INSIGHTS INTO USER BEHAVIOR IN DEALING WITH COMMON INTERNET ATTACKS

A THESIS

SUBMITTED TO THE DEPARTMENT OF COMPUTER ENGINEERING

AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

By

Utku Ozan Yılmaz

July, 2011

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Ali Aydın Selçuk (Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Özgür Ulusoy

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Mustafa Akgül

Approved for the Graduate School of Engineering and Science:

Prof. Dr. Levent Onural
Director of the Graduate School

ABSTRACT

INSIGHTS INTO USER BEHAVIOR IN DEALING WITH COMMON INTERNET ATTACKS

Utku Ozan Yılmaz
M.S. in Computer Engineering
Supervisor: Assist. Prof. Dr. Ali Aydın Selçuk
July, 2011

The Internet's immense popularity has made it an attractive medium for attackers. Today, criminals often make illegal profits by targeting Internet users. Most common Internet attacks require some form of user interaction such as clicking on an exploit link, or dismissing a security warning dialogue. Hence, the security problem at hand is not only a technical one, but it also has a strong human aspect. Although the security community has proposed many technical solutions to mitigate common Internet attacks, the behavior of users when they face these attacks remains a largely unexplored area.

In this work, we describe an online experiment platform we built for testing the behavior of users when they are confronted with common, concrete attack scenarios such as reflected cross-site scripting, session fixation, scareware and file sharing scams. We conducted experiments with more than 160 Internet users with diverse backgrounds. Our findings show that non-technical users can exhibit comparable performance to knowledgeable users at averting relatively simple and well-known threats (e.g., email scams). While doing so, they do not consciously perceive the risk, but solely depend on their intuition and past experience (i.e., there is a training effect). However, in more sophisticated attacks, these non-technical users often rely on misleading cues such as the "size" and "length" of artifacts (e.g., URLs), and fail to protect themselves. Our findings also show that trick banners that are common in file sharing websites and shortened URLs have high success rates of deceiving non-technical users, thus posing a severe security risk.

Keywords: Simulated attacks, internet security, user behavior.

ÖZET

YAYGIN İNTERNET SALDIRILARININ ÜSTESİNDEN GELMEDE KULLANICI DAVRANIŞINA İÇGÖRÜLER

Utku Ozan Yılmaz
Bilgisayar Mühendisliği, Yüksek Lisans
Tez Yöneticisi: Assist. Prof. Dr. Ali Aydın Selçuk
Temmuz, 2011

İnternetin yüksek popülaritesi, onu saldırganlar için çekici bir ortam haline getirmiştir. Günümüzde suçlular sık sık İnternet kullanıcılarını hedef alarak yasadışı kazanç sağlamaktadırlar. Birçok yaygın İnternet saldırısı bir istismar bağlantısına tıklamak ya da bir güvenlik uyarısı diyaloğunu azletmek gibi birtakım kullanıcı etkileşimlerine ihtiyaç duymaktadır. Bundan dolayı, eldeki güvenlik problemi sadece teknik olmayıp güçlü bir insani yöne de sahiptir. Güvenlik topluluğu yaygın internet saldırılarını azaltmak için birçok teknik çözüm önermiş olmasına rağmen, kullanıcıların bu saldırılarla karşılaştıklarındaki davranışları büyük ölçüde keşfedilmemiş bir alan olmayı sürdürmektedir.

Bu çalışmada, kullanıcıların yansıyan siteler arası betik yazma, oturum sabitleme, scareware ve dosya paylaşım dolandırıcılıkları gibi yaygın, somut saldırı senaryolarıyla yüzleştiklerindeki davranışlarını test etmek için kurduğumuz bir çevrimiçi deney platformunu tanımlıyoruz. Çeşitli geçmişlere sahip 160'tan fazla İnternet kullanıcısıyla deneyler yürüttük. Bulgularımız, teknik olmayan kullanıcıların görece basit ve iyi bilinen tehditleri (örneğin e-posta dolandırıcılıkları) önlemede bilgili kullanıcılarla kıyaslanabilir performanslar sergileyebildiklerini göstermektedir. Bunu yaparken tehlikeyi bilinçli bir şekilde idrak etmeyip, yalnızca sezgilerine ve geçmiş deneyimlerine (yani bir eğitim etkisi var) güvenmektedirler. Fakat, daha gelişmiş saldırılarda, bu kullanıcılar sıklıkla yapıların (mesela URL'ler) "boyut" ve "uzunluk"ları gibi yanıltıcı ipuçlarına dayanmakta, ve kendilerini korumakta başarısız olmaktadırlar. Bulgularımız ayrıca, dosya paylaşım sitelerinde yaygın olan hile afişlerinin ve kısaltılmış URL'lerin, teknik olmayan kullanıcıları kandırmada yüksek oranda başarılı olduklarını, bu nedenle ciddi bir güvenlik tehlikesi yarattıklarını göstermektedir.

Anahtar sözcükler: Benzetimli saldırılar, internet güvenliği, kullanıcı davranışı.

Acknowledgement

I would first like to thank my supervisor Assist. Prof. Dr. Ali Aydın Selçuk, who encouraged me to pursue my masters studies in the field of computer security and lended me his support and guidance throughout my studies.

I also gratefully thank Assist. Prof. Dr. Engin Kırda for mentoring me during my visit to Institute Eurécom, where this thesis study was carried out in part; without his supervision this work would not have been possible.

I owe special thanks to Assist. Prof. Dr. Davide Balzarotti and Kaan for their invaluable contributions that vastly improved this work.

I would also like to thank all the great people in Eurécom for their assistance and friendship.

Lastly, I thank The Scientific and Technological Research Council of Turkey for financing my masters studies through the "National Scholarship Programme for MSc Students".

Contents

1 Introduction **1**

2 Common Internet Attacks **5**

3 Related Work **8**

4 Design of the Experiments **10**

4.1 Demographic Information 12

4.2 Web-Based Attacks 13

4.3 Email-Based Attacks 14

4.4 File Sharing-Related Attacks 15

4.5 General Security Warnings 16

4.6 Online Banking 17

4.7 Processing of the Results 17

5 Analysis of the Test Results **21**

5.1 Demographics and Diversity 21

5.2 Security and Risk Perception 24

5.3 Test-Specific Results 28

5.3.1 Web Tests Including IP Addresses and Shortened URLs . . 28

5.3.2 Trick Banner Tests 30

5.3.3 General Security Warning Tests 30

5.3.4 Online Banking Tests 32

6 Discussion and Insights Gained 34

6.1 Security Training and Risk Perception 34

6.2 Size Matters 37

6.3 Loud and Clear Warnings 38

6.4 URL Shortening Services and Tools 39

6.5 Trick Banners 40

6.6 Second Authentication Channels 41

7 Conclusion 43

List of Figures

4.1 List of test suites 11

4.2 Home page of our online test platform 12

4.3 Short briefing screen at the beginning of a test 12

4.4 Facebook wall post with a reflected cross-site scripting attack . . 13

4.5 Spam email with a generic text and a suspicious link 14

4.6 Search results for a movie in Filestube with a bad file extension and a suspicious size. 16

4.7 Online banking website 18

4.8 SMS including a one-time code to authenticate an online banking transaction 19

5.1 Five-number summaries of participant familiarity with typical computer tasks and concepts 22

5.2 Five-number summaries and probability densities of total scores . 25

5.3 Relationship between risk perception and security scores 27

5.4 Correct clicks for the trick banner tests on pages reproduced from popular file sharing websites 29

List of Tables

4.1 Applicability criteria for each of the statistical tests we used . . . 20

5.1 Five-number summaries for each test suite and participant group . 23

5.2 Results showing participants' decisions when confronted with a shortened URL and a raw IP link 28

5.3 Summary of the online banking test suite 32

Chapter 1

Introduction

The Internet has grown dramatically in the last twenty years and has changed the way we communicate, do business, maintain friendships, and acquire information. In fact, a study in 2008 showed that today's youth are developing important social and technical skills online [44]. It concluded that the learning process today is becoming increasingly peer-based and networked, and that education needs to be reconsidered in the 21st century. Without doubt, the Internet has become a critical infrastructure, and any disruption in services adversely affects our lives and causes significant damage (e.g., the recent Amazon EC2 cloud outage that affected several Fortune 500 companies and millions of users [24]).

Clearly, as the importance of an information medium increases, so does its attractiveness for criminal activity with the aim of making quick, illegal financial gains. In fact, because of their high popularity and a user base that consists of millions of Internet users, web applications have become primary targets for attackers. According to SANS [37], attacks against web applications constitute more than 60% of the total attack attempts observed on the Internet. Many web applications are exploited every day to convert trusted websites into malicious servers hosting client-side browser exploits. Once the victim's machine has been infected with malware, the attackers can then start collecting sensitive information such as credit card numbers, passwords, and financial information. According to SANS, most website owners fail to scan their applications for common flaws.

In contrast, from an attacker's point of view, automated tools designed to target specific web application vulnerabilities simplify the discovery and mass infection of websites.

Note that, although some types of attacks are technically very difficult for users to detect and prevent (e.g., a stored form of a cross-site scripting attack [29] on a popular social networking website), most Internet attacks actually require the interaction of the user (e.g., by clicking on an exploit link, installing risky software, failing to recognize a phishing website, ignoring an SSL certificate warning, etc.). Hence, unfortunately, the user often becomes the weakest link in the chain, and the attackers often rely on social engineering techniques to trick victims into engaging in risky behavior, thus compromising their security.

To date, the security community has proposed many solutions to mitigate current Internet threats such as botnets (e.g., [17, 34, 35]), malware (e.g., [18, 39, 62]), cross-site scripting (e.g., [61]), cross-site request forgery (e.g., [19]), and drive by download exploits (e.g., [47]). However, although it is clear that the problems at hand are not only technical (i.e., there is a strong human aspect as some form of user interaction is typically required for many of these attacks to be successful), existing literature is sparse, and there have not been many works that attempt to shed light on how Internet users are able to cope with concrete attacks.

In [26], Dhamija et al. attempted to understand which phishing attack strategies work better in practice and why. The paper provided the first empirical evidence on which malicious strategies are successful at deceiving general users by conducting experiments with 22 users.

Recently, Sunshine et al. [53] presented an empirical study of SSL warning effectiveness, which showed that users do not behave as expected and that they often exhibit dangerous behavior. Based on the lessons that they were able to learn, the authors conducted experiments with 100 users and designed new warnings that performed significantly better than the SSL certificate warnings used in browsers today.

Besides Dhamija et al.'s and Sunshine et al.'s previous studies that solely focused on phishing and SSL certificate warnings, there have not been any comprehensive studies on how users are able to cope with current threats and popular attacks on the Internet, and how they behave and make decisions in an adversarial setting. This work presents the first empirical findings that shed light on how different user groups deal with, and react to, common Internet attacks such as cross-site scripting, session fixation, malware infections, and file sharing scams.

Our findings show that although many users suspect and may be able to detect a potential attack, they make wrong decisions with respect to their security as they are not aware of the consequences of their actions. Also, our experiments demonstrate that users often have a completely wrong understanding of security risks, and what may constitute an attack.

Our findings empirically confirm the general intuition that security education has a significant effect in practice in preventing the more complex Internet attacks, and that a general "security awareness" is critical for user protection. We believe that online test systems such as the one we have constructed are useful in educating students and users about popular attacks on the Internet.

This work makes the following contributions:

- We present an online security test system that presents 50 typical benign and malicious use cases to users, and records their behavior. We have conducted empirical experiments with a diverse set of more than 160 users. To the best of our knowledge, the study we present is the largest that has been conducted to date on current Internet threats such as cross-site scripting and file sharing scams.
- We show that, while exposure to common attacks such as email scams helps users get familiar with and avert similar threats in future encounters, more advanced attacks (e.g., session fixation) are still only detected through security training. We also show that users with a security education are better at assessing the consequences of a possible threat and making the right decisions.

- We provide empirical evidence that many users treat "length" and "size" as a sign of maliciousness (e.g., length of URLs and size of files).
- We show that clear, direct and intimidating security warnings are more effective in conveying risk to users compared to detailed technical explanations.
- We show that trick banners that are common in file sharing websites have a high success rate of deceiving technically unsophisticated users, and therefore, pose an important security threat.
- We provide empirical evidence that non-technical users are frequently tricked by shortened URLs, and are largely not aware of simple web-security tools and services such as shortened URL expanders.
- We show that technically insufficient users are mostly deceived by man-in-the-middle attacks during online transactions and although second authentication channels help technically inclined people to avoid the attack, non-technical users fail to benefit from them.

The rest of this thesis study is structured as follows: In Chapter 2 we provide a brief overview of common Internet attacks. In Chapter 3, we summarize the related work. In Chapter 4, we present our experiment platform and give details of each test we performed on the participants. In Chapter 5, we show the results we obtained from the tests and list our observations. In Chapter 6, we discuss these results, and summarize the insights we distilled. Finally, in Chapter 7, we briefly conclude the thesis.

Chapter 2

Common Internet Attacks

In this chapter, we provide a brief overview of the common Internet attacks and explain some related terms.

- *Phishing* can be explained as directing users to fraudulent web sites [26]. It is generally used in conjunction with link manipulation techniques.
- *Link manipulation* is the effort of making a link's destination look different from it really is. It can be achieved in many ways. One of the most commonly utilized methods is taking advantage of the Domain Name System hierarchy. `http://www.paypal.hostding.com` can be given as an example to this method. In fact, it points to the paypal subdomain of the hostding domain. However, it can easily be mistaken as a legitimate link to a page in the Paypal domain by unknowledgeable or careless users. Another widely used method is putting a reliable destination to the displayed text of a link, while the link points to another, potentially dangerous destination. Although most browsers display the real destination when a user hovers his mouse on the link, users usually don't check it. There are also other link manipulation techniques like internationalized domain name homographic attack or using links containing the '@' symbol.
- *Cross site scripting (XSS)* is injecting client side scripts (e.g., JavaScript)

into interactive web pages as an input. In the most common reflected type, which we used in our tests, the script is embedded into a link and when the user clicks it, the vulnerable target website executes the script.

- *Session fixation* is fixing a user's session ID before the user even logs into the target server [40]. In order to achieve this, the attacker should create a link with a valid session identifier and convice the user to click on the link including the fixed session identifier.

- *URL shortening* is using an HTTP redirect on a short domain name to a long URL. It can be used to hide the underlying URL, thus sometimes it is used for malicious purposes.

- *IP address links* are links that contain an IP address instead of a domain name. Just like shortened URLs, IP address links can be used for malicious purposes.

- *Advance fee fraud* is a fraud that attempts to trick prospective victims into parting with funds by persuading them that they will receive a substantial benefit in return for providing some modest payment in advance [50]. The most famous example knows as the Nigerian scam or 419 scam, which we used in our tests, involves a person who is "allegedly" looking for someone to help him move some funds abroad by providing a bank account, in return for a generous comission. If the victim takes the bait, the attacker will ask for some advance payment to deal with a complication, and thus the victim is scammed [28].

- *Scareware* is a software which scares users into taking action by using fake threat warnings. The action can be something like buying a useless software or visiting a malicious website in order to get rid of the bogus threat. Sometimes, scareware can be in the form of web pop-up windows.

- *Trick banner* is a banner ad that tries to trick people into clicking it. In the case of file sharing websites, trick banners try to disguise themselves as legitimate download links.

- *Man-in-the-middle attack (a.k.a. bucket brigade attack)* is an attack where the attacker compromises the connection between two parties and then impersonates the other party to both parties. While the two parties believe that they are communicating with each other, in fact the communication is totally controlled by the man in the middle.

- *Man-in-the-browser attack* is a type of man-in-the-middle attack, where the man-in-the-middle is a browser trojan. In this attack, the trojan can change the details of an online transaction (e.g., the target account of a money transfer) without the user's knowledge.

- *Two-factor authentication* is an authentication method that requires two independent pieces of information (i.e., factors) to establish identity and privileges [51]. If these factors are transported over two different channels, it can also be named as two-channel authentication, and the supporting channel can be referred to as the second authentication channel.

Chapter 3

Related Work

Although attacks such as cross-site scripting, session fixation and social engineering-based malware are common on the Internet, there has not been any comprehensive empirical study to determine the awareness level of Internet users about these attack vectors, and how they are able to deal with such attacks in practice.

As mentioned before, one well-known work that attempts to understand why phishing strategies work was conducted by Dhamija et al. [26]. In another related effort, Sunshine et al. [53] presented an empirical study of SSL warning effectiveness. Note that both studies focused on very specific aspects of security (i.e., phishing and SSL warnings, respectively) while our study has a much broader scope, covering general web attacks (e.g., reflected cross-site scripting), file sharing and e-mail scams (e.g., fake search results and attached malicious files), general security warnings (e.g., antivirus messages), and attacks against financially sensitive operations (i.e., man-in-the-middle attacks targeting online banking websites).

Friedman et al. [31] conducted a number of general interviews about web security and concluded that many participants could not reliably determine whether a connection is secure. The participants were shown screenshots of a browser connecting to a site and they had to decide if the connection was secure or not. In

another study [32], Friedman et al. interviewed participants about their concerns on risks and potential harms of web use. In comparison, our work focuses on concrete technical web attacks such as session fixation, cross-site scripting, and malicious links provided by URL shortening services. We report our findings on how users behave when confronted with such realistic attacks.

A recent work by Conti et al. [25] suggests a taxonomy of malicious interface techniques. The authors conducted a survey on a group of users to measure their frustration and tolerance when they encounter such interfaces. However, this study does not discuss the effectiveness of such techniques at deceiving users.

Other researchers have attempted to measure the effectiveness of social engineering attacks in social networks. For example, in [38], Jagatic et al. have performed realistic phishing attacks on undergraduate students based on the information they were able to harvest from social networking websites. In another work, Bilge et al. [20] were able to show that users tend to have a higher level of trust in messages they receive from their social networks.

Finally, there exist several studies on the usability of security solutions. For example, in [22], Chiasson et al. describe a usability study they have conducted on 26 users which shows that some previously proposed security solutions have serious usability problems. In [23], Clark et al. present another study on the usability of anonymous web browsing. Note that compared to general existing work on the usability of security solutions, our main focus in this paper is on determining and understanding how users react to common attacks.

Chapter 4

Design of the Experiments

In this chapter, we describe the setup we developed to test and simulate the typical security threats that users may encounter in their everyday Internet usage. In order to be able to reach a large number of users with diverse backgrounds, we conducted online experiments using an interactive test platform.

We designed the experiments as a within-group study in which all participants responded to a series of tests in various security contexts. After studying a wide range of common Internet attacks that require some sort of user interaction or decision (e.g., reflected cross-site scripting attacks, advertisement trick banners, scareware pop-ups, etc.), we created 50 security-related scenarios, grouped in five test suites: web-based attacks, email-based attacks, file sharing-related attacks, application security warnings and online banking (See Figure 4.1).

After reaching the homepage of our online test platform (See Figure 4.2), we informed the participants that they were going to take part in "an experiment to determine the security-awareness of Internet users" and asked them to provide an email address. We also informed the participants that the tests would take about an hour. Moreover, at the beginning of each test, we provided them with the estimated time to complete it and a short explanation on what they should do (See Figure 4.3).

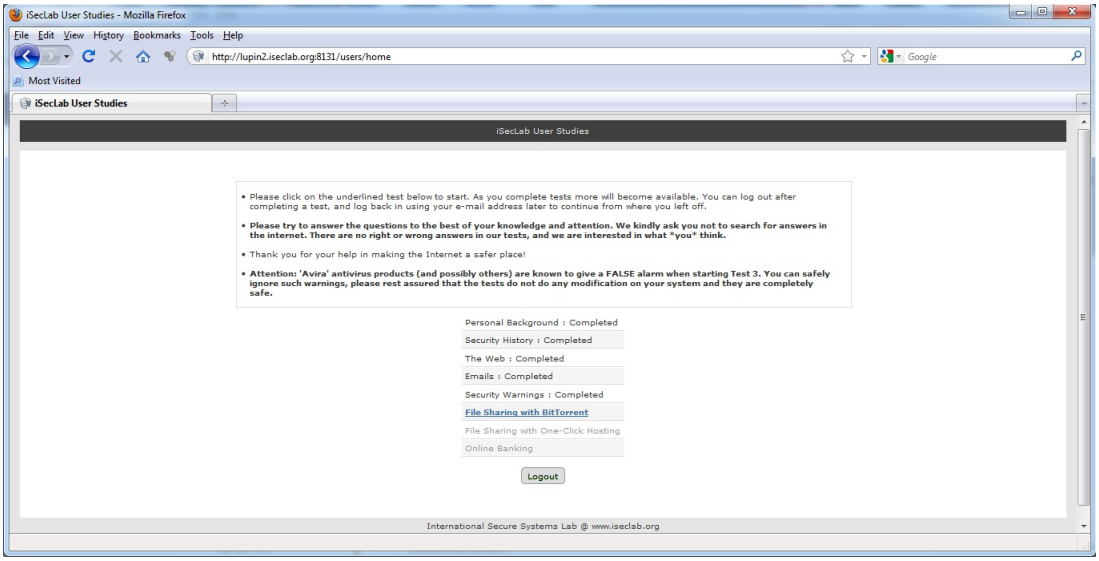

Figure 4.1: List of test suites

In order to mitigate the negative effects of motivational confounds and prevent inaccurate results due to loss of concentration, we gave participants the option to leave after completing any number of tests, and come back again later to continue from where they left off. Apart from their email addresses (which we used to uniquely identify participants in case they wanted to continue the tests later), we did not ask for any other personally identifiable information.

We recruited the participants through announcements on Twitter and Facebook, and by directly asking people in other disciplines (e.g., medicine, geology) to promote our test platform URL. After eliminating the data from 2 respondents due to their poor command of English, we completed our empirical study with 164 participants. However, the file sharing test suite was available only to the participants that reported previous experience with BitTorrent or with one-click-hosting services (91 and 97 participants, respectively). In addition, online banking test suite was optional and was completed by 140 participants. In the following sections, we describe in more detail the security tests we conducted.

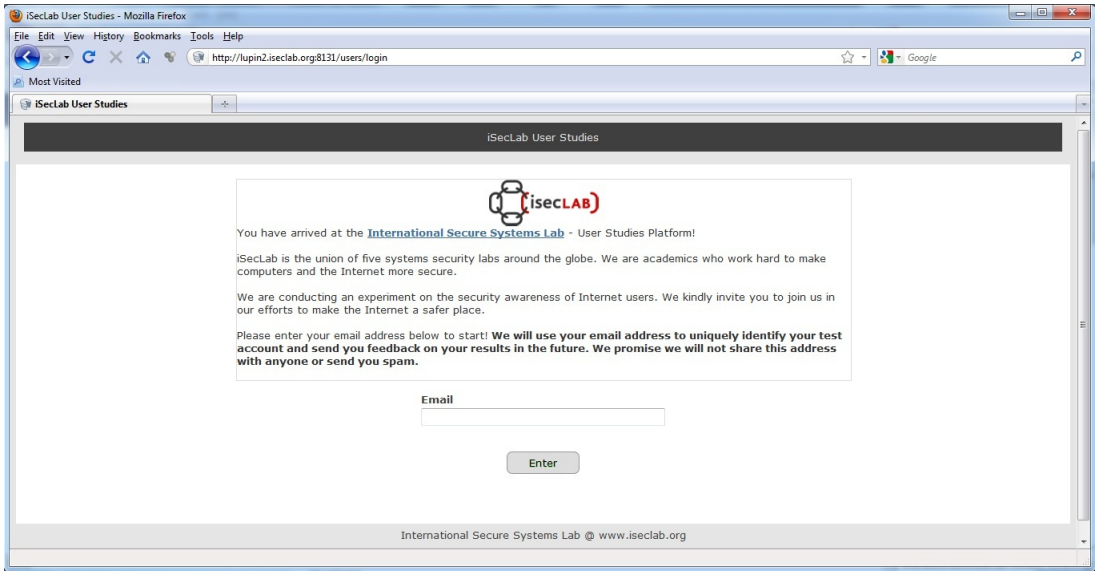

Figure 4.2: Home page of our online test platform

4.1 Demographic Information

Before starting the tests, we collected standard demographic information. In addition, since we were interested in observing the effects of technical background on the results of the experiments, we asked the participants several questions to estimate their security proficiency. More specifically, we asked them if they are comfortable with doing everyday tasks using their computers, if they have previous programming experience, if they are professionally involved in software/hardware development, if they have a degree in computer science or a related technical field,

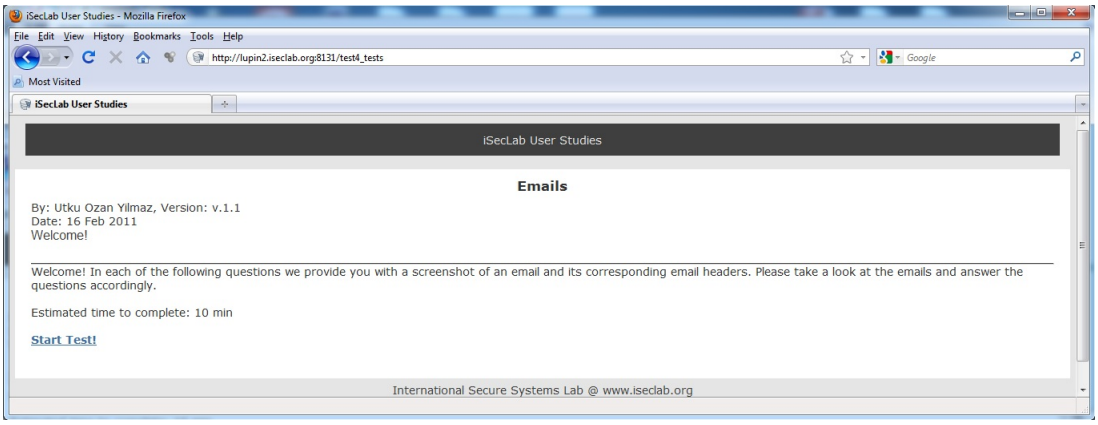

Figure 4.3: Short briefing screen at the beginning of a test

Figure 4.4: Facebook wall post with a reflected cross-site scripting attack

and if they have specialized security expertise.

4.2 Web-Based Attacks

The first test suite presented the participants with various URLs, and asked them to rate the "risk they perceived" for each link. That is, the participants were asked to rate how dangerous or safe they believed it would be to click on the links. The risk perception ratings were given in the 5-point Likert scale [43], ranging from "*Definitely safe*" to "*I cannot decide*", to "*Definitely dangerous*". After assessing the risk for the link, the participants were also asked if they would click on the link and were prompted to briefly explain the rationale behind their decision.

The test suite included both malicious and benign URLs. Malicious URLs exemplified common attacks such as cross-site scripting, session fixation and link manipulation tricks (see Figure 4.4 for an example). The rest of the tests included perfectly benign but somewhat unconventional links with, for instance, a very long parameter list, a non-HTTP protocol, and mixed-case characters.

Each URL and its related questions were presented on a separate page. The URLs were created using actual HTML anchor tags to allow the browsers to render them authentically, and to allow the participants to hover their mouse over the links to see the hyperlink destination in the browser's status bar.

For two URLs (one given as a raw IP address and the other one as a TinyURL [14] link), the participants were also given a third option: to verify

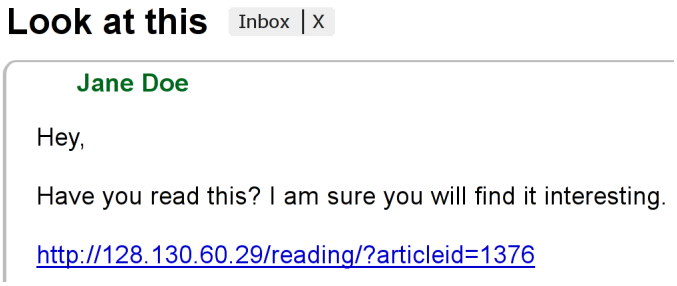

Figure 4.5: Spam email with a generic text and a suspicious link

the destination and then decide if they would follow the link. The participants who chose this option were asked to briefly explain how they would determine the destination.

Clicking on links was disabled for all tests by including an appropriate Javascript code. Note that there were two exceptions where we included a screenshot instead of embedding the link in the HTML page. These were attacks that required a context surrounding the link (e.g., the Facebook wall post shown in Figure 4.4).

4.3 Email-Based Attacks

In the email-based attacks test suite, we showed the participants screenshots of emails together with full header information. The suite included a PayPal phishing email, a spam email suggesting to click on an IP link, an ordinary eBay newsletter, a fake prize-giveaway notification, a malicious email with an executable attachment, an advance-fee fraud with the classic Nigerian connection text (e.g., [56]), an innocuous Amazon advertisement, and a phishing email crafted to look like it is sent from a bank (see Figure 4.5 for an example). Similar to the web-based attacks test suite, we asked the participants to rate the risk they perceived on a 5-point Likert scale. We also asked the participants whether they would react to the email by, for example, downloading the attachment, and asked them to state the reasons behind their decisions.

Again, each email and the related questions were shown on a separate page.

Prior to every question, the participants were informed that they should ignore the email headers, and only focus on the email content, if they did not know how to interpret the header information.

4.4 File Sharing-Related Attacks

In the file sharing test suite, our aim was to confront the participants with typical (but potentially risky) file sharing scenarios (e.g., the download of an executable file disguised as a movie). Clearly, making the participants go through a number of file sharing scenarios if they did not possess prior experience in the area would not have produced useful results. Therefore, prior to the file sharing tests, we asked the participants whether they know what BitTorrent or One-click-hosting are, and whether they had experience with the websites we used in our tests. The participants who responded negatively were directly taken to the next suite, skipping the rest of the tests.

We split the file sharing test suite in two parts: BitTorrent-related tests and One-click-hosting-related tests. In the BitTorrent tests, we walked the participants through a scenario in which they were trying to download their favorite movie. We showed the participants a set of screenshots of torrent search results and torrent detail listings, as presented by three popular BitTorrent hosters/meta-search engines: The Pirate Bay [12], isoHunt [7] and Torrentz [15]. Each of these included cues to the legitimacy (or the lack thereof) of the movie file such as file extensions, file sizes, torrent contents, number of people sharing the file, reputation of the uploader, torrent descriptions, and warnings in user comments (see Figure 4.6 for an example).

We then asked the participants to rate the risk they perceived for each search result we presented them, and decide whether they would proceed to download the file. Furthermore, we asked the participants which cues they used when they made a decision. While the screenshots showing the details of a single torrent were presented on a separate page, the search results were all given in the same

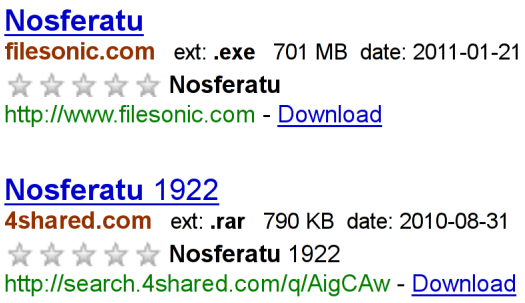

Figure 4.6: Search results for a movie in Filestube with a bad file extension and a suspicious size.

page in order for the participants to be able to compare the search hits to each other.

Once the questions were answered, we took the participants to fully interactive torrent download pages carefully reproduced from the ones of The Pirate Bay and isoHunt. We then asked them to click on the correct download button among the various advertisement banners disguising themselves as legitimate download links. Note that we did not introduce any artificial web banners for the purpose of our study. Instead, we faithfully copied the original content from the corresponding websites for more realistic observations.

We designed the One-click-hosting tests in a similar fashion. However, this time, we showed screenshots and reproduced pages from the popular websites Filestube [4], iFolder [6], Megaupload [10] and Megavideo [11].

4.5 General Security Warnings

In the general security warnings test suite, we showed the participants screenshots of different types of security warning windows: an invalid SSL certificate warning when trying to open Facebook, a blacklisted website warning in Firefox, a virus alert by an antivirus scanner, and a scareware virus alert web pop-up.

Since there is no definite answer to the legitimacy of some of these warnings, we only asked the participants whether they believe the warning is critical, and

to explain the rationale behind their decision. As always, each screenshot was presented on a separate page.

4.6 Online Banking

In the final test suite, we asked participants to transfer money to given account numbers using a simple website we built for this purpose (see Figure 4.7). After completing the transactions, we asked them to answer some questions about them. Moreover, we tried two-factor authentication with 50 of the participants by sending them one-time codes by SMS to authenticate their transactions (see Figure 4.8).

The task in this test suit was to perform two transactions, the first of which was a regular transaction. However in the second transaction, we altered the destination account number, simulating the behavior of some man-in-the-browser trojans [21]. This alteration could be seen by the participants in the transaction summary screen and in the SMSes they received.

4.7 Processing of the Results

We used R Programming Language [13] to make statistical computations and to produce the necessary graphs. However, we first needed to process the raw data we gathered from the users through some intermediate steps to make them into proper R input.

We started by producing keys for test suites. These keys were mainly used for quantifying the replies to questions with definite answers, which were in turn used to calculate security and risk perception scores. On the other hand, questions without definite answers (demographic questions, open-ended questions, etc.) were simply copied to output files. We used a Python script to produce output files by interpreting input and key files.

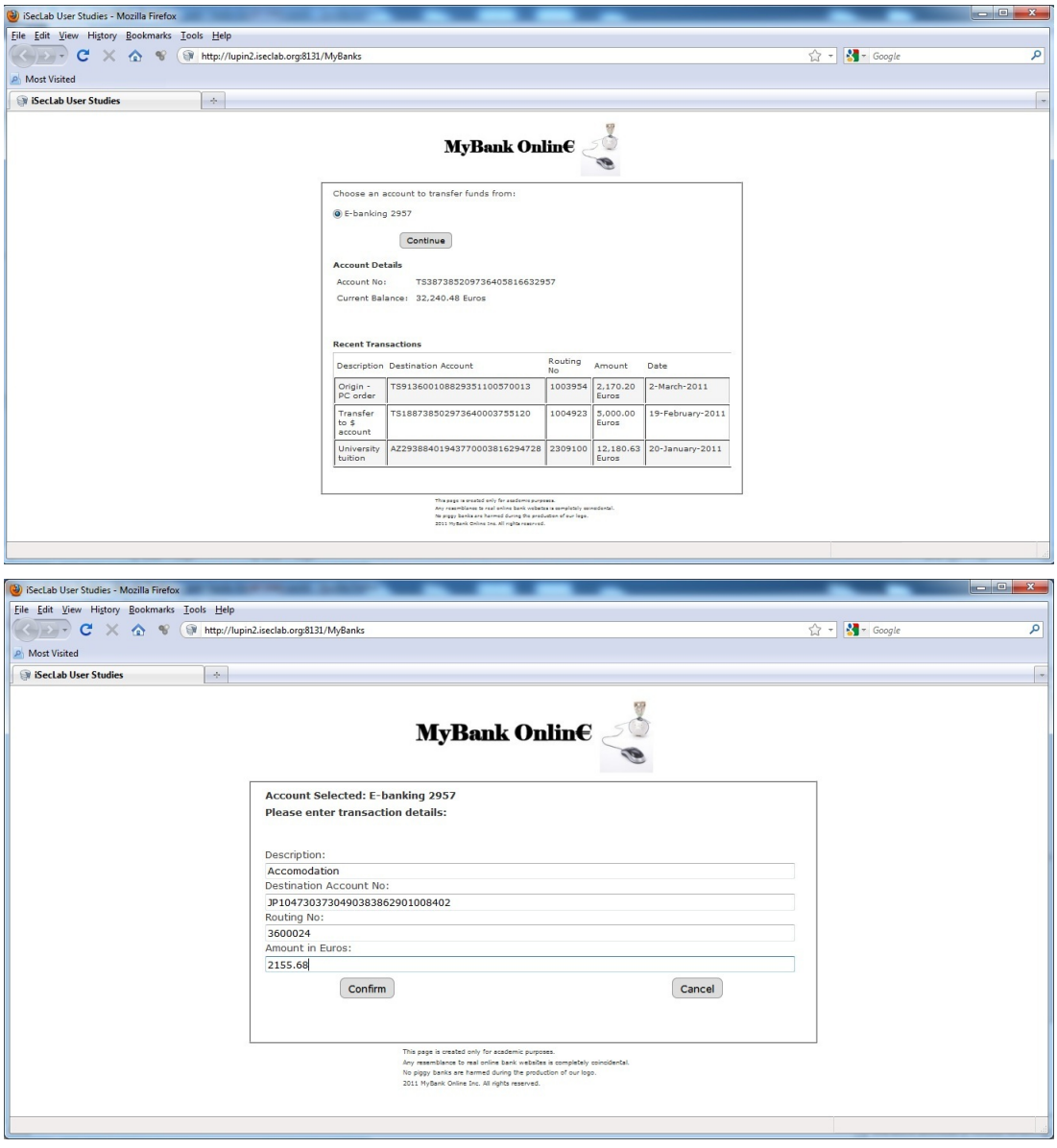

Figure 4.7: Online banking website

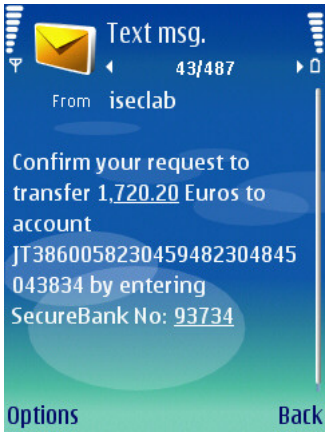

Figure 4.8: SMS including a one-time code to authenticate an online banking transaction

After that, we manually categorized the answers to open-ended questions and tagged them. We determined the following categories:

- *technical*: Means that the participant used a technical point of view in his answer (e.g., he investigated the header of an email).
- *wrong technical*: Means the participant approached the situation from a mistaken technical point (e.g., he doesn't realize the actual destination of a link is different from the one in its anchor text and technically examines the anchor text). We should emphasize that wrong refers to the method rather than the result. Taking the right action for the wrong technical reasons still falls under this category.
- *familiarity*: Signifies that the participant relies on his past experience while answering.
- *content/context/source*: Denotes that the participant depends on the content of the item in question (e.g., he says he never reads emails in "need your help" format), the context in which it appears (e.g., he says he doesn't follow links in emails), or its source (e.g., he says he would base his judgement on the reliability of the website which provides the link) while giving his answer.

Name of the test	Goal	Type of data
Kruskal-Wallis one way ANOVA	Compare three or more unmatched groups	Rank, score or measurement from non-Gaussian population
Spearman's rank correlation	Quantify association between two variables	Rank, score or measurement from non-Gaussian population
Pearson's product-moment correlation	Quantify association between two variables	Measurement from Gaussian population
Fisher's exact test	Compare two unpaired groups	Binomial (two possible outcomes)
Wilcoxon rank-sum test (a.k.a Mann-Whitney U test)	Compare two unpaired groups	Rank, score or measurement from non-Gaussian population
Pearson's chi-squared test	Compare three or more unmatched groups	Binomial (two possible outcomes)

Table 4.1: Applicability criteria for each of the statistical tests we used

- *intuition/guess/clueless*: Means the participant has no idea about the situation, or he feels like it is dangerous/safe but cannot show a reason as to why.

Then, we used a small Java program that produces csv files from these intermediate files. While doing so, it also removes answers to the open-ended questions, leaving only their tags. Finally, these csv files were used by R to produce statistical results. Table 4.1 shows the applicability criteria for each of the statistical tests we used [5].

Chapter 5

Analysis of the Test Results

In this chapter, we explain the strategy we used to evaluate the collected data, and we present the results obtained through the experiments. In Chapter 6, we interpret these results, and summarize the insights we distilled.

5.1 Demographics and Diversity

The test participants were 31.1% female and 68.9% male, their ages ranged from 19 to 69 (mean=26.52, s.d.=8.76, variance=76.79), and their nationalities spanned 17 different countries.

11.6% of the participants held a doctoral degree, 10.4% a master's degree, 45.8% a bachelor's degree, and 1.2% an associate's degree. 72.6% of our participants were continuing students of which 5.5% were pursuing a doctoral degree, 37.2% a master's degree, and 29.9% a bachelor's degree. 33.5% of them were professionally employed.

The majority of the participants reported being familiar and comfortable with common tasks and concepts such as surfing the web and using email services, social network applications, entertainment activities like watching movies and playing games, performing e-commerce and e-banking operations, installing and using

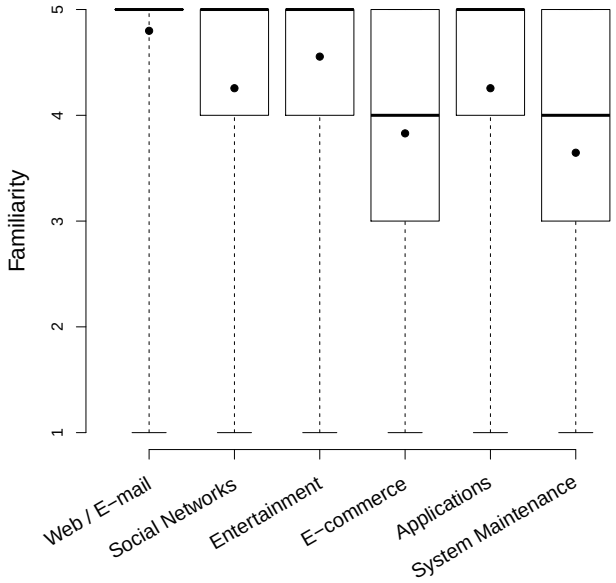

Figure 5.1: Five-number summaries of participant familiarity with typical computer tasks and concepts. Mean values are also displayed by the black dots.

applications, and doing basic system maintenance by configuring their operating system and recovering from simple errors (see Figure 5.1 for the corresponding five-number summaries).

65.6% of the test participants used Windows as their primary operating system, 17.7% used Linux, 15.6% used Mac OS X, and one participant used a BSD-variant. 41.5% of them preferred Firefox, 34.8% preferred Chrome, 14.6% preferred Internet Explorer, 6.1% preferred Safari, and the remaining 3.0% used various other web browsers.

Based on the participants' responses to the questions about their security background, we divided them into three expertise groups:

- *Non-techies* use computers at home or work on a regular basis. They are comfortable using basic applications to perform everyday tasks. Their professions are in non-technical fields, and they have little, or no programming experience. 42.7% of the test participants fall into this group.

		Security Scores					
		Min	1st Quartile	Median	Mean	3rd Quartile	Max
Web Tests	Non-techies	22.2	44.4	55.6	56.7	66.7	100.0
	Techies	44.4	55.6	66.7	68.0	77.8	100.0
	Experts	33.3	66.7	77.8	74.2	88.9	100.0
Email Tests	Non-techies	50.0	75.0	75.0	77.5	87.5	100.0
	Techies	62.5	75.0	87.5	85.2	100.0	100.0
	Experts	50.0	75.0	87.5	84.7	100.0	100.0
BitTorrent Tests	Non-techies	36.4	54.5	54.5	58.7	63.6	90.9
	Techies	36.4	45.4	63.6	63.6	75.0	90.9
	Experts	36.4	54.5	63.6	63.6	72.7	81.8
One-click Hosting Tests	Non-techies	20.0	60.0	60.0	64.8	80.0	100.0
	Techies	40.0	65.0	80.0	79.1	100.0	100.0
	Experts	40.0	60.0	80.0	76.1	100.0	100.0

		Risk Perception Scores					
		Min.	1st Quartile	Median	Mean	3rd Quartile	Max
Web Tests	Non-techies	46.7	60.0	63.3	64.1	68.9	82.2
	Techies	53.3	64.4	71.1	71.1	75.6	91.1
	Experts	53.3	68.9	75.6	74.3	79.4	91.1
Email Tests	Non-techies	40.0	65.0	72.5	71.7	80.0	97.5
	Techies	55.0	75.0	80.0	80.8	87.5	97.5
	Experts	60.0	75.6	81.2	81.4	90.0	92.5
BitTorrent Tests	Non-techies	49.1	56.4	60.9	60.9	63.6	76.4
	Techies	56.4	61.8	66.4	65.8	67.7	78.2
	Experts	54.5	61.8	67.3	66.4	70.9	80.0
One-click Hosting Tests	Non-techies	48.0	60.0	68.0	67.6	72.0	100.0
	Techies	48.0	69.0	76.0	74.5	84.0	92.0
	Experts	56.0	68.0	72.0	73.1	80.0	92.0

Table 5.1: Five-number summaries for each test suite and participant group. Mean scores are also given for comparison to the median.

- *Techies* are either computer scientists, or otherwise involved in a closely-related field of study or profession (such as engineering disciplines dealing with technology). These participants are knowledgeable on the intricacies of computer systems. However, their technical training does not focus on computer security. Techies constitute 19.5% of the participants.

- *Experts* are computer security professionals. In their studies or professions, they specialize in securing computer systems. They claim to have a deep understanding of security fundamentals, and have some practical experience in the field. 37.8% of the participants are experts.

5.2 Security and Risk Perception

In order to quantify the performance of the participants in the security tests, we computed two global scores: a *security score* and a *risk perception score*.

The *security score* is a measure of how good a participant is at averting attacks, while also refraining from erroneously discrediting non-threats as being dangerous. We compute the security score as the total number of questions answered correctly in this manner. We then normalize it to account for the participants who skipped any of the file sharing tests, and scale it to a value between 0 and 100.

The *risk perception score* is a measure of a participant's ability to recognize the severity and consequences of each situation. We compute it based on the 5-point Likert [43] scale, with questions that ask the participants how dangerous they think each scenario is. For an obvious threat, the participants who respond with 'definitely dangerous' receive 5 points, while those who answer 'definitely safe' only receive 1 point. For benign items, we reverse the scores accordingly. We then normalize and scale the score in the same way we calculate the security score.

Considering all the participants in all tests, the security scores ranged from 46.43 to 96.97 (mean=70.21, s.d.=10.84, variance=117.53, 1^{st} quartile=63.64, median=69.70, 3^{rd} quartile=78.57) and risk perception scores ranged from 48.18 to 90.59 (mean=70.48, s.d.=7.35, variance=54.09, 1^{st} quartile=66.29, median=70.45, 3^{rd} quartile=75.19). The distributions of scores for the three groups are summarized in the violin plots in Figure 5.2. Note that, since scores show what percentage of the questions were answered correctly, values close to the middle of the scale (i.e., 50) could possibly indicate no security awareness but merely random guesses.

Out of all the questions the participants answered incorrectly, 56.1% were benign samples misjudged as being malicious. 43.9% were attacks confused as being benign. This slight imbalance could be explained by the observation that

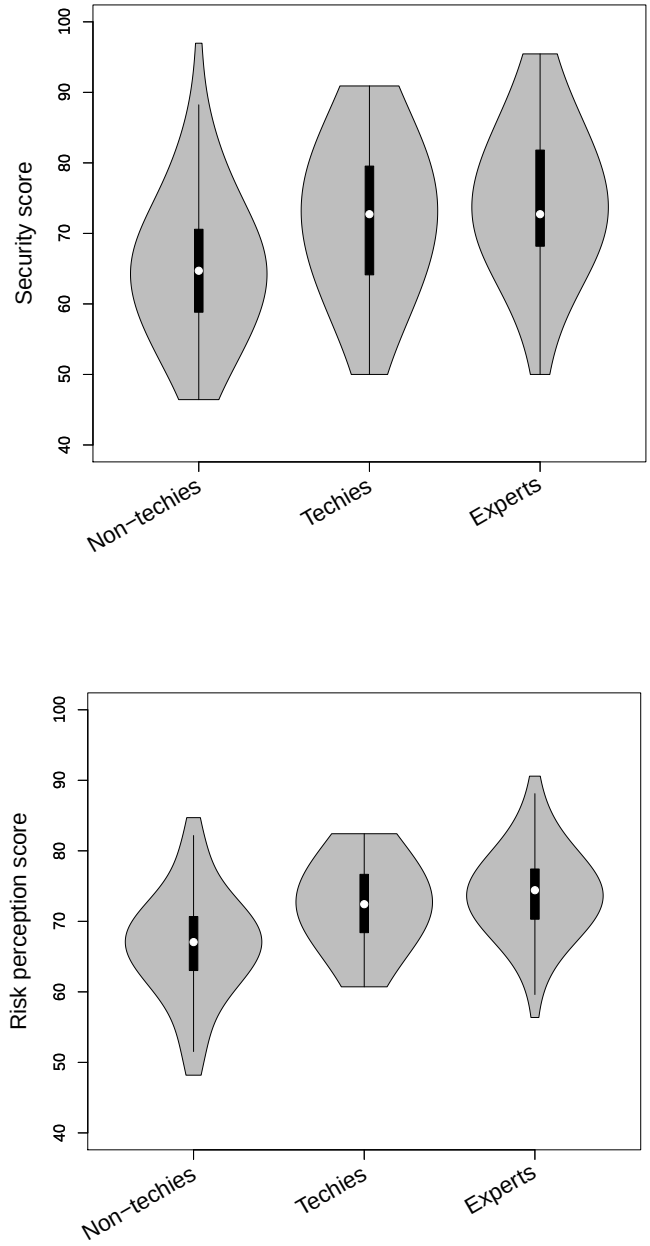

Figure 5.2: Five-number summaries and probability densities of total scores

participants were probably over-careful because they were expecting to encounter attacks in the tests.

The Kruskal-Wallis one-way analysis of variance tests [41] showed that both the security and risk perception scores differed significantly among the three participant groups (i.e., $H = 26.89$, $df = 2$, $p = 1.45 \times 10^{-6}$ for security scores and $H = 37.36$, $df = 2$, $p = 7.71 \times 10^{-9}$ for risk perception scores). Following these with multiple comparison post-hoc tests revealed that non-techies and experts differed significantly in both scores, while non-techies and techies, or techies and experts did not. This means that both security and risk perception considerably increased with security expertise.

We also analyzed the scores for each test suite separately (see Table 5.1), and checked the scores once again for potential differences among participant groups. Similarly, the risk perception scores differed significantly between non-techies and experts in every test suite. However, we observed that the security scores only differed significantly for the web-based attacks suite between non-techies and experts (i.e., $H = 25.42$, $df = 2$, $p = 3.01 \times 10^{-6}$). We did not observe a statistically significant difference in security scores for the email and the two file sharing test suites (i.e., $p > 0.54$, $p > 0.23$ and $p > 0.67$, respectively).

When we investigated the relationship between risk perception and security scores in each participant group (see Figure 5.3), an analysis with Spearman's rank correlation [52] revealed that the two types of scores are positively correlated for each group. In other words, for higher risk perception scores, the security scores show an increasing trend as well. However, this correlation is considerably weaker for non-techies (i.e., $\rho = 0.50$, $p = 9.99 \times 10^{-6}$) compared to techies and experts combined (i.e., $\rho = 0.70$, $p = 6.51 \times 10^{-15}$).

Moreover, we looked into the relationship between the familiarity with security terms and security and risk perception scores. In order to achieve this, we calculated a security terms familiarity score for each participant by asking them to rank their familiarity with the following security terms in a scale of 1 to 5 (5: very familiar, 1: not familiar at all): malware, virus, trojan, worm, spyware,

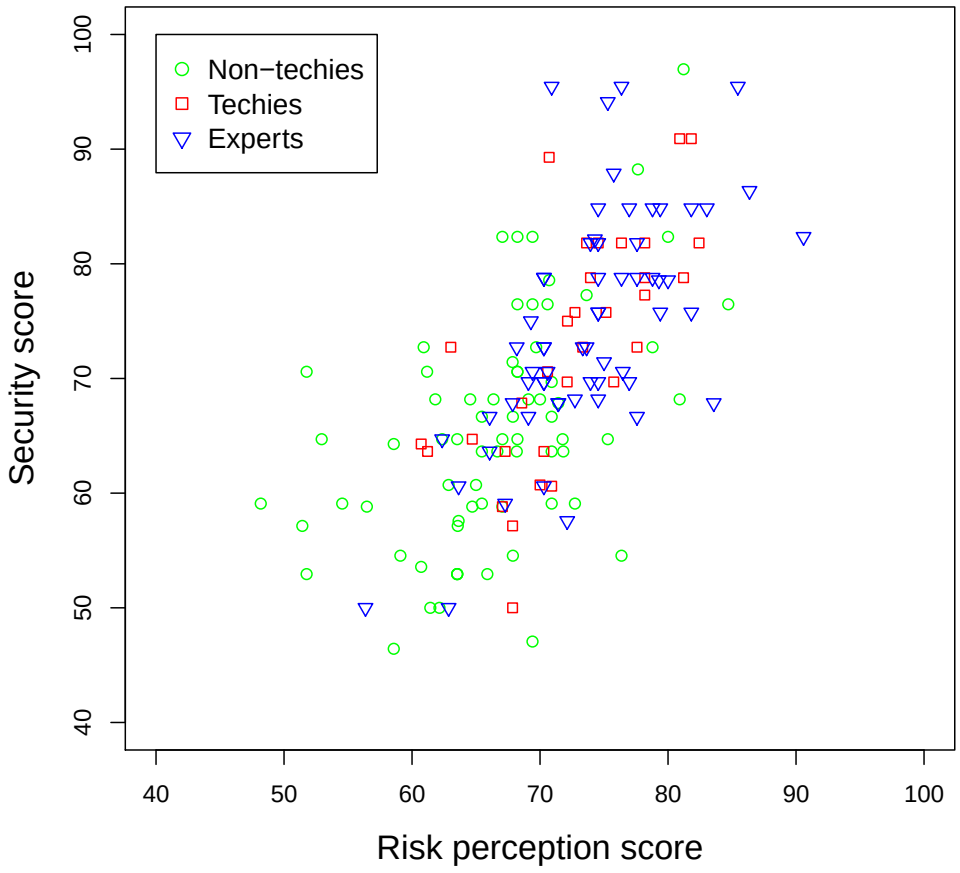

Figure 5.3: Relationship between risk perception and security scores

adware, scareware, botnet, spam, phishing, drive-by-download, SSL/TLS, certificate, https and social engineering. Then, we summed the answers for each participant and scaled the result to make it into a score out of 100. After analyzing the aforementioned relationship using Spearman's rank correlation, we found a correlation between the familiarity score and security and risk perception scores ($\rho = 0.38$, $p = 4.44 \times 10^{-7}$ for security scores and $\rho = 0.70$, $p < 2.2 \times 10^{-16}$ for risk perception scores). This means people who are more familiar with terms related to the internet attacks are more successful at both recognizing and averting them.

Finally, in our study group, we did not see a statistically significant correlation between the age and education level of the participants and either of the scores ($p > 0.50$, $p > 0.23$ for security scores and $p > 0.72$, $p > 0.52$ for risk perception

		Blindly Follow	Ignore	Technically Verify	Depends on Source	Not Familiar
TinyURL	Non-tech	17.1 %	74.3 %	0.0 %	8.6 %	35.7 %
	Tech	21.9 %	31.3 %	18.7 %	28.1 %	15.6 %
	Expert	16.1 %	32.3 %	32.3 %	19.3 %	4.8 %
Raw IP	Non-tech	15.7 %	80.0 %	0.0 %	4.3 %	28.6 %
	Tech	18.8 %	43.7 %	25.0 %	12.5 %	6.3 %
	Expert	17.7 %	38.7 %	33.9 %	9.7 %	3.2 %

Table 5.2: Results showing participants' decisions when confronted with a shortened URL and a raw IP link. The "Not familiar" column is not mutually exclusive with the others, i.e., some participants stated they do not know what a shortened URL is but decided to follow it anyway.

scores, respectively). The score medians differed significantly with sex, where females scored significantly lower compared to males. However, this was largely due to the females being concentrated in the non-techies. When testing for each group separately there was no significant difference in score medians within groups (i.e., $p > 0.43$, $p > 0.70$, $p > 0.49$ for security scores and $p > 0.38$, $p > 0.49$, $p > 0.32$ for risk perception scores, for each group, respectively).

5.3 Test-Specific Results

The test suites contained questions that cannot be effectively quantified with the previous scoring approach. Hence, we used case-specific evaluation strategies for these questions, as described in the following subsections.

5.3.1 Web Tests Including IP Addresses and Shortened URLs

In the web-based attacks suite, the legitimacy of the IP and shortened URL links (e.g., TinyURL) cannot be determined just by looking at the URL. Hence, we did not compute scores for them. Instead, we investigated how many participants were able to successfully verify the link destinations. For example, the participants could have fetched HTML headers, performed WHOIS and reverse DNS look-ups, or utilized URL expansion tools. A summary of these results is given

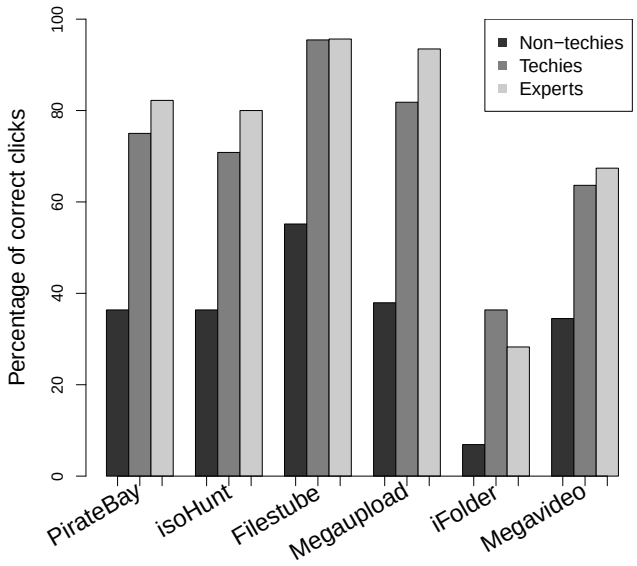

Figure 5.4: Correct clicks for the trick banner tests on pages reproduced from popular file sharing websites

in Table 5.2.

The most striking observation in the experiments was that none of the participants in the non-techie group said that they would attempt to verify the destination of either the IP, or the TinyURL link. Moreover, compared to the other groups, a considerably higher number of non-techies did not know what an IP address, or URL shortening was.

Note that many participants directly related to their previous surfing experience and were confused by similar links they had used in the past. In their answers, these participants demonstrated a completely wrong technical understanding of the use of IP addresses, or URL shortening services. For example, some answers we received indicated an IP to be an interface for a printer/router configuration screen, an index of photographs, and "*proxy code*". The TinyURL link was thought to be a YouTube video, a photo sharing service, a blog, and an MP3 download website.

5.3.2 Trick Banner Tests

For the simulated websites in the file sharing test suite, we counted the number of clicks on the correct download links, as well as on deceptive banners crafted with the aim of luring Internet users to click on them. The percentage of clicks on the correct links for all participant groups are shown in Figure 5.4. The high error rate in the iFolder test was in part because we reproduced the page in its original language (Russian). The aim was to observe the participants' navigation behavior when the website is in an unfamiliar language.

These results showed that while most experts and techies were able to recognize and avoid false banners, the majority of non-techies failed. That is, such participants did not realize that they were not clicking on the actual link (and were being tricked into clicking on potentially dangerous banners) even though they reported being familiar with the test website.

5.3.3 General Security Warning Tests

For the general security warning suite, we collected and analyzed the qualitative feedback we received from the participants after displaying them a number of specific warning screens, and asking them how they would react to them.

The first security warning in this suite, an SSL certificate error raised when visiting the Facebook homepage, generated mixed opinions, balanced by opposing views among the different participants groups. 46 participants told us that they would ignore the warning since they regularly encounter similar situations every day. In contrast, 37 participants said that they would trust their browser's warning, and immediately leave the website. 22 participants thought that it would be safe to ignore the warning since Facebook is very well-known while 17 participants believed that its popularity would make Facebook an obvious target for attacks, making it difficult to ignore such a warning. 11 participants said that they would have ignored the warning if they had visited the website in question before, and had not seen a warning message. In comparison, 10 participants responded in the

exact opposite way; they stated that they would know an attack is taking place if the website did not raise a warning in previous visits. Interestingly, among the techie and expert participants, only 9 participants could correctly recognize the possible threat scenario (such as a DNS-based attack). The rest provided wrong reasons for the displayed warning. For example, 7 participants thought that the fault would most likely be in their machines.

In the next part of the test suite, where we displayed a screenshot of a black-listed website warning in Firefox, the majority of the participants (i.e., 121) agreed that taking the warning seriously would be the best course of action and only 2 participants claimed that they had encountered false alarms in the past. From the participants' comments we can conclude that the text of the warning (i.e., `Reported Attack Page!`) was very important in recognizing and evaluating the danger. For example, many reported that it was frightening, and some non-techie participants explicitly stated that "*The red color is scary*" or that "*This error seems more serious [than the certificate error].*".

We obtained similar results for the malware scanner screenshot that warns the participants of a possible infection. 122 participants said they would comply with their antivirus software's recommendation, and delete the infected files. The common rationale is perfectly summarized by one of the comments: "*Isn't this why I pay for the antivirus software? If I don't trust it, why do I pay for it?*". However, 20 participants said they would ignore the warning if they recognized the reported files as being "benign".

The final warning screen simulated a scareware web pop-up (i.e., a browser pop-up window that imitates an alert from a virus scanner tool) with an exaggerated list of reported infected files. This test tricked 81 participants into believing that it was an authentic malware alarm. A participant who claimed to be a security expert stated his concern: "*With so many files infected, I would re-install the system*". Non-techies in particular were confused by the number of reported infections. One of them stated that he would definitely click on the web pop-up, and added: "*Easy decision with that many infections!*". Only 23 participants (of which only one was a non-techie) demonstrated familiarity with scareware, while

	Without second channel		With second channel	
	Noticed attack	Didn't notice attack	Noticed attack	Didn't notice attack
Non-techies	6.5 %	93.5 %	0 %	100 %
Techies	5.6 %	94.4 %	45.5 %	54.5 %
Experts	7.7 %	92.3 %	36 %	64 %

Table 5.3: Results showing percentage of participants who realized and who didn't realize the man-in-the-middle attack, based on their expertise groups and the utilization of second authentication channel.

58 participants in all groups averted the threat either because they thought web pop-up warnings were not convincing, or because they knew websites cannot scan their computers without their authorization.

5.3.4 Online Banking Tests

For the online banking suite, we investigated how many participants noticed the change in the account number during the second transaction. In order to improve participants' chance of noticing the change, we presented the destination account number they should use as an image file instead of a text. This forced them to manually enter the account number, which makes it easier to recognize the alteration afterwards. A summary of the results is given in Table 5.3.

An interesting observation was that none of the non-techies who used the second authentication channel (i.e., SMS) realized an attack was going on, even though some who didn't use the second channel realized the situation. On the other hand, utilization of the second authentication channel proved to be useful on techies and experts. An analysis with Fisher's exact test [30] demonstrates that the usage of the SMS codes helped the participants to perform significantly better (i.e., $p = 9.04 \times 10^{-4}$).

We also investigated the relationship between successfully realizing the attack and the frequency of online banking (How often do you use online banking?) and the frequency of transferring money using online banking (How often do you make money transfers using online banking?). A Wilcoxon rank-sum test (a.k.a. Mann-Whitney U test) [45] shows that while the relationship between

realizing the attack and frequently performing online banking operations isn't strong enough to be considered significant, frequently transferring money using online banking significantly improves the performance of the participants ($p = 0.6$ and $p = 0.3$, respectively).

Lastly, we looked into the relationship between a user's expertise group and his success in the online banking test. Pearson's Chi-squared test [46] showed that the success rate differed significantly among three groups (i.e., $X - squared = 7.41$, $df = 2$, $p = 0.02$). Following this with Fisher's exact tests between pairs of groups revealed that non-techies and experts differed significantly, while non-techies and techies, or techies and experts did not ($p = 0.011$, $p = 0.053$ and $p = 1$, respectively). This means that success rate of noticing the attack increased with security expertise.

Chapter 6

Discussion and Insights Gained

In this chapter, we provide detailed interpretations of the test results presented in Chapter 5, and list the insights we distilled from them.

6.1 Security Training and Risk Perception

As we have shown in Section 5.2, when the overall scores are considered, techies and experts performed significantly better than non-techie participants. In the individual test suites, as one would intuitively expect, the techies and experts also received higher scores. However, we only observed a statistically significant difference between these groups in the web security test suite. In other words, we did not see statistical proof that, when generalized to the whole population, the security experts would perform better in email and file sharing security scenarios compared to technically unsophisticated users. For example, our results confirm that even an average Internet user has sufficient experience to detect and mitigate common attacks that they regularly encounter (e.g., an email promoting the Nigerian scam).

Note that, the fact that users are able to detect an attack does not necessarily mean they also understand its intricacies, or its consequences. In fact, for

the email test suite, when we categorized the decision making strategies of the participants by looking at their explanations, only 2.9% of the responses given by non-techies provided meaningful technical insights, while the remaining 97.1% were based purely on intuition and past experience. In contrast, 23.4% of the techies and 30.6% of the experts directly looked for technical cues while detecting an attack (e.g., by investigating the email headers).

Our results empirically demonstrate that after continuous exposure to spam emails and common scams such as fake prize-giveaways, most of the non-techie users learn to instinctively avoid these attacks. This observation is in line with psychology literature that shows individuals fall back on their intuition when faced with complex information that they cannot process. That is, the guesses based on intuition could be correct since they draw from vast previous experience [27, 36].

The attacks we presented to the participants in the web security test suite (e.g., link manipulation and code injection tricks) were more sophisticated compared to classic email scam scenarios. Hence, detecting these attacks required specialized technical knowledge. Although the participants who were security experts always scored higher than non-techies, the web attack test suite was where their security training gave them a substantial advantage over technically unsophisticated users. The web security scores show that the majority of technically unsophisticated users are not equipped to deal with more elaborate attack scenarios. In fact, intuition and "folk wisdom" without any technical basis sometimes even misled non-techie participants: for example, exactly half of these users believed that a partially upper-case URL like `www.googLE.com` was a clever attempt at phishing.

The results we presented in Section 5.2 indicate that the difference in risk perception scores between non-technical participants and experts is statistically significant. Note that this observation holds even in the tests where the security scores do not significantly differ among these two groups. That is, non-techies and experts have different perceptions of the risk in a given situation (i.e., the risk perception scores differ significantly). Nevertheless, these groups reach similar

conclusions, and act in a similar manner (i.e., the security scores do not differ significantly).

Although non-technical participants can sometimes detect attacks, this does not imply that they realize the severity of a threat. In the tests where the security scores of the non-techies are closer to those of the experts, these technically unsophisticated participants do not really "perceive" the danger of the situation. Rather, they depend on their intuition and past experience. In most of the cases, the non-technical participants judged a malicious email (e.g., a Paypal phishing scam) as being "Definitely safe", or being "Most Probably Safe", stating that they cannot read email headers, and that they do not see anything wrong with the content. However, they chose to ignore the mail instead of clicking on the given link. One participant explained: "Looks good. But I don't trust it, I don't know why". Not being able to articulate the reasons behind a correct decision is a known indication of guesses based on intuition [27, 36].

Note that this observation is also supported by the relatively weaker correlation between the risk perception scores and the security scores of non-techies compared to those of techies and experts. Although such a correlation in no way implies a causality relationship between the two scores, it shows that for experts, a higher risk perception is associated with higher security, but much less so for non-technical participants. That is, non-techies base their decisions on uninformed guesses. When their intuition is "trained" by exposure to a certain type of attack (e.g., regular email scams), they do manage to mitigate the attack. However, when the attack is something they have not regularly seen (e.g., a session fixation attack), they fail.

The weak correlation between the participants' perception and their actions sometimes manifests itself in the exact opposite way as well. For example, one of the participants flagged a link to a blog with a script injection attack as being "Most probably dangerous", but decided to click on it anyway. The reasoning was: "*It is a blog. I would probably click anyway.*". While the participant understood that there was something wrong with this URL, she could not understand and assess the severity of the threat, and was not aware of the possible consequences

of her actions.

Our experiments confirm the general intuition that security training is the most effective method for reducing the risk posed by Internet threats. The more users are familiar with a threat, the higher their security awareness becomes. This is further supported by the correlation between familiarity of participants with the 15 internet attack-related terms we presented them, and their security and risk perception scores. Although it is unrealistic to expect an average Internet user to have the motivation to learn and understand security concepts, based on the feedback we received about our online security test platform, we believe that security games and online test platforms that are tailored towards non-technical people in order to familiarize them with common attack patterns (e.g., such as PhishGuru and Anti-Phishing Phil [42, 49] for anti-phishing training) are required. Such testing systems could help achieve a similar effect to the one we observed in the email-based attack tests. In fact, general psychology literature also supports the idea that intuition could be "taught" by repeated experience, and also by virtual simulations [36, 48].

6.2 Size Matters

When the participants did not have the necessary technical knowledge to make an informed decision for a test and had to rely on their intuition, a very common trend was to make a guess based on the "size", the "length", or the "complexity" of artifacts.

In the web-based attack tests, for example, a benign Amazon link was labeled as malicious by non-technical participants based on the fact that the URL contained a crowded parameter string. Some of the comments included: "*Too long and complicated.*", "*It consists of many numbers.*", "*It has lots of funny letters.*" and "*It has a very long name and also has some unknown code in it.*". Many of these participants later said they would follow a malicious Paypal phishing URL because "*It is simple.*", "*Easy to read.*", "*Clear obvious link.*" and it has a "*Short*

address". One participant further argued that "*This is not dangerous, address is clear. [Amazon link] was dangerous because it was not like this.*".

Interestingly, in some cases, the non-technical participants managed to avert attacks thanks to this strategy. For example, a number of participants concluded that a Facebook post containing a code injection attack was dangerous solely on the grounds that the link was "long" and "confusing".

Analogously, in the file sharing tests, the responses based on intuition mainly relied on arguments about the file size. For example, the participants who did not understand how BitTorrent works judged torrents merely on their expectations of a full-length movie's size. These participants often made misinformed decisions such as discrediting a 700MB RAR archive as being malicious as the size of the movie had not decreased after the compression (note that movie files are already heavily compressed), or marking a 790KB file as correct since it referred to a very old movie from 1922. In another example, several participants thought that a 70MB file could be a full 90 minute high-quality movie.

Again, these results underline the importance of a proper security training. In fact, when users are not able to make an informed decision about a possible threat, they fall back on judging the situation based on often misleading characteristics, such as an item's size and complexity.

6.3 Loud and Clear Warnings

The responses from participants in the general security warnings test suite demonstrate that warnings and alerts which are stated in simple, yet very direct terms are considerably more successful in conveying the associated risk. Also, it is useful to include intimidating visual cues into the alerts. For example, Firefox's blacklisted website warning, which clearly states in a frame with a red background (which again resides on a black background) that the page in question is an "attack page", was very effective.

Surprisingly, we also observed similar effects in the BitTorrent file sharing tests. That is, many participants believed that a perfectly legitimate search result we presented them was malicious content because it contained a "skull icon" next to the torrent name. Ironically, this icon indicates that the file was uploaded by a trusted Pirate Bay user. Similarly, in the isoHunt tests, several participants were disturbed by the "red font" used when listing some of the trackers, which was completely benign.

In line with the experiments of Sunshine et al. [53], the invalid SSL certificate warning we showed our participants was not very persuasive with its ordinary color theme and vague statement: "This connection is untrusted". Clearly, warnings need to be loud and clear in communicating the risk of a wrong decision by the user. In the case of Firefox, even though the warning attempts to inform users that their connection to the website could be tampered with, non-technical users who are completely oblivious to the consequences of a man-in-the-middle are not disheartened by this explanation.

Our findings support the idea that security warnings should be designed with an alarming look-and-feel. They need to express the impending threat clearly, and avoid a formal description of the issue. For example, instead of saying: "This website's identity cannot be verified", we believe that a message such as "This website is probably fake and it can steal any information you enter!" would be more effective.

6.4 URL Shortening Services and Tools

Our tests indicate that none of the non-technical participants attempted to verify the destination of a shortened URL (in our case, a TinyURL). As explained in Section 5.3.1, the majority of the non-techie group was not aware of the fact that a shortened URL could link to any destination on the web. Rather, they thought that TinyURL was the website that actually hosted the content. Even those participants who were aware of the risks stated that they did not know how

to verify the destinations of these links.

A wide variety and number of URL shortening services are available on the Internet today. Their frequent use in social networks such as Twitter make them ubiquitous. Unfortunately, the prevalence of shortened URLs also make them an effective way to distribute malware and lure users into scams. A recent study by Grier et al. [33] states that over 2 million links posted on Twitter point to attack pages and that through the use of nested URL shortening, blacklisting solutions can be circumvented. Our results support the fact that non-technical users are easily fooled by shortened URLs in practice.

There exists several online services (e.g., [3, 9]), and extensions for popular browsers (e.g., [2, 16]) that offer shortened URL expansion capabilities. While these tools would definitely help technically inclined people assess the risk before following a shortened URL, our experiments show that they are ineffective for non-technical users who do not have a firm grasp of the technology behind URL shortening.

Analogous to the recent integration of website blacklists and phishing detection heuristics into popular browsers (e.g., the anti-phishing features in IE as of Version 7), we believe that URL expansion and threat detection capabilities (e.g., [8]) need to be integrated into browsers as soon as possible.

6.5 Trick Banners

In the interactive tests featuring reproduced download websites, the false click rates for non-techies were considerably higher compared to experts. In 5 of the 6 tests, more than half of the non-technical participants clicked on a banner instead of the real link to download the item in question. That is, even if these participants were able to differentiate between a legitimate and a malicious search result displayed by the file sharing website, they still would not have managed to complete the download successfully.

Using deceptive banners to trick Internet users into visiting a website is a well-known advertisement strategy [54]. However, there have also been recent attacks on the advertisement networks of popular websites where attackers have legitimately bought banner space [55, 57, 60] or exploited bugs in ad servers (e.g., such as a recent attack on The Pirate Bay [58]). In such attacks, the attackers typically use banners to serve malware. Additionally, some malware have utilized trick banners for committing fraud [59]. Our study empirically confirms that trick banners are very effective (attack) techniques in influencing the click behavior of non-technical users. From a user's point of view, a possible defense technique in dealing with such tricks would be utilizing ad-blockers (e.g., [1]). Hence, it is important to inform and train users about the use of such tools, especially when visiting certain classes of websites.

6.6 Second Authentication Channels

We have shown in Section 5.3.4 that experts performed significantly better than non-techies in the online banking test suite. Moreover, even though usage of a second authentication channel significantly improves the chances of realizing an attack, none of the non-technical participants who used a second authentication channel detected our simulation of a man-in-the-middle attack. Combined with the fact that 6.5% of the non-technical participants who didn't use a second authentication channel noticed an attack was taking place, we can argue that second authentication channels may even enhance the chance of success of an attack by inducing a false sense of security in the technically unsophisticated users, let alone prevent attacks. Further studies on this subject are needed to prove or disprove this point.

This is also in line with our observations in Section 6.4. Just as online services and browser extensions that offer shortened URL expansion capabilities help technically inclined people while they are ineffective for non-technical users, second authentication channels help technically competent people while failing to help those without the necessary technical knowledge. This once again proves the

importance of a proper security training in mitigating attacks on the internet. As long as users stay security-unaware, it is impossible for security tools to achieve their full potentials.

Chapter 7

Conclusion

In this thesis study, we describe an experiment platform for observing the behavior of users when they are confronted with typical benign and malicious interaction scenarios on the Internet. We present the results of a study we conducted on 164 Internet users who possessed diverse backgrounds and varying degrees of computer security knowledge. Our results empirically confirm the general intuition that security training has a considerable positive impact on a user's ability to make correct security decisions and assess risks. This observation especially holds when the threats involve technically complex attacks. The thesis also shows that for relatively simpler and common threats (e.g., well-known e-mail scams), non-technical users can exhibit comparable performance to knowledgeable users by solely depending on their intuition and past experience (i.e., there is a training effect).

We observed that many users consider unusual "size" and "length" characteristics of URLs and downloaded files as indicators of risk. These users are often highly susceptible to attack strategies that exploit shortened URLs, raw IP addresses, and trick banners. We also observed that clear and intimidating warning screens are more effective in conveying risk as opposed to vague technical messages.

Recently, security tools that aim to assist users in revealing the real destinations of shortened URLs have been introduced, such as Longshore [8]. Our findings suggest that such security services are largely ineffective for non-technical users since they are not able to use them, or understand the concepts behind URL shortening services. Likewise, we observed that second authentication channels used during online transactions fail to enhance the security of non-technical users.

Although it is clear that the security problems on the Internet do not only have a technical aspect (i.e., there exists the human factor), existing literature is sparse, and there have not been many studies by the security community that attempt to shed light on how Internet users are able to cope with current attack vectors. We hope that this work will pave the way for similar works in the future.

Bibliography

[1] Adblock Plus. http://adblockplus.org/en/, 2011.

[2] ChromeMUSE - Multi-URL Shortener/Expander. https://chrome.google.com/extensions/, 2011.

[3] Clybs - Url expander. http://www.clybs.com/urlexpander, 2011.

[4] Filestube. http://www.filestube.com/, 2011.

[5] How to choose a statistical test. http://www.graphpad.com/www/book/choose.htm, 2011.

[6] iFolder. http://www.ifolder.com/, 2011.

[7] isoHunt. http://isohunt.com/, 2011.

[8] Long-Shore. http://long-shore.com/, 2011.

[9] LongURL. http://longurl.org/, 2011.

[10] Megaupload. http://www.megaupload.com/, 2011.

[11] Megavideo. http://www.megavideo.com/, 2011.

[12] The Pirate Bay. http://thepiratebay.org/, 2011.

[13] The R Project for Statistical Computing. http://www.r-project.org/, 2011.

[14] TinyURL. http://www.tinyurl.com/, 2011.

[15] Torrentz. `http://torrentz.eu/`, 2011.

[16] Xpnd.it! short URL expander. `http://addons.mozilla.org/en-us/firefox/addon/xpndit-short-url-expander/`, 2011.

[17] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster. Building a dynamic reputation system for dns. In *Proceedings of the 19th USENIX conference on Security*, USENIX Security'10, pages 18–18, Berkeley, CA, USA, 2010. USENIX Association.

[18] M. Bailey, J. Oberheide, J. Andersen, Z. M. Mao, F. Jahanian, and J. Nazario. Automated classification and analysis of internet malware. In *Proceedings of the 10th international conference on Recent advances in intrusion detection*, RAID'07, pages 178–197, Berlin, Heidelberg, 2007. Springer-Verlag.

[19] A. Barth, C. Jackson, and J. C. Mitchell. Robust defenses for cross-site request forgery. In *Proceedings of the 15th ACM conference on Computer and communications security*, CCS '08, pages 75–88, New York, NY, USA, 2008. ACM.

[20] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. All your contacts are belong to us: automated identity theft attacks on social networks. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 551–560, New York, NY, USA, 2009. ACM.

[21] CAcert. Concepts against Man-in-the-Browser Attacks. `http://www.cacert.at/svn/sourcerer/CAcert/SecureClient.pdf`, January 2007.

[22] S. Chiasson, P. C. van Oorschot, and R. Biddle. A usability study and critique of two password managers. In *Proceedings of the 15th conference on USENIX Security Symposium - Volume 15*, Berkeley, CA, USA, 2006. USENIX Association.

[23] J. Clark, P. C. van Oorschot, and C. Adams. Usability of anonymous web browsing: an examination of tor interfaces and deployability. In *Proceedings of the 3rd symposium on Usable privacy and security*, SOUPS '07, pages 41–51, New York, NY, USA, 2007. ACM.

[24] CNN. Amazon EC2 outage downs Reddit, Quora. `http://money.cnn.com/2011/04/21/technology/amazon_server_outage/index.htm`, 2011.

[25] G. Conti and E. Sobiesk. Malicious interface design: exploiting the user. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 271–280, New York, NY, USA, 2010. ACM.

[26] R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, CHI '06, pages 581–590, New York, NY, USA, 2006. ACM.

[27] A. Dijksterhuis, M. W. Bos, L. F. Nordgren, and R. B. van Baaren. On making the right choice: The deliberation-without-attention effect. *Science*, 311:1005–1007, February 2006.

[28] E. Edelson. The 419 scam: information warfare on the spam front and a proposal for local filtering. *Computers & Security*, 22(5):392–401, 2003.

[29] D. Endler. The Evolution of Cross Site Scripting Attacks. Technical report, iDEFENSE Labs, 2002.

[30] R. Fisher. On the interpretation of χ 2 from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society*, 85(1):87–94, 1922.

[31] B. Friedman, D. Hurley, D. C. Howe, E. Felten, and H. Nissenbaum. Users' conceptions of web security: a comparative study. In *CHI '02 extended abstracts on Human factors in computing systems*, CHI EA '02, pages 746–747, New York, NY, USA, 2002. ACM.

[32] B. Friedman, D. Hurley, D. C. Howe, H. Nissenbaum, and E. Felten. Users' conceptions of risks and harms on the web: a comparative study. In *CHI '02 extended abstracts on Human factors in computing systems*, CHI EA '02, pages 614–615, New York, NY, USA, 2002. ACM.

[33] C. Grier, K. Thomas, V. Paxson, and M. Zhang. @spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security*, CCS '10, pages 27–37, New York, NY, USA, 2010. ACM.

[34] G. Gu, R. Perdisci, J. Zhang, and W. Lee. Botminer: clustering analysis of network traffic for protocol- and structure-independent botnet detection. In *Proceedings of the 17th conference on Security symposium*, pages 139–154, Berkeley, CA, USA, 2008. USENIX Association.

[35] G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee. Bothunter: detecting malware infection through ids-driven dialog correlation. In *Proceedings of 16th USENIX Security Symposium on USENIX Security Symposium*, pages 12:1–12:16, Berkeley, CA, USA, 2007. USENIX Association.

[36] R. M. Hogarth. *Educating Intuition*. Univ. of Chicago Press, 2001.

[37] S. Institute. Top Cyber Security Risks, September 2009. `http://www.sans.org/top-cyber-security-risks/summary.php`.

[38] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Commun. ACM*, 50:94–100, October 2007.

[39] E. Kirda, C. Kruegel, G. Banks, G. Vigna, and R. A. Kemmerer. Behavior-based spyware detection. In *Proceedings of the 15th conference on USENIX Security Symposium - Volume 15*, Berkeley, CA, USA, 2006. USENIX Association.

[40] M. Kolšek. Session fixation vulnerability in web-based applications. *Acros Security*, page 7, 2002.

[41] W. Kruskal and W. A. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, pages 583–621, 1952.

[42] P. Kumaraguru, S. Sheng, A. Acquisti, L. F. Cranor, and J. Hong. Teaching johnny not to fall for phish. *ACM Trans. Internet Technol.*, 10:7:1–7:31, June 2010.

[43] R. Likert. A Technique for the Measurement of Attitudes. In *Archives of Psychology*, volume 140, 1932.

[44] MacArthur-Foundation. New Study Shows Time Spent Online Important for Teen Development. `http://www.macfound.org/site/c.lkLXJ8MQKrH/`

b.4773437/k.3CE6/New_Study_Shows_Time_Spent_Online_Important_for_Teen_Development.htm, 2008.

[45] H. Mann and D. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 18(1):50–60, 1947.

[46] K. Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302):157–175, 1900.

[47] P. Ratanaworabhan, B. Livshits, and B. Zorn. Nozzle: a defense against heap-spraying code injection attacks. In *Proceedings of the 18th conference on USENIX security symposium*, SSYM'09, pages 169–186, Berkeley, CA, USA, 2009. USENIX Association.

[48] M. E. P. Seligman and M. Kahana. Unpacking intuition: a conjecture. *Perspectives on Psychological Science*, 4:399–402, July 2009.

[49] S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L. F. Cranor, J. Hong, and E. Nunge. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 3rd symposium on Usable privacy and security*, SOUPS '07, pages 88–99, New York, NY, USA, 2007. ACM.

[50] R. Smith, M. Holmes, P. Kaufmann, and A. I. of Criminology. *Nigerian advance fee fraud*. Australian Institute of Criminology, 1999.

[51] SonicWALL. Two-Factor Authentication. www.sonicwall.com/downloads/SSL_VPN-Two_factor_Authentication.pdf, 2011.

[52] C. Spearman. The proof and measurement of association between two things. *American Journal of Psychology*, 15:88–103, 1904.

[53] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying wolf: an empirical study of ssl warning effectiveness. In *Proceedings of the*

18th conference on USENIX security symposium, SSYM'09, pages 399–416, Berkeley, CA, USA, 2009. USENIX Association.

[54] M. Terms. Trick Banner. `http://www.marketingterms.com/dictionary/trick_banner/`.

[55] ThreatPost. Major Ad Networks Found Serving Malicious Ads. `https://threatpost.com/en_us/blogs/major-ad-networks-found-serving-malicious-ads%-121210`, December 2010.

[56] C. Tive. *419 Scam: Exploits of the Nigerian Con Man*. iUniverse.com, 2006.

[57] TorrentFreak. Yahoo! pimping malware from banner ads. `http://www.theregister.co.uk/2008/04/28/yahoo_serves_rogue_ads/`, April 2008.

[58] TorrentFreak. Hackers Target and Exploit Pirate Bay Ad Server. `http://torrentfreak.com/hackers-target-and-exploit-pirate-bay-ad%-server-100913/`, September 2010.

[59] Trusteer. Zeus Adds Investment Fraud to its Bag of Tricks. `http://www.trusteer.com/blog/zeus-adds-investment-fraud-its-bag-tricks/`, April 2011.

[60] Wired. Rogue Anti-Virus Slimeballs Hide Malware in Ads. `http://www.wired.com/epicenter/2007/11/doubleclick-red/`, November 2007.

[61] Y. Xie and A. Aiken. Static detection of security vulnerabilities in scripting languages. In *Proceedings of the 15th conference on USENIX Security Symposium - Volume 15*, Berkeley, CA, USA, 2006. USENIX Association.

[62] H. Yin, D. Song, M. Egele, C. Kruegel, and E. Kirda. Panorama: capturing system-wide information flow for malware detection and analysis. In *Proceedings of the 14th ACM conference on Computer and communications security*, CCS '07, pages 116–127, New York, NY, USA, 2007. ACM.