insights into the efficacy of AV software and human risk factors. AV performance was found to be lower under real-life conditions compared to tests conducted in controlled conditions. Moreover, computer expertise, volume of network usage, and peer-to-peer activity were found to be significant correlates of malware attacks. We assert that this work shows the viability and the merits of evaluating security products, techniques, and strategies to protect systems through long-term field studies with greater ecological validity than can be achieved through other means.

CCS Concepts: • **Security and privacy** → Usability in security and privacy; • **Human-centered computing** → **HCI design and evaluation methods**; *Field studies*; *User studies*;

Additional Key Words and Phrases: Computer security, clinical trial, malware, antivirus, risk factors

ACM Reference format:
Fanny Lalonde Lévesque, Sonia Chiasson, Anil Somayaji, and José M. Fernandez. 2018. Technological and Human Factors of Malware Attacks: A Computer Security Clinical Trial Approach. *ACM Trans. Priv. Secur.* 21, 4, Article 18 (July 2018), 30 pages.
https://doi.org/10.1145/3210311

This project was funded by Trend Micro, Fonds de recherche du Québec - Nature et technologie (FRQNT), through a doctoral research scholarship (first author), and Canada's Natural Sciences and Engineering Research Council (NSERC), through the Inter-networked Systems Security Network (ISSNet) Strategic Research Network, the Discovery Grant program, the Alexander Graham Bell Canada Graduate Scholarship Doctoral program (first author), and a Canada Research Chair (second author).

Authors' addresses: F. L. Lévesque and J. M. Fernandez, École Polytechnique de Montréal, 2900 Edouard Montpetit Blvd, Montreal, QC, Canada, H3T 1J4; emails: {fanny.lalonde-levesque, jose.fernandez}@polymtl.ca; S. Chiasson and A. Somayaji, Carleton University, 1125 Colonel By Dr, Ottawa, ON, Canada, K1S 5B6; emails: {chiasson, soma}@scs.carleton.ca.

https://doi.org/10.1145/3210311

1 INTRODUCTION

Malicious activity on the Internet is continuously evolving; the nature of threats changes rapidly. Modern malware authors adapt their techniques to exploit new vulnerabilities, take advantage of new technologies, and evade security products. Users may be enticed to take direct (or indirect) actions that lead to the infection of their computers. Some actions, such as opening an email attachment or visiting a malicious web site may occur immediately prior to infection. Others, such as not updating system or willingly installing software whose true intention is masked, may occur over time so that a combination of actions lead to a vulnerable system state.

Meanwhile, antivirus (AV) products have evolved in response. 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 and system behaviour monitoring [42]. 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; these databases are in turn fed by the continuous reporting of millions of AV client installations [1, 42]. AV products have thus evolved into complex "anti-malware" software, or rather complex software systems involving several semi-independent components with which the user must occasionally interact. While many AV vendors try to make the installation and operation of their product as usable as possible, the truth is that the AV's operation and performance still depends on the user. This dependence is due to user configuration of the many AV features and to other user-driven factors such as how often the machine is connected to the Internet, how often its software and signatures are updated and, most importantly, how users 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 AV products and malware not only includes the machine they are trying to protect/penetrate but also the network that connects it to the rest of the world and its *user*. Indeed, the human is part of the operating environment of the machine, along with the software that attempts to execute on it or protect it. It thus seems natural to adopt a *human-in-the-loop* approach to evaluate the performance of AV products and the susceptibility of users to getting their machines infected. This change of paradigm is fundamental if we want to better understand what role users really play in the process of malware infection. In particular, it becomes paramount to understand how human factors, such as demographics, computer literacy, perception of threat, and user behaviour may affect the risk of malware infection.

This philosophy of *human-in-the-loop* is also in sharp contrast with current AV evaluation methods, which are largely based on automated tests performed in controlled environments [8, 13, 15, 32]. While these tests are adequate to evaluate AV products under specific scenarios, they do not measure the "real-world" efficacy of AV products as deployed on machines operated by real users [24]. Even the most advanced tests, which include automated user profiles [40], cannot accurately capture all user behaviour and other external factors, such as evolving malware threats or different system configurations, that may affect how AV perform.

To address these shortcomings and to understand the influence of human factors in malware attacks, an alternative method, *computer security clinical trials*, was proposed in 2009 by Somayaji et al. [44]. We conducted the first such experiment at the École Polytechnique de Montréal in 2011–2012, involving 50 home users using their own computers in everyday life for 4 months. This journal version of our work details the methodology [22, 29] and extends the preliminary results already presented in earlier work both in terms of AV product evaluation [30] and human risk factors [26]. In particular, this article (i) provides a new evaluation of the AV product including the efficacy (Section 5.3) and the user experience (Section 5.4), (ii) reinterprets and updates statistical analysis on human factors (Section 6), (iii) investigates some unstudied user behaviour factors (Section 6.2), such as system activity, applications usage, network usage, files downloaded, peer-to-peer activity, and level of vigilance, and (iv) extends the discussion to include the implications of these new results.

The remainder of the article is structured as follows. In Section 2, we present related work in AV product evaluation and human factors related to computer threats. Section 3 details the concept of computer security clinical trials. Section 4 describes its methodology. In Section 5, we discuss the results of the study in terms of threat detections by AV, missed detections, and user experience. Section 6 identifies potential risk factors related to user characteristics, demographics, and behaviour. We discuss limitations of our work in Section 7. Finally, we conclude and summarize the results and implications of this work in Section 8.

2 RELATED WORK

Numerous studies evaluating the performance of AV products and the influence of human factors on information technology (IT) security have been conducted in recent years. We describe in Section 2.1 the current state of affairs regarding AV product evaluation, and we present in Section 2.2 previous research related to humans factors (user characteristics, demographics, and behaviour) and computer threats.

2.1 Antivirus Product Evaluation

Antivirus products offer important information system protection against current threats. Testing how these products are effective at protecting end-users and their systems is therefore crucial. We present here a critical review of current evaluation methods and discuss their limitations.

Controlled Conditions. Typical evaluations by commercial testing labs (see, e.g., References [2, 40]) are based on automated tests conducted in controlled environments. For example, file scanning tests (also called "static" or "on-demand") are based on scanning collected or synthesised malware samples along with legitimate programs. As there is no file execution, i.e., there is no software behaviour to analyze, static tests cannot adequately reflect the performance of products using active and proactive detection. Dynamic tests consist either of executing files or exposing antivirus products to known bad URLs [8]. While this latter type of tests does evaluate the performance of AV products as a whole (and not that of individual features), they may not be representative of what is typically experienced by users "in the wild."

One major issue with in-lab tests is that the sample malware collection is often too small, inappropriate, and not validated [16, 20]; this problem is often referred to as the *sample selection problem*. This can easily bias the test results, whether consciously or not, and can thus severely limit their usefulness. To this end, the WildList Organization International has proposed in 1993 the WildList [47], a cooperative listing of malware. This list, which only contains malware observed in the wild, has the main advantage of being validated by security professionals. However, it may not be representative of the most common malware in real-time, as it is only updated monthly. To partially address this shortcoming, the Anti-Malware Testing Standards Organization (AMTSO) created in 2013 the Real Time Threat List (RTTL) to provide a real-time view of threats as they are found in the wild [53]. The RTTL allows testers to conduct evaluations based on malware samples that represent the current state of the malware ecosystem. Although AV tests against such data sets are more realistic, they are not ecologically valid in that the effect of the human factor is not being measured.

Some researchers [34, 51] suggested emulating user interaction with scripts and creating user-specific testing scenarios. For example, testing for a gaming profile should prioritize network latency or reduction in frame rate, while testing for a worker profile should emphasize on downloading files from a server or audio/video file editing. As a first attempt, PC Security Labs conducted in 2013 an AV test [40] to measure the defence efficiency of AV solutions against seven different types of user profiles: Internet addict, network businessman, socializer, basic user, gamer, self-presenter and infrequent user. Their test confirmed that AV solutions perform differently depending on the user profile. However, though this testing approach simulates a more realistic operational conditions, it is impossible to capture all user behaviour, and external factors that may affect AV efficacy in real-life.

Real-life Conditions. One complementary approach to tests performed under controlled environments is to conduct evaluations in an holistic environment where the system, the AV product, and the user are included. For example, some observational field studies of AV products have been conducted. In such tests, AV products are not installed on systems. Rather, systems are monitored with their actual protection without any intervention. Blackbird et al. [3] used data from the Malicious Software Removal Tool (MSRT) on millions of systems to evaluate how AV protection state impacts infection rates. Lalonde Lévesque et al. [27] also used MSRT data to measure the overall performance of the AV ecosystem over a four-month period. In another study from Lalonde Lévesque et al. [24], the authors used data collected from the MSRT and Microsoft's Windows Defender on millions of systems to conduct a large-scale comparative test of AV products. Their findings showed that AV performance varies significantly as a function of external factors, such as user factors, environmental factors, and malware types.

Another potential way to assess AV performance in real-life is to conduct experimental field studies. For example, Somayaji et al. [44] proposed in 2009 conducting computer security clinical trials inspired by the same methodology as used in medical trials. In this method, security products are randomly deployed on specific populations and are monitored to assess their real-world performance in normal use. However, to the best of our knowledge, there has been no such studies of AV products published in the literature other than our previously published work [22, 30].

2.2 Human Factors and Computer Threats

In this section, we present a review of past work that studied how human factors, such as user demographics, characteristics and behaviour, correlate with computer threats.

Subjective Research Methods. One approach to studying the human factors in computer security is to adopt a subjective research method. This type of approach seeks to explore the perception and the attitude of users when they are facing computer security decisions. It primarily uses qualitative methods such as surveys, interviews and observations to understand how and why participants interact with computer systems.

For example, Milne et al. [33] applied protection motivation theory and social cognitive theory to understand online customers' risky behaviour and protection practices. They conducted a national online survey of 449 non-student respondents in 2009 and confirmed that age and gender are significant correlates of online risky behaviours; males and younger users were found to be more likely to adopt risky behaviours online. This assertion was also confirmed in 2010 by Sheng et al. [43] who conducted an online study with 1001 users to evaluate their susceptibility against phishing attacks. They concluded 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 and that young people aged between 18 and 25 are more susceptible to phishing. Using a sample of 295 college students, Ngo and Paternoster [35] applied the general theory of crime and lifestyle/routine activities framework to assess the effects of individual and situational factors on seven types of cybercrime victimization, including computer virus infection. They conducted a self-assessment survey in 2011 and deduced that age is a significant risk factor for computer virus infection, with older respondents being less likely to get infected. In another study, Bossler et al. [4] applied a routine activities framework to explore the causes and correlates of self-reported data loss from malware infection. The authors administered a survey on a sample of 788 college students in 2006 and investigated, among others, the effect of gender, age, race, and employment status. They found that being a female and being employed increases the odds of data loss compared to male and unemployed users, respectively. However, age was not identified as a significant predictor of self-reported data loss from malware infection. Similarly, Reyns et al. [41] also applied the routine activity theory to study online crime. Using data from a sample of 5,985 participants from 2008 to 2009, they investigated the relationship between individual's online routines, characteristics (age, gender, employment, income) and identity theft victimization. Results suggested that age, gender, employment, and income were significant correlates, where older respondents, males, employed respondents, and those with higher incomes were more at-risk. The authors also found that using the Internet for banking, shopping, communicating (e-mail, instant messaging), and downloading, is associated with increases in the likelihood of identify theft. Onarlioglu et al. [37] conducted a survey in 2011 on 164 Internet users who possess diverse backgrounds and varying degrees of computer security knowledge. Results confirmed the general intuition that technical security knowledge has a considerable positive impact on user ability to assess risk, especially when the threats involve technically complex attacks. Finally, Grimes et al. [14] surveyed 207 participants in 2007 to study how computer-related characteristics, online behaviours, and demographics (age, gender) correlate with spam attitudes and actions. The authors found no significant association between demographics and self-reported reception of spam. However, they did found some evidence linking specific online behaviours, such as purchasing online, making a web page, or posting in a newsgroup, and self-reported reception of spam.

Objective Research Methods. Another complementary approach to study human factors is to conduct studies based on an objective research method. While subjective studies will allow researchers to better understand user perception and perspective regarding computer threats, an objective method, either based on qualitative or quantitative data, will allow to study and measure user behaviour regarding computer security. For example, one approach to identify potential risk factors related to malware infection is to conduct observational or experimental studies based on real-life data, as self-reported data may lack ecological validity to represent actual user behaviour.

Maier et al. [31] performed in 2011 an empirical study based on network traces from residential users to analyse the relationship between security hygiene (AV and OS software updates) and potential risky behaviour. They found that computer hygiene has little correlation with observed behaviour, but that risky behaviour, such as accessing blacklisted URLs, can more than double the likelihood that a system will manifest security issues at the network level, e.g., sending spam, performing address scans or communications with botnet command-and-control (C&C) servers. Canali et al. [5] performed a comprehensive study on the effectiveness of risk prediction based on the web browsing behaviour of users in 2013. Their results showed that the more web sites a user visits, the higher is his exposure to threats. Ovelgonne et al. [38] leveraged 2009-2011 telemetry data from the Symantec's Worldwide Intelligence Network Environment (WINE) project [7] to study the relationship between user behaviour and cyber attacks. They created four user profiles (gamers, professionals, software developers, and others), and studied how seven machine features (number of binaries; fraction of unsigned, downloaded, low prevalence, and unique binaries;

number of ISPs to which the user connected) correlate with the number of attempted malware attacks by host machine. The authors found all features to be significant contributing factors, suggesting that heavy downloading of binaries, traveling a lot, and downloading rare pieces of code could increase the risk of malware attacks. In addition, they found software developers to be the most prone to malware attacks.

Some researchers have focused on phishing susceptibility. Jagatic et al. [18] launched in 2005 a real (harmless) phishing attack targeting 581 university students to quantify how reliable social context would increase the success of victimisation. Through their analysis, they found that females were more likely to fall victim of the social phishing attack. The attack was also slightly more successful with younger targets. Kumaraguru et al. [21] conducted in 2008 a real-world study to evaluate phishing training effectiveness, and investigate how users' demographic factors influence training and phishing susceptibility. Their results showed no significant difference between males and females. However, they found participants in the 18–25 age group to be consistently more vulnerable to phishing attacks than older participants. In another study, Oliveira et al. [36] investigated spear phishing susceptibility as a function of user age, gender, weapon of influence (scarcity, authority, commitment, etc.), and life domain (financial, health, social, etc.). The authors performed a 21-day study involving 158 participants, which took place in the participants' homes from 2015 to 2016. After exposing participants to experimentally controlled spear phishing emails, researchers found that women, particularly older women, were more susceptible to phishing attacks. Moreover, their results highlighted the extent to which younger and older participants differ in their susceptibility to various weapons of influence (scarcity, authority, commitment, etc.).

Other studies have adopted a methodological approach based on the concepts and methods of epidemiology. This approach refers to the likely causes and risk factors for infection, understanding the spread of malware and, where appropriate, the methods to remedy it. For example, Carlinet et al. [6] designed a case-control study in 2006 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 heavily using web applications and streaming are major risk factors of malware infection. Lee et al. [28] also conducted in 2010 a case-control study of academic malware recipients to identify putative factors that are associated with targeted attack recipients. The experiment revealed that specific individual profiles, such as individuals working in Eastern, Asiatic, African, American, and Australian Languages, Literature and Related Subjects, and Social Studies, especially Economics, are at a statistically significant elevated risk of being subjected to targeted attacks compared with others. Following the same methodology, Thonnard et al. [48] designed a case-control study to identify organizational and individual risk factors of targeted attacks. Based on a large corpus of targeted attacks blocked by an email scanning service from 2013 to 2014, they showed that directors and high-level executives are more likely to be targeted and that specific job roles such as personal assistants are even more at risk of targeted attack compared to others. Lalonde Lévesque et al. [23] conducted another case-control study specifically designed to evaluate the independent effect of age and gender on the risk of malware victimisation. Using data collected from Microsoft's Windows Defender on a sample of three million devices in 2015, the authors found that both age and gender are significant contributing factors for malware encounters. Men, and young men in particular, were found to be more susceptible to malware attack than women, and younger users to be more at risk than their older counterparts. Interestingly, results also suggested that the effect of age and gender is not constant across different types of malware; women were slightly more susceptible to encounter adware, and older users were more susceptible to rogue malware and ransomware. Also inspired by the epidemiology approach, Yen et al. [52] conducted in 2013 a study of malware encounters in a large, multi-national enterprise. They coupled malware encounters with Web activities and demographic information. Their results suggested that user demographic and behaviour features can be used to infer the likelihood of malware encounters; males and people with technical expertise were found to be more likely to encounter malware.

3 COMPUTER SECURITY CLINICAL TRIALS

One potential way to study technological and human factors of malware attacks is through conducting clinical trials of software, as proposed in 2009 by Somayaji et al. [44]. With such 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.

For this first experiment, we evaluated one single AV product, and we fixed some of the external factors that could affect a computer's likelihood of being infected by malware. For instance, all users were selected in the same geographic area. They all had the same laptop and system configuration, as those factors could affect the AV performance in protecting the system. The main reason behind such decisions was to minimize the number of free variables and reduce the complexity of designing, conducting and analysing the results of this first proof-of-concept study. Moreover, the data collected during the experiment considered many of the other reasonable factors that could influence malware attacks such as user profiling, user behaviour, host configuration, and environment.

4 STUDY DESCRIPTION

This first experiment of its kind was conducted from November 2011 to February 2012 as a proof-of-concept study involving 50 participants. 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 AV products in a real-world environment;
(2) Determine how malware infects computer systems and identify source of malware infections;
(3) Determine how phenomena such as the system configuration, the environment in which the system is used, and user behaviour affect the probability of infection of a system.

4.1 Ethics Clearance

The project was examined and cleared by the two relevant university entities: the *Comité d'évaluation des risques informatiques* (CERI, i.e., the computer security risks evaluation committee) and the *Comité d'éthique de la recherche* (CER, i.e., research ethics review committee).

Computer Risks. We provided users with an AV product that was centrally managed on our own server to guarantee high-availability. The AV software was updated daily and configured to perform a full scan of the computer every day, to provide an equal or better level of protection than average corporate or home users would normally have. Should the AV detects an infection, it would be automatically neutralized. Conversely, in the event that our diagnostics tools detected an infection on the computer that had been undetected by the AV, a procedure was given to users so they could neutralize the threat by themselves.

Giving that the experiment implied manipulation of malware files, special precautions were taken to protect the university IT infrastructure. All malicious or potentially malicious files were first encrypted and copied to DVDs before being stored in the high security zone of the laboratory.

Moreover, all computers were analysed by being connected to an isolated network to prevent any contamination of the university network.

Ethical and Privacy Considerations. Following the computer security risks evaluation committee clearance, the research ethics review committee cleared our recruiting procedures, the experimental protocol, as well as the measures adopted for user anonymity and confidentiality of the data collected.

To ensure the anonymity of users, we assigned each user a unique identification (ID) number associated with his computer. The only personal information kept for administrative and financial purposes was the participant's name, email address, and telephone number. This information was only accessible by the project leader and was destroyed 3 months after the end of the study. All raw data and statistics generated during the experiment were sanitized. The data was stored in a locked cabinet in the high-security zone of the laboratory, which is protected with three-factor authentication (biometrics, PIN, and ID card). This work zone is completely isolated from the Internet and the university network. 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 information security standards.

Only authorized personnel within the context of the project was able to access the data. In the event we wanted to share the anonymized data with other researchers, they had to agree to comply to the university computer risks and ethics policy. Moreover, the data collection was bound to the purpose of the project's research objectives. Finally, if we had inadvertently discovered information leading a reasonable person to believe that a (serious) crime had been committed or was about to be committed, we would have been required by law to advise the appropriate authorities (law enforcement agencies, etc.). Fortunately, this was not the case in this experiment.

4.2 Equipment

The laptops provided to the subjects all had identical configurations, with the following software installed: Windows 7 Home Premium; Trend Micro's OfficeScan 8.0; monitoring and diagnostic tools including HijackThis, ProcessExplorer, Autoruns, SpyBHORemover, SpyDLLRemover, tshark, WinPrefetchView, WhatChanged; and custom Perl scripts developed for this experiment. These tools and their use in our experiment are described in Section 4.3.

Scripts were used to automate the execution of the tools and compile statistical data about system configuration, the environments in which the system was used, and the manner of use. The data compiled by our scripts included:

- The list of applications installed;
- The list of applications for which updates were available;
- The number of web pages visited per day;
- The number of web pages visited by categories per month;
- The number and type of files downloaded from the Internet;
- The number of different hosts to which the laptop communicated;
- The list of the different locations from which the laptop established connection to the Internet;
- The number of hours per day the laptop was connected to the Internet;
- The number of hours per day the laptop was powered on.

Before deployment, we profiled the laptops to establish a baseline data set to compare at later date the variation in infection rates induced by AV and hardware choices vs. that generated by variation in demographics, behaviour, and software configuration. The recorded information included: (i) a hash of all files plus information about whether the files were signed; (ii) a list of auto-start programs; (iii) a list of processes; a list of registry keys; a list of browser helper objects (BHO); (iv) a list of the files loaded during the booting process; and (v) a list of the pre-fetch files.

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 AV clients installed on the laptops were thus sending relevant information to our server about any malware detected or suspected infections as they occurred.

4.3 Experimental Protocol

Subject Recruiting. We recruited by advertising the experiment on the Université de Montréal campus (which includes the engineering school and the business school) using posters and newspapers. Even though the recruiting process was centered on the university campus, the study was open to everyone. Interested participants were invited to visit a designated web site to obtain more details and fill a short on-line questionnaire that we used to collect initial demographic information such as gender, age, status and field of expertise. The only inclusion criteria was to be at least 18 years old.

Given our limitation on study sample size (number of laptops available), an important issue was to select a sample of 50 users who were as representative as possible of the general population of Internet users. Due to the over-representation of students and the limited number of candidates, we selected users based on a cluster sampling technique where users were grouped by their demographic characteristics. While this approach was suitable for a first study, recruiting for larger-scale trials should be more rigorously structured, as is the case for medical clinical trials.

In-person Sessions. Users were required to attend five face-to-face sessions: an initial session where they received their laptop and four monthly sessions where we collected the data and analyzed the computer. Participants were invited to book their appointments via an on-line calendar system hosted on our server. To encourage subjects to remain in the study, we paid them for each session attended. If they completed all sessions, then a bonus was paid out; in total, if a subject attended all sessions they would receive a sum equivalent to the cost of the laptop, along with a small additional compensation.

Initial Session. The intent of this short session was to obtain each user's informed consent and provide them with their laptop. Each user had to read and sign the informed consent form to confirm their participation in the study. Thereafter, the laptop was sold at a below retail-market price to the users, with laptops staying in users' possession after the study. This option was chosen for legal reasons and to foster user ownership of their computer, in the hope of reducing experiment bias in user behaviour. The only restrictions imposed were that they were not allowed to do the following during the study: (i) format the hard drive, (ii) install another operating system, (iii) delete our tools and the data collected, (iv) install another AV product, and (v) create a new disk partition. In addition, users were asked to answer an initial questionnaire to collect general information for their profile, such as gender, age group, status (worker, student, unemployed), field of expertise (computer science, natural science, art, and humanities), and self-reported level of computer expertise.

Monthly Sessions. During the monthly sessions, users answered an online questionnaire. The aim of this questionnaire was to assess user experience and opinion of the AV product, gain insights about how the computer was used, determine their level of security awareness, and their reported due diligence exerted to secure their computers. Meanwhile, statistical data compiled by the scripts were collected on the computer by the experimenter. The computer was also analyzed following a strict, fixed protocol, to look for malware missed by the AV product. The following diagnostic tools were used:

- HijackThis: gives the list of auto-loading programs and services, BHOs, plugins, toolbars, and so on;
- ProcessExplorer: shows the list of active processes;
- Autoruns: gives the complete list of programs configured to run during system bootup or login;
- Sigcheck: shows file version number, timestamp information, and digital signature details, including certificate chains;
- SpyBHORemover: gives the list of installed BHOs and classifies them in four categories (dangerous, suspicious, safe, unrated);
- SpyDLLRemover: gives the list of loaded DLLs and classifies them in three categories (dangerous, safe, unrated);
- Whatchanged: scans for modified files and registry entries;
- Winprefetchview: reads prefetch files and displays information stored in them.

We classified each element in four categories (safe, dangerous, suspicious, unrated) using external on-line resources, such as *www.systemlookup.com*, *www.processlibrary.com*, VirusTotal [50], and Anubis [17]. Computers with files identified as dangerous or suspicious were suspected to be infected, and any unrated files were subject to an in-depth investigation to see if they had malicious purposes. If the AV product detected malware over the course of the month, or if our diagnostic tools indicated that the laptop was infected or suspected to be, then users were asked to answer an additional questionnaire. This specific questionnaire collected more information regarding the potential means and sources of the infection, and on any behavioural changes observed on the computer. Moreover, additional consent was requested from the users to collect specific data, such as the browser history, network traffic data from tshark log files, and the suspicious file(s). These data were collected to help us identify the vector and the source of the infection.

Final Session. The final session was similar to the other monthly sessions. However, users answered a post-experiment questionnaire about their overall experience in the study. This final survey helped us identify activities or mindsets that may have unduly affected the experimental results. We also requested that users keep their experiment data for an additional period of 3 months in the event we might need to perform more in-depth analysis of their computer. Finally, we provided procedures to stop the automatic collection of the data, delete the data and the tools we installed, and reinstall the operating system, if they wanted to do so.

5 ANTIVIRUS EVALUATION

To evaluate the AV product, we analysed the detections (blocked malware attacks) and the missed detections (successful malware attacks) occurring over the course of the experiment. Additionally, users' questionnaire responses were compiled to provide an overall picture of the AV's subjective performance.

5.1 Threats Detected by Antivirus

During the 4-month study, 380 suspicious files were detected on 19 different user machines by the AV product being evaluated. However, some of these files were detected multiple times on the same user machine. Removing these repetitions, we obtain a total of 95 unique detections. Figure 1 shows the frequency of unique detections. The minimum number of detections observed per user is 0, the maximum is 28, and the average number of detections per user is 1.19 ($SD = 4.46$). Among

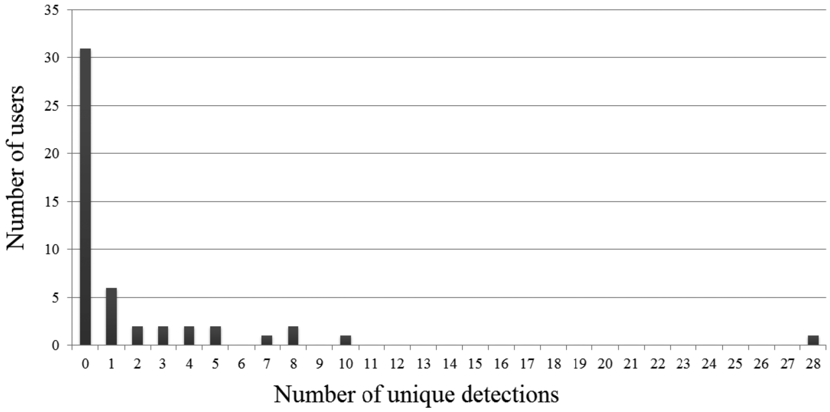

Fig. 1. Frequency histogram of unique detections.

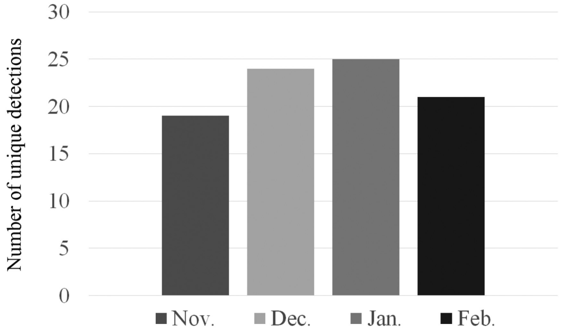

Fig. 2. Unique malware detections per month.

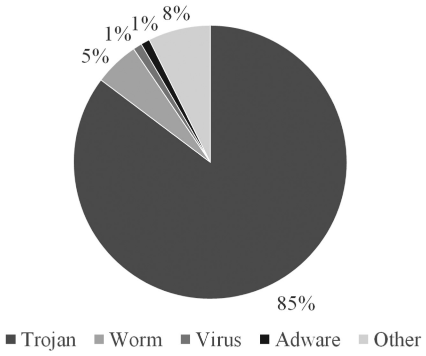

Fig. 3. Malware detections by type.

those 95 unique detections, we were able to trace the source of infection and determine that 17 of these propagated through portable storage devices (USB key or external hard drive).

In terms of overall virulence, 38% of the users were exposed to computer threats over a period of 4 months. More importantly, however, these results indicate that, if they are representative of the whole Internet population, one out of three newly installed machines would have been infected within 4 months if they had not had an AV installed. This figure aligns with the Eurostat Annual Report [10] indicating 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 (e.g., AV, anti-spam, firewall). Regarding the evolution of detections over time, the level of monthly detections is quite stable, as shown in Figure 2.

Detections were classified based on information provided by the AV product. As illustrated in Figure 3, most detections were classified as trojans, while viruses and adware had a relatively weak representation. These results are somewhat similar to those reported for overall detections by other AV vendors for the same period. For example, the 2011 Annual Report from Panda Security [39] indicates that trojans account for most detections with a ratio of 73%, while worm, virus, adware and other have respective ratios of 8%, 14%, 3%, and 2%. The differences with our results could be partially attributed to differences in the classification methods. For example, a file might 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 95 samples, while Panda Security has access to thousands of different samples and a user base of several millions users.

5.2 Missed Threats

The process of identification and classification of missed detections was based on user reporting of suspicious machine behaviour, monthly analysis of logs from the diagnostic tools, and results of automated queries to on-line sources with respect to processes and files found on the machine, and start-up programs (obtained automatically by scripts that we wrote).

Overall, 20 possible infections were detected on 10 different machines. The most useful diagnostic tool was HijackThis, which was involved in identifying 18 of the suspected infections. SpyBHORemover uncovered one additional infection. The last suspected infection was reported by the user, who contacted the project manager when he suspected that his machine had been infected. Except for the user-reported suspected infection, all suspicious files were captured during the monthly visits. 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. From this analysis, we classified 12 samples as unwanted software, seven as adware, and one as rogueware, for a total of 20 missed threats.

The 12 detected unwanted software and 7 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. However, it was unclear if the adware samples were indeed malicious, in that they show additional behaviour (e.g., theft of personal/private information) that might have further consequences for the user. The last sample was identified as rogueware—a software that pretends to be an AV program but does not provide any security. As previously mentioned, the corresponding user informed the project manager that his laptop was probably infected. It turned out that the laptop was infected with a fake AV product (AV Security Scanner). Warning 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. To get rid of these infections, the user was invited by the rogueware to register and provide his contact and payment information. At that moment, the user suspected that he may be infected and contacted the project manager. 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 could detect this threat *a posteriori*.

Overall, 20 missed threats were detected on 10 machines, which represents 20% of users. If we consider only missed malware, i.e., the seven adware and the rogueware, then 12% of users got infected. One point of comparison is the 2009 SurfRight report on real-world malware statistics [45]. Over a period of 55 days, 107,435 users scanned their machine with the Scan Cloud product. Among those, Scan Cloud found that 32% of protected machines were infected, compared to 46% of unprotected machines. In comparison with our 20% and 38% ratio, it would appear that our users were 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 users already suspect that their machines are infected, therefore resulting in an important self-selection bias. In all cases, direct comparison with our study is difficult given the fact that the time-period and the definition and classification methods for threats are quite different.

5.3 Antivirus Efficacy

The *efficacy* of the AV product (AE) is a function of the number of actual threats detected, i.e., the *true-positives* (TP), and the number of threats missed, i.e., the *false negatives* (FN):

$$AE = \frac{TP}{TP + FN}. \tag{1}$$

If we add the 20 threats that were not detected by the AV (FN) to the 95 unique detections (TP), then the AV has been exposed to a total of 115 threats:

$$AE \approx 0.8261. \tag{2}$$

Therefore, the AV product provided an efficacy of 83% . More specifically, this result represents the sensitivity of the AV to properly identify threats, including malware, and potentially unwanted software. If we only consider missed malware, i.e., the seven adware and the rogueware, then the efficacy raises to 92%. In comparison, the test performed by PC Security Labs [40] for the same period reported an efficacy of 99% for Trend Micro, and AV-Comparatives [2] reported an efficacy of 98% for the same product and period. These differences in performance between our test and the commercial tests suggest that AV protection differs between real-life conditions and controlled conditions. In other words, AV *field efficacy*, i.e., how the AV performs in actual use, is lower than AV efficacy from in-lab evaluations where testers have greater control of the testing environment.

5.4 User Experience

We evaluated user experience with the AV product through monthly surveys. We assessed their opinion regarding the level of interference, the information provided by the AV, the perceived level of protection provided, and their attitude toward the AV product. In addition, every time malware was detected by the AV, or the laptop was suspected to be infected by our diagnostic tools, we also collected insights on how users interact with the AV and potential computer threats while using their system.

Experience with the AV. Descriptive statistics relating to user perceptions for the 4 months (M1, M2, M3, M4) of the study are presented in Table 1. We computed the relative frequency of responses for questions Q1 to Q4, and the arithmetic mean (AM) and the standard deviation (SD) for questions Q5 to Q8.

Overall, 1/3 of users mentioned that the level of interaction (Q1) required by the AV was not enough and 2/3 judged that the required level was adequate. Only a few users found the level of interaction too high. Those findings are also confirmed by the monthly results on the level of interference (Q5), which ranged from 3.0/10 to 3.4/10, where 1 meant no interference and 10 meant high interference. It is worth mentioning that the AV evaluated (i.e., Trend Micro OfficeScan) was a business product that was configured to be silent. Pop-up windows would only appear in the case of a detection, which could explain why one third of the users found that the level of interaction was not enough. One potential explanation could be that for those users, interaction with the AV provides reassurance that they are protected [12].

When evaluating the information provided by the AV (Q2), half the users found that the level of information was adequate and half found it was not enough. The product was configured to only give the name and the path of the file detected, the action from the AV, and generic information on the family. For half the users, this information was not sufficient, meaning that a minimalist design might not be appropriate for all users. The level of usefulness (Q6) of the information provided

Table 1. User Experience and Opinion Per Month

	M1	M2	M3	M4
Q1 Level of interaction required by the AV				
Too frequent	2%	4%	4%	2%
Adequate	68%	59%	55%	60%
Not enough	30%	37%	41%	38%
Q2 Amount of information provided by the AV				
Too much	2%	2%	0%	0%
Adequate	47%	51%	49%	48%
Not enough	51%	47%	51%	52%
Q3 Response to a pop-up window from the AV				
I read it and follow its suggestions	60%	59%	65%	68%
I read it but don't follow its suggestions	11%	6%	10%	4%
I close it without reading it	6%	16%	4%	12%
Other (specify)	23%	19%	21%	16%
Q4 Feeling when an AV's pop-up window appears				
Comforted to know that the AV is working	60%	70%	63%	72%
Annoyed that the AV is interrupting me	19%	4%	10%	4%
Don't notice	15%	10%	10%	8%
Other (specify)	6%	16%	17%	16%
	AM (SD)	AM (SD)	AM (SD)	AM (SD)
Q5 Level of interference (1 to 10)	3.4 (2.8)	3.2 (2.7)	3.0 (2.3)	3.4 (2.6)
Q6 Level of usefulness of the information (1 to 10)	5.7 (2.2)	6.1 (2.5)	6.0 (2.4)	6.3 (2.4)
Q7 Level of protection (1 to 10)	7.8 (2.0)	7.7 (1.9)	7.5 (2.1)	7.7 (1.8)
Q8 Level of understanding of the information (1 to 10)	6.8 (2.2)	6.8 (2.3)	7 (2.5)	6.7 (2.6)

ranged from an average of 5.7/10 to 6.3/10, where 1 meant useless and 10 meant very useful. We also evaluated if the information provided by the AV was presented in a manner that could be easily understood by users (Q8). Monthly average results ranged from 6.8/10 to 7.0/10, where 1 meant difficult to understand and 10 meant easy to understand.

We also asked users how they felt (Q4) and reacted (Q3) when they saw a pop-up window from the AV. More than 2/3 of users said they feel comforted to know that the AV is working. The last third was either annoyed that the AV was interrupting them or did not notice any pop-up. Most users who answered "Other" mentioned that they did not get any pop-ups from the AV. Examples of other answers included: "I don't want to see the pop-up when I am watching a movie, but other times, I don't care," "I don't understand what happens," "It doesn't bother me," or "I feel annoyed because there is a virus on my computer." Regarding how users reacted when they saw a pop-up window (Q3), almost 2/3 reported that they read and followed the suggestions of the AV. Over the 4 months, between 4% and 11% of the users said they read the pop-ups but ignored the suggestions. And between 4% and 16% of users said they closed the pop-ups without reading them. Most participants who answered "Other" mentioned that they did not experience any pop-up from the AV. Some "Other" answers included: "I read it as well as the suggestions but I take the action

I want," "I read it and sometimes I follow its suggestions," "It depends," "I read it quickly and if it's important I follow the suggestions, if it's not, I close it," and "I ask someone else to take care of it."

The perceived level of protection (Q7) provided by the AV was also evaluated over the study. The monthly averages ranged from 7.5/10 to 7.8/10, where 1 meant very low protection and 10 meant high protection. We computed the average level of perceived protection provided over the 4 months for users that had at least one detection from the AV, and for users that had no detection. Users that experienced a detection by the AV over the study had an average level of perceived protection of 7.38/10 ($SD = 1.66$), and users that had no detection had an average of 7.83/10 ($SD = 1.48$). A Mann-Whitney U-test was conducted and no statistically significant difference was found between users with and without detections; $U = 243.00$, two-tailed exact p-value $= 0.31$. We also conducted a Mann-Whitney U-test to investigate if there was a significant difference between users that had at least one missed threat by the AV ($A.M. = 7.07, SD = 1.37$), and users that had no infection ($A.M. = 7.81, SD = 1.57$). Results of the test indicated that there is no significant difference; $U = 133.50$, two-tailed exact p-value $= 0.11$. Although no significant difference was found, users that experienced no detection or missed threats over the study reported a marginally higher level of perceived protection.

Experience with Computer Threats. Additional information on user experience with potential computer threats were collected during the monthly sessions when the AV detected malware or when we suspected the laptop to be infected. All concerned users agreed to answer the additional questionnaire and provide more specific data about the system's activity.

As part of the questionnaire, users were asked to report any strange computer behaviour they might have experienced over the last month. Of the 40 reports, 22 users said they had not observed strange behaviours, 2 said they did not know, and 16 answered yes. Examples of strange behaviour included annoying pop-ups, music starting to play, new web browser home page, web search redirections, and changes in computer performance (e.g., crashes or slowdowns). Among the reports that were related to missed threats and not threats detected by the AV, seven users said they did not observe strange behaviour, and eight said they did. While half the users observed behaviour that are known to be warning signs of malware infection, the other half did not notice anything abnormal on their computer even though they were infected with some form of computer threat that was missed by the AV.

We also asked users if they remember receiving any security-related messages from the system or the AV. Interestingly, only half of users answered yes. Among those, six said they felt comforted, four mentioned that they felt annoyed by the interruption, three were confused, six were worried about the security of their computer, and one answered "Other" ("I was concerned about the computer's security, but I would like to proceed on that"). In addition, we asked users to report what they were doing when the message appeared. Only five users said they did not remember. The other 15 responses included: visiting web sites (N = 3), downloading software/files from the Internet (N = 9), using a portable storage device (N = 2), and watching a movie (N = 1). While those comments are not sufficient to establish the exact transmission vectors used, they suggest potential user involvement in the infection process, whether blocked by the AV or successful.

6 USER PROFILING AND BEHAVIOUR

We examined whether user demographics, characteristics, and certain types of user behaviour led to a higher probability of malware attack. We first divided users in two groups. The first group contains *high-risk* users, which are those who experienced at least one malware attack, whether blocked or successful, and the second group contains *low-risk* users who had no malware attack

Table 2. Proportion of Users for Each Factor

Factor		Total sample (N = 50 users)	High-risk group (N = 23 users)	Low-risk group (N = 27 users)
Gender	Male	60%	61%	59%
	Female	40%	39%	41%
Age	18–24	38%	35%	41%
	25–40	46%	61%	33%
	41+	16%	4%	26%
Employment status	Student	64%	70%	59%
	Worker	30%	26%	33%
	Unemployed	6%	4%	8%
Field of expertise	Computer Science	26%	22%	30%
	Natural Science	52%	48%	56%
	Arts/Humanities	22%	30%	14%
Computer expertise	High	18%	30%	7%
	Low	82%	70%	93%

during the experiment. Table 2 shows the user distribution between the total sample, the *high-risk* group, and *low-risk* group, based on user characteristics and demographic factors.

The risk analysis was determined based on the calculation of the odd ratio (OR)—a measure of the degree of association between a risk (or protective) factor and an outcome. It represents the ratio between the probability that an outcome will occur in a group exposed to a factor of interest and a reference group that is not exposed. Given that A is the number of individuals in the exposed group who developed the outcome, B is the number of individuals in the exposed group who did not developed the outcome, C is the number of individuals in the reference group who developed the outcome, and D is the number of individuals in the reference group who did not develop the outcome, the OR can be calculated as follows:

$$OR = \frac{A * D}{B * C}. \tag{3}$$

An OR larger than 1 indicates that the factor of interest is a risk factor. An OR smaller than 1 means that the exposure is a protective factor. And if the OR equals 1, the outcome is equally likely in both groups. The confidence interval (CI) in which the true value of the OR is likely to be must also be taken into account when interpreting the OR. Hence, if 1 is included in the CI, nothing can be said on the association between the factor and the outcome.

6.1 Characteristics and Demographic Factors

Risk analysis through OR was performed to assess if particular user characteristics and demographics increase the odds of malware attack. Malware attack was used as the outcome, indicated by either 1 or 0, depending on whether the user experienced any malware attack during the experiment. The factors of interest were gender, age, status, field of expertise, and self-reported level of computer expertise. Female, 18–24 age group, unemployed, arts/humanities, and low self-reported level of computer expertise were used as the reference groups for gender, age, status, field of expertise, and self-reported level of computer expertise respectively. Results of the analysis are summarised in Table 3. For each factor, we computed the OR, the 95% CI, and the p-value as a measure of statistical significance. For the purpose of our analysis, items marked with * were considered as statistically significant at p-value < 0.05.

Table 3. Odds Ratio of User Characteristics and Demographic Factors

Factor		OR	(95% CI)	p-value
Gender	Male vs. Female	1.06944	(0.34339–3.33061)	0.90778
Age	25–40 vs. 18–24	2.13889	(0.62071–7.37039)	0.02934 *
	41+ vs. 18–24	0.19643	(0.01999–1.92938)	0.07169
Employment status	Worker vs. Unemployed	1.33333	(0.09772–18.19174)	0.94323
	Student vs. Unemployed	2.00000	(0.16442–24.32783)	0.46665
Field of expertise	Computer Science vs. Arts/Humanities	0.72917	(0.15303–3.47431)	0.55542
	Natural Science vs. Arts/Humanities	1.16667	(0.30169–4.51161)	0.58426
Computer expertise	High vs. Low	5.46875	(1.00696–29.70058)	0.04907 *

Gender. The total sample included 30 males and 20 females, which gives a proportion of 60% and 40% respectively. Table 2 shows that the gender distribution among the 23 *high-risk* users is very similar to the total sample, indicating that gender may not be a significant risk factor for malware attack. This was supported by the statistical analysis where no significant difference between males and females (Table 3) was found with respect to the risk of malware attack.

In comparison with previous studies that investigated the effect of gender, six out of eight studies reported a significant gender effect. Some researchers [23, 41, 52] found that males were more at risk than females, and others [4, 18, 36] found that females were more susceptible to computer threats than males. While our results are in line with the studies that reported no significant effect [14, 21], direct comparison is not possible; we studied malware attacks while Grimes et al. [14] used (self-reported) reception of spam and Kumaraguru et al. [21] investigated phishing susceptibility. When looking only at studies that focused on malware attacks [23, 52], males were found to be more at risk of encountering malware than females. This discrepancy with our results could be attributed to differences in study design, target population, and sample size; Yen et al. [52] studied malware encounters of corporate users within a large enterprise, and Lalonde Lévesque et al. [23] based their study on malware encounters of millions of Windows users. Although prior work [23, 52] suggests that gender is a significant correlate of malware attack, further studies should be conducted to validate the direction of the aforementioned correlation, if any.

Age. We divided 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 2 shows that the proportion of 18 to 24 year olds in the *high-risk* group is almost the same as for the total sample. For those 25 to 40, the proportion in the *high-risk* group (61%) is higher than for the total sample (46%), which could suggest that this age group is more susceptible to malware attack. And for the 41+ age group, we observe a decrease of 12% in the proportion between the total sample (16%) and the *high-risk* group (4%). Results from the analysis (Table 3) revealed a significant difference between the 25–40 age group and the reference group (18–24). However, as the value 1 is included in the 95% CI, nothing can be said on the nature of the association, that is whether it is a risk factor or a protective factor.

Similarly to most prior work that included the effect of age, our statistical results mildly suggest it may be a contributing factor associated with the risk of malware attack. In comparison, some researchers found younger users to be more susceptible to phishing [18, 21, 43] and malware attacks [23, 35], while [41] found older users to be significantly more at risk of (self-reported) identity theft. Bossler et al. [4] and Grimes et al. [14] reported no significant age effect on (self-reported) data loss from malware infection and (self-reported) reception of spam respectively.

These discrepancies can be explained, first, because the experimental methods are quite different: some studies involved surveys of users where susceptibility levels are evaluated through user self-declarations of previous incidents, and not from actual observation. Second, these results are not (all) specific to malware attacks. Finally, the granularity of the age data recorded is different so it is hard to precisely compare 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 of malware attack and age. Large-scale studies based on alternate data sources, other time frames, and different analysis methods will be required to settle the issue of age as a contributing factor for malware attack.

Employment Status. Users were classified in three self-declared categories: student, worker, or unemployed. Table 2 indicates that the proportion between the total sample and the *high-risk* group is quite similar for each category, suggesting that employment status may not be a contributing factor of malware attacks. This was confirmed by the risk analysis (see Table 3) where no statistically significant difference is shown between the different categories.

In contrast, prior work that studied the effect of employment status found employed users to be at higher risk of (self-reported) data loss from malware infection [4] and (self-reported) identity theft [41]. Given that unemployed users represented only 6% of our study, it is possible that our sample was simply too small to observe any significant difference.

Field of Expertise. We recruited users based on their field of work or study to have a heterogeneous sample. As shown in Table 2, 26.5% of users were self-identified as being in computer science, 47% in natural science, 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 risk analysis (Table 3) show no statistically significant effect for the field of expertise.

To the opposite, Yen et al. [52] found that job types have a significant impact on the risk of malware encounters; jobs from the top of the enterprise organizational tree and jobs requiring higher technical expertise had a greater likelihood of malware encounter. Similarly, Thonnard et al. [48] identified directors, high-level executives, and personal assistants to be at higher risk of targeted attacks compared to other jobs. Finally, Lee et al. [28] found that some areas of work are associated with increased risk of being subjected to targeted phishing attacks, suggesting that it is the area of expertise that leads users to be of interest to attackers. Although prior studies found that the field of expertise may be a contributing factor, direct comparison with our results is not possible as we studied *home users* while they focused on non home-user domains (e.g., industry, government, academia).

Computer Expertise. We assessed computer expertise by asking users about their proficiency with certain technical tasks. Users were considered to have a high self-reported 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 self-reported computer "experts" for the purposes of our analysis. As observed in Table 2, those with high expertise were nearly twice as likely to be in the *high-risk* group when compared to the total sample. This may indicate that a high level of expertise increases the risk of malware attack, which was confirmed by the statistical analysis. More specifically, users with high self-reported level of computer expertise were found to be 5.47 times more likely to experience malware attack than users with low expertise.

Although our results are somewhat counterintuitive, they are consistent with the work of Ovelgonne et al. [38] and Yen et al. [52]. Ovelgonne et al. [38] identified software developers to be more prone to malware attack, and Yen et al. [52] found people with technical expertise to be

Table 4. Odds Ratio of Behavioural Factors

Factor		OR	(95% CI)	p-value
System activity		1.00047	(0.99972–1.00121)	0.22066
Applications installed		1.00678	(0.99449–1.01922)	0.28083
Outdated applications		1.03763	(0.75098–1.43369)	0.82282
Connection time		1.00369	(1.00044–1.00697)	0.02618 *
Hosts contacted		1.00002	(1.00000–1.00005)	0.04969 *
Default web browser	Firefox vs. IE	1.83333	(0.39238–8.57580)	0.74626
	Chrome vs. IE	5.10714	(1.17708–22.15903)	0.03005 *
Web pages visited		1.00007	(1.00002–1.00013)	0.00697 *
Files downloaded		1.00007	(0.99956–1.00196)	0.21369
P2P activity	Yes vs. No	13.63636	(2.60209–71.46171)	0.00199 *

more at-risk of encountering malware. In opposition, Onarlioglu et al. [37] found *computer security expertise* to be a protective factor. A possible explanation is that self-reported expert users are more at risk of malware attack, because they know just enough to get themselves into trouble. For example, they may have a false sense of self-confidence that leads them to engage in more risky behaviours. Another potential explanation could be that users with high computer expertise have a high risk-seeking profile, which lead them to engage in risky behaviours. One last explanation could be that expert users are *heavy* computer users (they spend more time online, they download more applications from the Internet, etc.), which contributes, intentionally or not, to increasing their odds of getting exposed to malware.

Summary of User Characteristics and Demographic Factors. In summary, we found little evidence linking user demographics and characteristics to increased risk of malware attack. Gender, student/employment status and field of expertise showed no statistically significant differences. However, we did find partial support linking age and self-reported level of computer expertise to the risk of malware attack.

6.2 Behavioural Factors

To assess if specific user behaviour led to a higher risk of malware attacks, we focused our analysis on the following factors: system activity, application installs, network usage, web browser usage, web pages visited, files downloaded, peer-to-peer (P2P) activity, and level of vigilance. Data was collected through scripts on the computer and self-reported questionnaires. Using a similar approach to that described in Section 6.1, we conducted a risk analysis based on the calculation of the OR. In the case of continuous variables, the OR is interpreted in terms of each unit increase on the variable; for each increase by one unit, the odds of the outcome is multiplied by the OR. Table 4 summarises the statistical results; items marked with * were considered statistically significant at p-value < 0.05.

System Activity. The activity of the system was measured by scripts using the number of hours per day the laptop was on. To study its impact on the risk of malware attack, we computed the total number of hours the laptop was on for the entire duration of the study. The total system usage ranged from 109 to 2,882h, with an average of 1,629h (SD = 778). When comparing groups, *high-risk* and *low-risk* users had their laptop on for an average of 1,793h (SD = 656) and 1,522h (SD = 863), respectively. Results from the analysis in Table 4 show no significant relationship between the system activity and the risk of malware attack. Hence, our analysis suggest that the system

Table 5. Type of Applications Installed Per Month

	High-risk group				Low-risk group			
	M1	M2	M3	M4	M1	M2	M3	M4
Most of the applications are games	5%	4%	0%	0%	0%	11%	8%	4%
Most of the applications are not games	90%	73%	61%	70%	88%	63%	60%	59%
No application was installed	0%	9%	30%	26%	8%	22%	28%	37%
Other	5%	14%	9%	4%	4%	4%	4%	0%

Table 6. Type of Applications Installed by Others Per Month

	High-risk group				Low-risk group			
	M1	M2	M3	M4	M1	M2	M3	M4
Most of the applications are games	14%	5%	5%	4%	0%	8%	4%	0%
Most of the applications are not games	19%	18%	17%	22%	15%	18%	8%	26%
I don't know	0%	5%	17%	9%	8%	4%	11%	0%
No one besides me has installed applications	62%	72%	61%	61%	73%	70%	77%	70%
Other	5%	0%	0%	4%	4%	0%	0%	4%

activity—as measured by the number of hours the system was on—does not seem to be a significant factor for malware attack.

Application Installs. We monitored using scripts the daily number of applications installed by each user. To assess the potential effect on the risk of malware attack, we computed for each user the total number of applications installed over the 4 months. Users installed between 2 and 177 applications, with an average of 70 (SD = 47) applications. The *high-risk* group installed on average 75 (SD = 46) applications, while the *low-risk* group installed 61 (SD = 47) applications on average. However, this difference was not found to be significant from the risk analysis (see Table 4). In contrast, Ovelgonne et al. [38] found a significant positive correlation between the number of binaries installed, and the number of attempted attacks per host. For comparison, we computed the correlation between the number of unique malware attacks and the number of applications installed. The Gamma statistic, a non-parametric correlation coefficient, was used, because our data on malware attacks contains many tied observations. Similarly to Ovelgonne et al., we found a weak significant positive relationship (G = 0.24, p-value = 0.04, N = 50) between the number of applications installed and the number of malware attacks. This seems plausible as installing many applications can contribute to increased probability of being exposed to malware, either by a malicious application, or by a legitimate application that may install unwanted software.

We also investigated the type of applications that were installed. Users were asked through the monthly survey what type of applications they installed the most (see Table 5), and what type of applications was installed by other people (see Table 7). Table 5 shows that the majority of applications installed over the study were not reported as games. Moreover, there does not seem to be major differences in the type of applications between users in the *high-risk* group and in the *low-risk* group, given the high level categorization used in the questionnaire. From Table 6, we see that the majority of users reported that no one besides them has installed applications on their computer. When comparing *high-risk* and *low-risk* groups, *high-risk* users more frequently reported that others had installed applications on their computer, which could suggest that *high-risk* users are more likely to let other people use their computer.

Table 7. Most Frequently Used Applications per Month

	High-risk group				Low-risk group			
	M1	M2	M3	M4	M1	M2	M3	M4
Web browser	86%	82%	91%	83%	88%	92%	96%	88%
Office Suite	0%	5%	9%	13%	8%	4%	0%	4%
Mail application	0%	0%	0%	0%	0%	0%	4%	4%
Games	0%	0%	0%	0%	0%	0%	0%	4%
Other	14%	13%	0%	4%	4%	4%	0%	0%

Table 8. Second Most Frequently Used Applications per Month

	High-risk group				Low-risk group			
	M1	M2	M3	M4	M1	M2	M3	M4
Web browser	14%	23%	17%	26%	8%	11%	8%	15%
Office Suite	43%	45%	61%	52%	46%	48%	54%	63%
Mail application	10%	9%	4%	0%	19%	15%	12%	7%
Games	10%	5%	4%	4%	15%	22%	19%	11%
Other	23%	18%	14%	18%	12%	4%	7%	4%

The survey also asked the most frequent and the second most frequent type of applications used. From Table 7, we can see that between 82% and 96% of users used a web browser most frequently. Half of participant reported that the Microsoft Office suite was second most frequently used (Table 8), followed by web browser and other. Comparison between the *high-risk* and the *low-risk* groups does not suggest major differences; they both reported web browser and Office suite as their most and second most frequently used applications.

In addition, we also investigated the number of applications for which updates were available, as outdated applications may increase the odds of malware infection. We computed the 4-month average number of outdated applications per user. Overall, users had on average between 3 and 11 outdated applications, with a mean of 7 (SD = 2) outdated applications. When looking at the *high-risk* and the *low-risk* group, both had on average 7 outdated applications. Based on the risk analysis (Table 4), the average number of outdated applications does not seem to be a significant risk factor.

Network Usage. User network activity was evaluated in terms of time spent online, number of different hosts contacted, and reported primary connection location. To assess the relationship between the time online and the risk of malware attack, we computed using scripts the total number of hours each laptop was connected to the Internet for the entire duration of the study. The connection time varied from 11 to 992h, with an average of 242h per user (SD = 229). *High-risk* users were connected on average 328h (SD = 273), while *low-risk* users were connected on average 169h (SD = 155). Results from the risk analysis in Table 4 show a weak significant positive association between the connection time and the risk of malware attack (OR = 1.00369). For each 100h connected online, the odds of malware attack increase by 1.048 (1.00369^{100}).

The daily number of different hosts contacted by the laptop was also collected over the 4-month period. For each user, we computed using scripts the total number of hosts contacted during the entire study. Users contacted between 18 and 1,508,833 hosts during the 4 months, with an average of 60,433 hosts per user (SD = 211,244). *High-risk* users contacted a higher number of hosts during the study than *low-risk* users; they respectively contacted on average 107,268 (SD = 309,065) and 20,536 (SD = 21,867) hosts. From the risk analysis in Table 4, there is a weak significant association

Table 9. Primary Location from Which the Laptop was Connected to the Internet per Month

	High-risk group				Low-risk group			
	M1	M2	M3	M4	M1	M2	M3	M4
Home	81%	82%	78%	70%	81%	78%	85%	86%
University campus	9%	18%	18%	26%	11%	15%	4%	7%
Work	0%	0%	4%	0%	8%	7%	11%	7%
Coffee shop	5%	0%	0%	4%	0%	0%	0%	0%
Other	5%	0%	0%	0%	0%	0%	0%	0%

Table 10. Installed Web Browsers

	Total sample	High-risk group
IE	78%	70%
Firefox	58%	65%
Chrome	66%	78%

Table 11. Most Frequently Used Web Browser

	Total sample	High-risk group
IE	30%	17%
Firefox	30%	26%
Chrome	40%	57%

between the number of hosts contacted and the risk of malware attack. However, as the value 1 is included in the CI, nothing can be said about the nature of the association.

We also asked users through the monthly survey the primary location from which the laptop was connected to the Internet. When looking at the results in Table 9, between 70% and 86% of users answered home as their primary connection location, followed by university campus (4%–26%), and work (4%–11%). Both *high-risk* and *low-risk* groups reported home as their primary location, suggesting that primary location may not be a contributing factor to malware attack.

Web Browser Usage. Each month, users were asked which web browser was installed, which one they used most and if they have changed the default security and privacy settings of their browsers. For each factor, except for the question related to the default settings, we also collected real data usage from scripts during each monthly meetings. We therefore prioritized, when possible, real data usage for our analysis instead of self-reported data obtained through surveys.

Table 10 presents the proportion of users that installed each web browser during the study, and Table 11 summarises the proportion of users who used each web browser. An increase of 17% is observed between the total sample and the *high-risk* group for Chrome. In contrast, the proportion decreases for Firefox and Internet Explorer (IE). When looking at the risk analysis in Table 4, Chrome was identified as a significant risk factor. Users with Chrome as their default browser were found to be 5.11 times more likely to experience malware attacks than users of IE. While these results suggest that having Chrome as a default web browser is a significant correlate of malware attacks, they do not imply that using Chrome is in itself a contributing risk factor. Possible explanations could be differences in browser's architecture or threats landscape. Another potential explanation could be differences in users. For example, Chrome users might have a high risk-seeking profile, or be *heavier* computer users compared to IE users.

As many web browser offer advanced security and privacy settings, such as anti-phishing or anti-malware protection, we also investigated the effect of those changes on the risk of malware attack. Out of 50 users, only 4 changed the default security and privacy settings of their main browser. One disabled cookies for Chrome, another asked Chrome to remember all of his passwords, and the last one decided not to keep his IE temporary files. Since only a small proportion of users changed their default settings (see Table 12), we cannot draw any conclusion on the effect of those changes.

Table 12. Security and Privacy Default Settings

	Total sample	High-risk group
Using default settings for all browsers	94%	96%
Made changes for Internet Explorer	2%	0%
Made changes for Firefox	0%	0%
Made changes for Chrome	4%	4%
Other	0%	0%

Web Pages Visited. The number of web pages visited was also recorded for the entire duration of the study to evaluate the impact on the risk of malware attack. This factor was computed from the browser history using scripts and represents the total number of web pages visited by user. In total, users visited on average 18,531 (SD = 17,008) web pages. The *high-risk* group visited on average 26,624 (SD = 20,822) web pages while the *low-risk* group visited on average 11,637 (SD = 8,426) web pages. The risk analysis (see Table 4) reveals a weak positive association between the total number of web pages visited and the risk of malware attack (OR = 1.00007); for each 100 web pages visited, the odds of malware attacks increase by 1.007.

Our results confirm the general trend that the more a user surfs the web, the more likely he is to be exposed to computer threats. In comparison with previous work, Canali et al. [5] also found that visiting many web pages increases the chance of visiting a malicious web page. In another study, Carlinet et al. [6] reached a similar conclusion: heavy web activity, as measured by the web traffic, increases the likelihood of generating malicious traffic.

We further analysed if particular categories of web pages were more prone to be associated with malware attacks. To this end, each web page visited was classified using the Site Safety Center of Trend Micro [49]. Overall, 70 different categories of web pages were found. We performed a risk analysis based on the calculation of the OR using the 22 most popular categories (see Table 13). In total, 10 categories were found to be significant: streaming media/MP3, peer-to-peer, social networking, software downloads, email, personal network storage/file download servers, search engines/portals, games, entertainment, and computers/Internet. Among those, peer-to-peer, software downloads and personal network storage/file download servers were identified as the more risky. For each 100 web pages visited in these categories, the odds of malware attacks are multiplied respectively by 15.58, 10.60, and 7.56.

In comparison, Symantec [46] identified the following 10 web site categories as the most *at-risk* of being malicious in 2011: blogs/web communications, hosting/personal web site, business/economy, shopping, education/reference, technology and Internet, entertainment and music, automobile, health and medicine, and pornography. Our findings are also similar to the results of Yen et al. that identified six web site categories as being associated with higher risk of encountering malware; chat, file transfer, freeware, social networks, and streaming. In another study, Canali et al. [5] also identified that specific web site categories, such as pornography and adult content, were at higher risk of being malicious. Overall, those results suggest that *high-risk* categories are not limited to what common sense traditionally associates with higher risk, such as hacking and pornography.

Files Downloaded. For each user, we collected using scripts the number of files downloaded from the Internet over the study. During the 4 months, users downloaded between 19 and 3,341 files, with an average of 496 (SD = 588) files downloaded per user. Over the study, *high-risk* users (AM = 604, SD = 488) downloaded more files from the Internet than *low-risk* users (AM = 386, SD = 638). Though this may indicate that the volume of files downloaded from the Internet is a contributing

Table 13. Odds Ratio by Web Page Categories

Factor	OR	OR^{100}	(95% CI)	p-value
Streaming media/MP3	1.00168	1.18277	(1.00032–1.00305)	0.01582 *
Peer-to-peer	1.02784	15.57943	(1.00089–1.05551)	0.04276 *
Social networking	1.00018	1.01816	(1.00002–1.00034)	0.02440 *
Software downloads	1.02388	10.59024	(1.00642–1.04165)	0.00716 *
Pornography	1.00299	1.34791	(0.99702–1.00901)	0.32590
Email	1.00054	1.30234	(1.00005–1.00102)	0.02890 *
Personal network storage/ file download servers	1.02044	7.56393	(1.00455–1.03658)	0.00697 *
News/media	1.00072	1.07463	(0.99984–1.00161)	0.11084
Shopping	1.00037	1.03769	(0.99959–1.00115)	0.35423
Chat/Instant messaging	1.00626	1.86647	(0.98623–1.02669)	0.54266
Search engines/portals	1.00056	1.05758	(1.00001–1.00110)	0.04485 *
Internet infrastructure	1.00788	2.19221	(0.99985–1.01598)	0.05454
Games	1.00736	2.08195	(1.00046–1.01431)	0.03642 *
Government/legal	1.00389	1.47439	(0.99933–1.00847)	0.09495
Entertainment	1.00409	1.50406	(1.00051–1.00767)	0.02508 *
Travel	1.00091	1.09522	(0.99906–1.00277)	0.33586
Blogs/web communications	1.00669	1.94794	(0.99861–1.01483)	0.10476
Financial services	0.99934	0.93611	(0.99745–1.00124)	0.49574
Business/economy	1.00104	1.10954	(0.99954–1.00254)	0.49574
Politics	0.99404	0.55003	(0.97494–1.01352)	0.54603
Computers/Internet	1.00127	1.13533	(1.00007–1.00246)	0.03688 *
Education	1.00055	1.05652	(0.99963–1.00147)	0.23837

Table 14. Odds Ratio by Type of Files Downloaded

Factor	OR	(95% CI)	p-value
docx	1.14990	(0.82516–1.60244)	0.40939
rar	1.34096	(0.84383–2.13097)	0.21444
zip	1.03869	(0.98748–1.09256)	0.14115
pdf	1.00359	(0.99866–1.00855)	0.15379
exe	1.06230	(1.00846–1.11908)	0.02276 *
doc	0.95267	(0.83164–1.09134)	0.48439
ppt	1.15924	(0.93941–1.43051)	0.16839
jpg	1.01833	(0.99514–1.04207)	0.12224
gif	1.06268	(0.94728–1.19213)	0.29995

factor of malware attacks, this factor was not found to be significant from our risk analysis (see Table 4).

We further investigated if specific types of files were associated with higher risk of malware attacks. We computed the OR for each file extension that had more than 100 files downloaded (see Table 14). Among the 9 types of files, only the extension *exe* was found to be a significant risk factor (OR = 1.06230). In comparison, Ovelgonne et al. [38] also that found a positive association between the percentage of downloaded binaries from the web and the number of attempted malware attacks per host. Given that many malware are distributed via the Internet, it seems plausible that heavy

Table 15. Computer Security Expertise

	High-risk group	Low-risk group
Configured a firewall	17%	19%
Secured a wireless network	43%	44%
Changed the default security settings of a web browser	39%	48%
None of the above	22%	22%

downloading of executable files contributes to increased risk of being exposed to malware. The remaining question is whether those executable files were downloaded by the users or if they were silently downloaded from the Internet as a result of drive-by-download attacks.

P2P Activity. As part of the monthly survey, we asked users to report how often they have used peer-to-peer networks to download audio, video files, or other software on the laptop. Overall, 14 users reported having engaged in peer-to-peer activities during the study. Among those, 12 were in the *high-risk* group and 2 in the *low-risk* group; suggesting P2P activity could be a risk factor. This was confirmed by the risk analysis in Table 4 where a strong significant association was identified between P2P activity and the risk of malware attack (OR = 13.63636). Users that reported engaging in P2P activity were found to be 13.64 times more likely to experience malware attack than users who did not. Our finding provides evidence that engagement in P2P activity might be a contributing risk factor of malware attack. This seems plausible as P2P networks are known to be a popular medium for spreading malware [19].

Level of Vigilance. User level of vigilance was evaluated based on the level of security awareness and the measure of due diligence they exert to secure their laptops. Each month, users were required to report which of the following tasks they had previously completed: configured a firewall, secured a wireless network, and changed the default security and privacy settings of a web browser. Overall, 18% of users configured a firewall, 44% secured a wireless network, 44% changed the default security and privacy settings of a web browser, and 40% completed none of the above. As shown in Table 15, both groups reported similar expertise in computer security. Based on the number of tasks each user had previously completed, we computed a computer security score ranging from 0 to 3. From there, we performed a Mann-Whitney U-test and found no significant difference between both groups; U = 239.00, two-tailed exact p-value = 0.74. Though we found computer expertise to be a significant risk factor, this was not the case for computer *security* expertise.

We also evaluated through the monthly survey users' level of concern about the security of their laptop. The level of concern ranged from 1 to 10, where 1 meant low concern and 10 meant high concern. The 4-month average for the total sample was 7.27 (SD = 2.06). The *high-risk* group and the *low-risk* group reported similar level of concern; they respectively had an average level of concern of 7.14 (SD = 2.23) and 7.38 (SD = 1.94). In addition, we asked users to report on the tasks they performed, if any, to secure their laptop. Table 16 shows that a higher proportion of users in the *high-risk* group answered that they are concerned but they do not know what to do to secure their laptop from being compromised. In contrast, a higher proportion of users in the *low-risk* group said that they know what to do and they actively perform theses tasks. The most common tasks mentioned were in order: avoid visiting dangerous and suspicious web sites, perform updates, avoid illegal downloading from the Internet, regularly scan computer, do not open suspicious files from the Internet, and perform risky actions in a virtual machine. Overall, we found no evidence linking the level of concern and the risk of malware attack. Rather, results suggest that being concerned is not sufficient if not combined with the adoption of safe computer behaviour.

Table 16. Concern About the Computer's Security Per Month

	High-risk group				Low-risk group			
	M1	M2	M3	M4	M1	M2	M3	M4
Typically not concerned	13%	13%	9%	13%	7%	11%	4%	11%
Concerned but do not know what to do	43%	52%	48%	48%	41%	30%	33%	30%
Know what to do but too busy	22%	17%	13%	22%	11%	4%	19%	15%
Know what to do and perform these tasks	13%	13%	26%	13%	37%	48%	41%	41%
Other	9%	4%	4%	4%	4%	7%	4%	4%

Summary of User Behaviour. We have identified six significant factors related to user behaviour: volume of network usage, number and types of web pages visited, default web browser, types of files downloaded from the Internet, and P2P activity. A high volume of network usage, as estimated by the time spent online and the number of hosts contacted, was identified as a risk factor. Similarly, visiting many web pages as well as certain categories of web pages were found to be a contributing risk factor. We also found an association between the main web browser used and the risk of malware attack. Finally, downloading executable files from the Internet, and engaging in P2P activity were both found to increase the risk of malware attack.

7 STUDY LIMITATIONS

The results we presented and discussed here are subject to certain limitations and potential bias that may threaten the internal and external validity of our study. Internal validity refers to the strength of the inferences from the study, that is the extent to which no other variables except the one we studied caused the results. While external validity refers to the ability to generalize the results to a more universal population.

First, the AV performance evaluation is limited to only 95 detected threats – a very small number compared to the numerous threats in the wild, especially considering that some of these may be false positives. As those threats were detected by an antivirus product, they depend on the efficacy of the later, which may lead to an underestimation of malware detections. In addition, the false negative number might also be underestimated, because we cannot guarantee that our protocol caught all malware missed by the AV. In other words, we do not have absolute ground truth.

Second, even though we were able to identify several factors correlated with the risk of malware attacks, these factors in themselves are not sufficient to explain the *causal link* leading to malware infection. To this effect, a more detailed analysis of the collected data is required to determine the sources and means of infection for each of the 115 detected threats. Only then will we be able to determine which of these factors are causes of infection, and which are consequences of other factors that were not included in this study. Moreover, another limitation of our study is its susceptibility to confounding. Although we included in our analysis many variables that could influence the risk of malware attacks, and we fixed some of the external factors (same AV, laptop, OS, geographic area), we cannot guarantee that our results were not affected by other unknown extraneous variables that may confound the results. It would be interesting in future work to consider additional variables, such as culture, risk averseness, or risk propensity of users.

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

One obvious limitation to the external validity of our study derives from our studied sample. First, subjects were located in the same geographic area. Second, their demographics (age and gender) and characteristics (status, field of expertise, computer expertise) distribution differ from that of the global Internet population. Third, we studied home users. Hence, results in terms of AV evaluation and risk factors may be different for non home-user domains (e.g., industry, government, academia). For example, corporate users may be exposed to different computer threats, or be targeted based on their corporation's characteristics. Fourth, our studied sample is limited to Windows 7 laptops protected by one antivirus product. Hence, our findings do not provide insight into other versions of Windows (e.g., Windows Mobile, Vista, Windows 10, etc.), non-Windows systems such as MacOS and Unix-based OS, other AV products, and other types of devices (e.g., tablet, mobile, desktop).

In addition, our findings may not be representative of other time frames. As security data are known to be dynamic, a similar study conducted at another time-period may lead to different results. This could be particularly true as malware, computer defences, and users evolve over time. Finally, our study was limited to mass market malware attacks. That is, we did not intended to study targeted attacks and zero-day attacks.

8 DISCUSSION AND CONCLUSION

In this article, we presented the results from the first computer security clinical trial of AV software performed with real users in non-laboratory conditions. Similar to clinical trials in medicine, we evaluated the real-life performance of AV software in protecting systems and studied how users interact with the AV, the system and malware attacks as they occurred in the wild. While the studied sample 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 malware attacks risk factors and defence effectiveness.

In terms of AV evaluation, our results show that 38% of users were exposed to a malware attack blocked by the AV, indicating that at least 38% of the users could 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 computer threats that was not detected by the AV. Interestingly, half of these users did not observe strange behaviour on their laptop even though they were infected. While AV *field efficacy* was estimated at 92%, this performance is below the protection reported by commercial tests for the same product and period. Perhaps this is like vaccine efficacy: Since real-life conditions are frequently suboptimal compared with clinical conditions, vaccine protection is often lower than in clinical tests. A similar dynamic may also be taking place with AV product where AV protection is lower with real-life conditions compared to in-lab evaluations [24]. Finally, the evaluation of the user experience with the AV product revealed variance in results, indicating that one single AV and/or configuration may not accommodate all types of users [9].

In terms of risk factors, our results indicate that age, gender, field of expertise, and employment status are not significant correlates of malware attacks. However, we found partial support linking self-reported level of computer expertise to the risk of malware attacks. Users who self-reported high level of computer expertise were found to be more susceptible. Regarding user behaviour, we identified six significant factors; volume of network usage, number and types of web pages visited, default web browser, types of files downloaded from the Internet, and P2P activity. High volume of network usage, and web pages visits were associated with increased risk of malware attacks. We also observed some surprising patterns in web usage, with seemingly innocuous categories of sites such as search engines/portals and computers/Internet being associated with a higher rate of malware attack while more "shady" sites such as those containing pornography content were less so. In addition, using Chrome as default web browser, downloading executables files from the Internet, and engaging in P2P activity were also found to increase the odds of malware attacks. Overall, results suggest that malware attacks may be more a function of frequency and type of online behaviour, rather than based on user characteristics and demographic factors.

Beyond the contribution of these results, this work demonstrates that computer security clinical trials have potential implications for the AV industry. First, it could provide AV testers a viable and complementary approach to tests conducted in controlled environments. Given the realism of the environment and the independence of the malware selection process, tests performed in real-life conditions are less prone to controversy and ethical issues, such as the creation of malware samples. While studies comparing multiple AV or other security products will require more users to get statistically significant results, increasing use of automation should allow such tests to be performed at relatively modest cost. Second, such studies could be suitable for AV vendors seeking to: (i) understand how their products perform in real-world usage, (ii) identify which aspects of the product (user interface, detection, remediation, etc.) could be further improved, and (iii) identify user groups for which they are more (or less) effective at preventing malware infections. A better understanding of what works best in real-life for specific user groups could help support the design of successful *user-tailored* AV products [24].

In addition, computer security clinical trials are of potential utility to help understand what user characteristics, demographic factors, and behaviour lead to higher risk of malware attacks. This knowledge could be used to improve the content and targeting of user education and training [23, 36], as well as support the development of user risk model [5, 25, 52] for the cyberinsurance industry. To this end, it is important that further research be conducted to assess the multi-level factors of malware attacks. More studies performed in real-life conditions, such as the Security Behavior Observatory [11], are needed to validate our findings and investigate factors that were not included in our study. We hope the work presented here illustrates the merits of future larger-scale computer security clinical trials.

REFERENCES

[1] Shahid Alam, Ibrahim Sogukpinar, Issa Traore, and Yvonne Coady. 2014. In-cloud malware analysis and detection: State of the art. In *Proceedings of the 7th International Conference on Security of Information and Networks*. ACM, 473.

[2] AV Comparatives. 2013. *File Detection Test of Malicious Software*. Technical Report. AV Comparatives.

[3] J. Blackbird and B. Pfeifer. 2013. The global impact of anti-malware protection state on infection rates. In *Proceedings of the Virus Bulletin International Conference*.

[4] Adam M. Bossler and Thomas J. Holt. 2009. On-line activities, guardianship, and malware infection: An examination of routine activities theory. *Int. J. Cyber Criminol.* 3, 1 (2009), 400.

[5] Davide Canali, Leyla Bilge, and Davide Balzarotti. 2014. On the effectiveness of risk prediction based on users browsing behavior. In *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security*. ACM, 171–182.

[6] Y. Carlinet, L. Mé, H. Débar, and Y. Gourhant. 2008. Analysis of computer infection risk factors based on customer network usage. In *Proceedings of the 2nd International Conference on Emerging Security Information, Systems and Technologies (SECURWARE'08)*. 317–325.

[7] Tudor Dumitras. 2011. Field data available at Symantec research labs: The worldwide intelligence network environment (WINE). In *Proceedings of the ASPLOS Exascale Evaluation and Research Techniques Workshop*.

[8] Simon P. G. Edwards. 2013. Four Fs of anti-malware testing: A practical approach to testing endpoint security products. In *Proceedings of the Workshop on Anti-malware Testing Research (WATeR'13)*. IEEE, 1–9.

[9] Serge Egelman and Eyal Peer. 2015. The myth of the average user: Improving privacy and security systems through individualization. In *Proceedings of the New Security Paradigms Workshop*. ACM, 16–28.

[10] Eurostat. 2011. Nearly one-third of internet users in the EU27 caught a computer virus. Retrieved from http://ec.europa.eu/eurostat/documents/2995521/5028026/4-07022011-AP-EN.PDF/22c742a6-9a3d-456d-bedc-f91deb15481b.

[11] Alain Forget, Saranga Komanduri, Alessandro Acquisti, Nicolas Christin, Lorrie Faith Cranor, and Rahul Telang. 2014. Building the security behavior observatory: An infrastructure for long-term monitoring of client machines. In *Proceedings of the Symposium and Bootcamp on the Science of Security*. ACM, 24.

[12] Steven Furnell. 2010. Usability versus complexity—Striking the balance in end-user security. *Netw. Secur.* 2010, 12 (2010), 13–17.

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

[14] Galen A. Grimes, Michelle G. Hough, and Margaret L. Signorella. 2007. Email end users and spam: Relations of gender and age group to attitudes and actions. *Comput. Human Behav.* 23, 1 (2007), 318–332.

[15] David Harley. 2009. Making sense of anti-malware comparative testing. *Info. Secur. Tech. Rep.* 14, 1 (2009), 7–15.

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

[17] International Secure Systems Lab. 2013. Anubis malware analysis for unknown binaries. Retrieved from https://anubis.iseclab.org/.

[18] Tom N. Jagatic, Nathaniel A. Johnson, Markus Jakobsson, and Filippo Menczer. 2007. Social phishing. *Commun. ACM* 50, 10 (2007), 94–100.

[19] Andrew Kalafut, Abhinav Acharya, and Minaxi Gupta. 2006. A study of malware in peer-to-peer networks. In *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement*. ACM, 327–332.

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

[21] Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Theodore Pham. 2009. School of phish: A real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security*. ACM, 3.

[22] Fanny Lalonde Lévesque and José M. Fernandez. 2014. Computer security clinical trials: Lessons learned from a 4-month pilot study. In *Proceedings of the 7th USENIX Conference on Cyber Security Experimentation and Test*. USENIX Association.

[23] Fanny Lalonde Lévesque, José M. Fernandez, and Dennis Batchelder. 2017. Age and gender as independent risk factors for malware victimisation. In *Proceedings of the 31th International British Human Computer Interaction Conference*.

[24] Fanny Lalonde Lévesque, José M. Fernandez, Dennis Batchelder, and Glaucia Young. 2016. Are they real? Real-life comparative tests of anti-virus products. In *Proceedings of the 26th Virus Bulletin International Conference*. 25–33.

[25] Fanny Lalonde Lévesque, Jose M. Fernandez, and Anil Somayaji. 2014. Risk prediction of malware victimization based on user behavior. In *Proceedings of the 9th International Conference on Malicious and Unwanted Software: The Americas (MALWARE'14)*. IEEE, 128–134.

[26] Fanny Lalonde Lévesque, Jude Nsiempba, José M. Fernandez, Sonia Chiasson, and Anil Somayaji. 2013. A clinical study of risk factors related to malware infections. In *Proceedings of the ACM SIGSAC Conference on Computer & Communications Security*. ACM, 97–108.

[27] Fanny Lalonde Lévesque, Anil Somayaji, Dennis Batchelder, and Jose M. Fernandez. 2015. Measuring the health of antivirus ecosystems. In *Proceedings of the 10th International Conference on Malicious and Unwanted Software (MALWARE'15)*. IEEE, 101–109.

[28] Martin Lee. 2012. Who's next? Identifying risks factors for subjects of targeted attacks. In *Proceedings of the Virus Bulletin International Conference*. 301–306.

[29] Fanny Lalonde Lévesque, C. R. Davis, J. M. Fernandez, S. Chiasson, and A. Somayaji. 2012. Methodology for a field study of anti-malware software. In *Proceedinsg of the Workshop on Usable Security (USEC'12)*. LNCS, 80–85.

[30] Fanny Lalonde Lévesque, C. R. Davis, J. M. Fernandez, and A. Somayaji. 2012. Evaluating antivirus products with field studies. In *Proceedings of the 22th Virus Bulletin International Conference*. 87–94.

[31] Gregor Maier, Anja Feldmann, Vern Paxson, Robin Sommer, and Matthias Vallentin. 2011. An assessment of overt malicious activity manifest in residential networks. In *Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer, 144–163.

[32] Andreas Marx. 2000. A guideline to anti-malware-software testing. In *Proceedings of the 9th Annual European Institute for Computer Antivirus Research Conference*. 218–253.

[33] G. R. Milne, L. I. Labrecque, and C. Cromer. 2009. Toward an understanding of the online consumer's risky behavior and protection practices. *J. Consum. Affairs* 43 (2009), 449–473.

[34] Igor Muttik and James Vignoles. 2008. Rebuilding anti-malware testing for the future. In *Virus Bulletin Conference*.

[35] Fawn T. Ngo and Raymond Paternoster. 2011. Cybercrime victimization: An examination of individual and situational level factors. *Int. J. Cyber Criminol.* 5, 1 (2011), 773–793.

[36] Daniela Oliveira, Harold Rocha, Huizi Yang, Donovan Ellis, Sandeep Dommaraju, Melis Muradoglu, Devon Weir, Adam Soliman, Tian Lin, and Natalie Ebner. 2017. Dissecting spear phishing emails for older vs. young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, 6412–6424.

[37] Kaan Onarlioglu, Utku Ozan Yilmaz, Engin Kirda, and Davide Balzarotti. 2012. Insights into user behavior in dealing with internet attacks. In *Proceedings of the Network and Distributed System Security Symposium (NDSS'12).*

[38] Michael Ovelgönne, Tudor Dumitras, B. Aditya Prakash, V. S. Subrahmanian, and Benjamin Wang. 2017. Understanding the relationship between human behavior and susceptibility to cyber attacks: A data-driven approach. *ACM Trans. Intell. Syst. Technol.* 8, 4 (2017), 51.

[39] Panda Security Labs. 2011. Panda Labs Annual Report 2011 Summary. Retrieved from https://www.pandasecurity.com/mediacenter/src/uploads/2012/01/Annual-Report-PandaLabs-2011.pdf.

[40] PC Security Labs. 2013. *Security Solution Review on Windows 8 Platform.* Technical Report. PC Security Labs.

[41] Bradford W. Reyns. 2013. Online routines and identity theft victimization further expanding routine activity theory beyond direct-contact offenses. *J. Res. Crime Delinq.* 50, 2 (2013), 216–238.

[42] Imtithal A. Saeed, Ali Selamat, and Ali M. A. Abuagoub. 2013. A survey on malware and malware detection systems. *International Journal of Computer Applications* 67, 16 (2013), 25–31.

[43] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs. 2010. Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'10).* 373–382.

[44] A. Somayaji, Y. Li, H. Inoue, J. M. Fernandez, and R. Ford. 2009. Evaluating security products with clinical trials. In *Proceedings of the USENIX Workshop on Cyber Security Experimentation and Test (CSET'09).*

[45] SurfRight. 2009. Real-World malware statistics: October/November 2009. Retrieved from http://files.surfright.nl/reports/HitmanPro3-RealWorldStatistics-OctNov2009.pdf.

[46] Symantec Corporation. 2012. Internet security threat report 2011 trends. Retrieved from http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_2011_21239364.en-us.pdf.

[47] The WildList Organization International. 2017. The WildList. Retrieved from https://www.wildlist.org/.

[48] Olivier Thonnard, Leyla Bilge, Anand Kashyap, and Martin Lee. 2015. Are you at risk? Profiling organizations and individuals subject to targeted attacks. In *Proceedings of the International Conference on Financial Cryptography and Data Security.* Springer, 13–31.

[49] Trend Micro. 2012. Website classification. Retrieved from http://solutionfile.trendmicro.com/solutionfile/Consumer/new-web-classification.html.

[50] Virus Total. 2013. Virus total. Retrieved from https://www.virustotal.com.

[51] J. Vrabec and D. Harley. 2010. Real performance? In *Proceedings of the European Institute for Computer Antivirus Research Annual Conference (EICAR'10).*

[52] Ting-Fang Yen, Victor Heorhiadi, Alina Oprea, Michael K. Reiter, and Ari Juels. 2014. An epidemiological study of malware encounters in a large enterprise. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security.* ACM, 1117–1130.

[53] Righard Zwienenberg, Richard Ford, and Thomas Wegele. 2013. The real-time threat list. In *Proceedings of the 23rd Virus Bulletin International Conference.*

Received October 2017; accepted April 2018