"Not able to resist the urge" : Social Insider Attacks on Facebook

by

Wali Ahmed Usmani

Bachelors of Science, Computer Science, Lahore University of Management Sciences, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Computer Science)

The University of British Columbia
(Vancouver)

December 2016

Abstract

Facebook accounts are secured against unauthorized access through passwords, and through device-level security. Those defenses, however, may not be sufficient to prevent *social insider attacks*, where attackers know their victims, and gain access to their accounts using the victim's device. To characterize these attacks, we ran two Amazon Mechanical Turk studies geographically restricting participant pool to US only. Our major goal was to establish social insider attack prevalence and characteristics to justify a call to action for better protective and preventative countermeasures against it.

In the first study involving 1308 participants, we used the list experiment, a quantitative method to estimate that 24% of participants had perpetrated social insider attacks, and that 21% had been victims to it (and knew about it).

In the second, qualitative study with 45 participants, we collected stories detailing personal experiences with such attacks. Using thematic analysis, we typified attacks around 5 motivations (fun, curiosity, jealousy, animosity and utility), and explored dimensions associated with each type. Our combined findings indicate a number of trends in social insider attacks. We found that they are common, they can be perpetrated by almost all social relations and often have serious emotional consequences. Effective mitigation would require a variety of approaches as well as better user awareness.

Based on the results of our experiments, we propose methodological steps to study the perception of severity of social insider attacks. In this procedure, we include an experimental design of the study and its possible limitations. The study consists of presenting stories collected in the previously mentioned second study to a new cohort of participants. It the asks them to provide a Likert Scale rating

and justification for how severe they perceive the attack in the story to be if they were the victim as well as how likely they feel they might be a victim to such an attack.

Lastly, we discuss possible future work in creating countermeasures to social insider attacks, their viability and limitations. We conclude that no single technique is complete solution. Instead mitigation will require a number of techniques in combination to be effective.

Preface

Research for social insider attacks was done as part of the ThirdEye Project, and funded by the Office of the Privacy Commissioner Canada. I would like to thank them for their support. This thesis is derived from the publication “Characterizing Social Insider Attacks on Facebook” set to appear in Proceedings of SIG CHI 2017 in Denver, Colorado and was a combined result of the following authors that I would like to acknowledge:

- Diogo Marques, LaSIGE, Faculdade de Ciłncias, Universidade de Lisboa (University of Lisbon)
- Ivan Beschastnikh, NSS, The University of British Columbia
- Tiago Guerreiro, LaSIGE, Faculdade de Ciłncias, Universidade de Lisboa (University of Lisbon)
- Konstantin Beznosov, LERSSE, The University of British Columbia
- Luis Carrio, LaSIGE, Faculdade de Ciłncias, Universidade de Lisboa (University of Lisbon)

This text expands upon the work covered in the previously mentioned publication.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I dedicate this thesis to:

My parents, who loved and supported me through this long journey especially at times where I was less than an ideal son, especially when I would be too distracted to pick up calls. Even though they were not here to witness all the highs and lows of my time in Canada, never in all my time did I feel that they were not there for me, no matter what time it was or how busy they were. Mom, thank you for convincing me to take this journey on. It made a better person out of me. Dad, thank you for your moral support and assurance. I certainly needed it.

Dr Ivan Beschastnikh, my friend, guide and mentor without whom this would not have been possible. I came to Canada with high hopes and amazing dreams and he helped me realize each and every one of them. I would certainly have been lost without his guidance. At no point during my time here did I feel like I didnt have someone to advice from, from topics of computer science to life.

Dr Konstantin Beznosov, an inspiration who pushed me on through the trenches. I got to know him during my first term's first course. I was intimated by how difficult Grad School can be but Kosta was a consistent force and taught me to project my ideas confidently and clearly; lessons I do not plan to forget.

Lastly, I would like to thank Dr Fareed Zaffar, without whom I would not be in the computer science field. After a tough 4th semester when all my colleagues were just starting to get into research, few believed in me. Dr Zaffar was one of them and as a result, I got to dabble in computer science research for the first time. It was the beginning of a long and fruitful journey. He helped me hone my skills, focus on my targets and gave me a platform to prove myself. I don't think I can ever repay that belief.

I would also like to acknowledge my second reader, Dr Karon Maclean whose feedback was immensely helpful in improving upon this work and helping me get over the finish line.

Chapter 1

Introduction

Over the last decade, Facebook has become the most popular social networking service with over 1.6 billion users worldwide [35]. Users often share and maintain personal and potentially sensitive information on their accounts, including messages, pictures and videos [20], which could negatively impact them if an unauthorized party gained access to it. As long as this information could be of potential value to an *adversary*, they may try to obtain it without the owner's consent. Adversaries who are *insiders* have a social relationship with the account owner and are of special concern. The proximity between the victim and an insider makes it easier for the insider to obtain unauthorized access to the victim's device and Facebook account.

Insider attacks can be both physical and digital and in terms of computer security, can be difficult to address due to privileged position of the adversary. In digital insider attacks, the adversary has specific insider knowledge of the victim's security measures and can leverage it to cause harm without physical access to the target system. One example can be that of a disgruntled employee remotely attacking an employer's system either through access as an employee or through insider knowledge of the system's weaknesses. Social insider attacks on the other hand are when the adversary uses the victim's device to perform the attack. It is also refereed to as a 'lunch-time' attack [11] synonymous with conducting the attack on a co-worker's computer during lunch time in an office setting.

Insider attacks have been combated on an enterprise level in the past by fol-

lowing security principles of least-privileged access control, device specific information access,location tagging and non-repudiable event logging [17]. However, little work has been done on such attacks in the online social network (OSN) context. OSNs commonly use 'something you know' or 'something you have' driven mechanisms such as passwords and cookies with authentication tokens to federate access control. To increase usability and not force users to log in each time they want to access their profile, most services allow authentication credential caching. While such security mechanisms may be effective to combat general adversaries, user devices are particularly vulnerable to insiders who can gain physical access to them. Such adversaries need neither training nor any special skills to gain access to the victim's device and due to their insider knowledge are in position to cause severe harm to the victim.

In this work, we focus on *Facebook social insider attacks*, when an insider adversary gains physical accesses to the Facebook account of a victim using Facebook's end-user interfaces, like the web or a mobile application, on the victim's device without the victim's permission. We consider a victim's device to be one that is regularly controlled by the victim. This includes personal devices such as smartphones, but also, for instance, work computers, or shared devices in a household.

Although often overlooked, social insiders attacks can have adverse effects. For instance, posting potentially embarrassing material using the victim's account (an act sometimes referred to as 'facejacking' or 'frape [8, 9]) is often dismissed as a prank. However, acts functionally similar to such pranks, have been regarded as defacement worthy of criminal prosecution [12].

Aside from anecdotal evidence, little is known about the nature and prevalence of social insider attacks on Facebook accounts. The lack of structured knowledge about the issue hinders the capacity to address it. For instance, how much effort should be expended on educating people on how to protect themselves, if social insider attacks are very rare, or of little consequence? And if they are not rare, how could we design defenses that are effective against the spectrum of attacks that might exist, if this spectrum is not properly understood?

This thesis helps to bridge our gaps in knowledge of this attack by quantitatively and qualitatively characterizing social insider attacks against Facebook ac-

counts. Specifically, this work examines whether a call to action for mitigation techniques against such attacks is justifiable. We do not, however, have predetermined benchmarks or thresholds from which we can make a unequivocal conclusions. While it is not possible to make a direct comparison, prevalence of other computer security threats to users such as the emails scams such as the 419 Advance Fee scams (commonly known as the "Nigerian Prince Scam") can be used a rough proxy. According to a report by EKOS Research Associates for the Government of Canada [1], 7% of online Canadians have replied to spoof of phishing emails.

In our first study, we estimated the prevalence of attacks with a survey conducted on Amazon's Mechanical Turk service (MTurk). Since direct questions about attacks are sensitive, we opted for the *list experiment* format [3, 30]. In list experiments, participants are presented with a list of statements and asked to indicate how many, instead of which ones, they agree with. Estimates of behaviors can be obtained by comparing average responses between lists with varying items. We ran this study with 1,308 US adult participants who reported being Facebook users, and found that social insider attacks are prevalent. We estimated that 24% of participants carried out Facebook social insider attacks, and that 21% were knowing victims. We describe our experimental procedure in detail to demonstrate the necessity of the list experiment as well as how we identified and addressed possible sources of error.

In our second study, we used a qualitative approach to understand what social insider attacks look like in more detail. We asked MTurk workers to write free-form and anonymous stories about past experiences with social insider attacks, and used thematic analysis to extract salient dimensions. We report on several themes that emerged from our analysis, including the relationships between perpetrators and victims, attack vectors, the role of premeditation, and others. We further organize social insider attacks on Facebook accounts by the types of motivation, and discuss how attacks tend to unfold.

In order to understand how social insider attacks are perceived by people, we outline an experimental procedure which incorporates aspects from previous usable security research as well as findings from our previously mentioned studies. Primarily quantitative in nature, the procedure presents new participants with sto-

ries collected in the second, previously mentioned study. It then asks them to rate severity according to their own perception of risk if they were the victim of the attack as well as how likely they think they could possibly be a victim of a social insider attack such as the one described in the story.

Our findings suggest that social insider attacks are common enough to argue for better countermeasures. Furthermore, they are mostly opportunistic and have a range of motives, including fun, curiosity, jealousy, animosity, and utility each of which have distinct attack patterns. We conclude that mitigating such attacks will require a multi-pronged approach. Promising avenues of future research could be education of users about the threat of social insiders, investigation of better deterrence of perpetrators, and improving technology support for detection and investigation of attacks by the victims themselves.

Chapter 2

Related Work

Information theft and unauthorized access is not a rare phenomenon. A 2013 Pew survey found that 21% of internet users have had an email or social networking account compromised or taken over by someone else without permission, and 86% had taken steps to protect themselves or mask their digital footprint [28]. The study also showed that people were concerned about data leakage, with 51% being very concerned for their data to only be accessible to them and those they authorize.

However, social insider attacks have seldom been a target of research. In contrast, attacks by outsiders, even targeted remote attacks, are much more well understood. For instance, the main characteristics of manual "hijacking" on Google accounts have been studied [5] with the explicit exclusion of attacks in which the attacker knows the victim personally. In those instances of outsider attacks, the motivation, and the way attacks unfold, follow a pattern of exploitation for financial gain, which is not comparable to insider attacks on OSN accounts. The experiences of victims of remote hijacking was studied in a 2014 survey of 89 people who had experienced compromise of a personal email or social networking account [32]. Although this study did not exclude insider perpetrators, only 5 participants were at least moderately confident that the compromise was caused by someone they knew. Nevertheless, the survey indicates that even if consequences for victims are not harmful in practice (e.g. spam to contact list), the negative feelings associated with being a victim are striking. Participants expressed anger, fear, embarrassment, and a sense they had been violated. In our research, which focus on physical

attacks, rather than on remote attacks, we found corroboration for the emotional consequence of being victim of an attack, to an even higher degree.

Previous research on social insider attacks has focused on identifying internal threats within an enterprise using machine learning techniques to differentiate authorized users and possibly malicious insiders [14]. Deployed in a single department within an organization, the system recorded a trace of employee activity, such as logging on and off, sending emails and access to restricted files. The US Government's Computer Emergency Readiness Team (CERT) was asked to artificially inject 'insider threat' data. Specifically, they injected artificially generated traces of malicious insiders attempting to extract sensitive enterprise data. For a more rigorous comparison, benign artificial data was also inserted to avoid the system's accuracy by differentiating between artificially created traces and recorded ones. The traces were then analyzed by an anomaly detection engine to identify possibly malicious behavior. The study reported an receiver operating characteristic (ROC) of 70%. Similar work is called Beneficial Intelligent Software for Identifying Anomalous Human Behavior (BENWARE) [26], focused on detecting human insider behavior in a closed organization's IT department. Benware used Support Vector Data Description (SVDD), a technique similar to Support Vector Machines (SVM) and trained on computer usage patterns such as log-on times, files accessed and web requests made to model standard behavior and then picks out anomalies. Benware took approximately 3 days to detect insiders.

Analysis of behavioral patterns in the context of OSNs has focused on detecting bots, or autonomous programs infiltrating the social network for various purposes from phishing to collection of private information. A variety of techniques have been used to perform such an analysis. With a focus in spam bots, one applied technique is Principal Component Analysis (PCA) [39] to model 'normal' behavior on Facebook based on account usage patterns and attempt to single out profiles that significantly deviate from this behavior. Techniques used in to detect anomalous behavior on OSN accounts show promise in being applied to detect social insider attacks which also leave an anomalous behavioral trace behind.

To our knowledge, of the several possible types of social insider attacks on Facebook, only "fraping" – impersonating a user, for comical (or humiliating) effect – has been studied in some detail. In a 2016 interview study with 46 OSN

users, fraping appeared to be mostly restricted to younger generations, to be seen as a practical joke, and even to have some positive effects, as a factor of in-group bonding [24]. However, fraping may sometimes be interpreted as a form of cyberbullying, and may resort to what amounts to online hate speech, for instance presenting homosexuality in a negative way [16]. As in the case of younger people using the word "drama" to refer to some online interactions which adults would classify bullying, using the word "frape" may allow ambivalence between serious and frivolous attacks, as a way to avoid framing incidents as instances of victimization [22].

Research on privacy perceptions of Facebook users suggest there is particular concern with insiders having access to information they are not supposed to have. In a 2012 study, among 260 participants, 86% were not concerned with the threat of strangers on Facebook [19]. In that instance, however, strangers were other Facebook users who were not directly in the user's social circle, viewing their content primarily due to Facebook's privacy settings. A sizable proportion of participants, 37%, showed concern with some people in their circles viewing their profile or social content, which, at least at the time, was problematic, since Facebook's privacy settings were mostly geared towards preventing strangers from having access to information. While our work also aims to understand activities of insiders, we focus on social insiders rather than digital insiders.

Unlike social insider attacks on Facebook, knowledge has been building on social insider attacks on smartphones. A 2013 investigation on concerns over social insider attacks on smartphones found that users are aware of the insider threat [25]. The study included a survey of 724 participants, of which 12% reported having had experiences of unauthorized data access, and 9% reported having had engaged in social insider attacks on a device belonging to someone else. However, since that study relied on self-reports, and the questions posed to participants were sensitive, those statistics are likely to underestimate the problem, due to social desirability bias [37]. A recent list experiment study of snooping attacks (the subset of smartphone social insider attacks in which the objective is limited to snooping), found much higher rates of prevalence, with an estimated 30% of participants having been perpetrators in a 1-year period [21]. Our research differs by focusing on Facebook instead of sensitive data on smartphones and by including all platforms

by which a social insider can conduct an attack, such as desktop computers, laptops and tablets.

Chapter 3

Social Insider Attack Prevalence Study

3.1 Background

A security threat is of general interest if it is both probable to materialize, and harmful when it does. With this first study, we wanted to understand how common (or uncommon) social insider attacks on Facebook are right now, as a proxy to the probability with which these attacks might occur in the future, all else remaining equal.

There are, however, challenges in obtaining such measurements. Asking users if they are victims or perpetrators of social insider attacks is likely to lead to underestimation. Victims may be unaware of intrusions, when attacks are unnoticed; or they may be unwilling to report them. Perpetrators may not want to self-incriminate, or may be led to give socially-desirable answers just by the use of language associated with privacy and security [4]. In this context, socially-desirable answers would aim to reduce the severity of the perpetrators actions, down play consequences or lie to cover up involvement in a social insider attack.

To minimize the social desirability bias, various indirect questioning techniques have been proposed such as the Three Card Method [10], Nominative technique [23], Item Count technique and Randomized Response technique (RRT) [6].They focus on designing measurement instruments with anonymity as a core principal

rather than an augmentation, with the hope that strict guarantees of answer confidentiality will encourage participants to answer truthfully. For example, RRT is mostly used during in-person structured interviews. A sensitive question Yes-No is asked of the participant and the researcher asks them to privately flip a coin. The participant is asked to answer 'Yes' if the coin is tails and truthfully if the coin is heads. Only the participant knows the result of the coin toss, thus whether the answer reflects the truth or is an affirmative one due to the coin toss is hidden from the researcher. This gives participants more confidence to respond truthfully if the coin toss result is heads. Since the chances of the coin flip being one of heads or tails in a large sample is half, half the population will respond 'Yes' irrespective of the truth and the other half would have answered truthfully. Thus whatever proportion of the group said "no", the true number of those who disagree is double that, because we assume the two halves are probably the same as it is a large randomized sampling. For example, if 20% of the population surveyed said "no", then the true fraction that disagree with the statement is 40%.

In this research, we decided to use the *list experiment* technique [30]. We opted to closely follow the method in Marques et al.'s [21] recent study of snooping on smartphones, as the type of attacks in that study and in ours have similarities: they both involve unauthorized physical access to devices, and, in both cases, perpetrators are likely to be social insiders.

In list experiments, participants are randomly split into a *control*, and a *treatment* groups. Participants are presented with a *list question*, which is a set of items, typically formulated as statements, and a prompt to indicate *how many* they agree with, but not which ones. List questions presented to the control and treatment groups are similar, both containing a set of *control items*, that is, statements that are of no interest to the research question. However, the set of items presented to the treatment group has an additional *treatment item*. Assuming that participants in the control and the treatment groups select, on average, the same number of control items, the difference in the mean number of statements selected per group is, then, the estimated proportion of participants who selected the treatment item.

While list experiments may not be the most effective technique in reducing the social desirability bias [31], the advantages of this method are that it is easy to explain to participants and highly deployable in online surveys without requiring

Now I'm going to read you three things that sometimes make people angry or upset. After I read all three, just tell me HOW MANY of them upset you. I don't want to know which ones, just how many.

(1) The federal government increasing the tax on gasoline
(2) Professional athletes getting million-dollar-plus salaries
(3) Large corporations polluting the environment

In the affirmative action condition, an identical question was read, except that the list contained four rather than three items, with the additional item reading:

(4) Black leaders asking the government for affirmative action

Figure 3.1: An example of the item count technique used by Gilens, Sniderman and Kuklinski [15] to measure the perception of affirmative action in the American population

the survey to be conducted in-person. In contrast to list experiment the randomized response method (RRT) [3] requires time and attention from participants; resources which are in short supply on surveying platforms like Amazon Mechanical Turk [27]. In online scenarios it is difficult to convince respondents that randomizing methods like coin flips are not secretly being recorded decreasing the likelihood of truthful answers.

For the purposes of this research, and unlike in Marques et al., we opted to have two treatment groups. One group would be shown a treatment item that identified participants having been victims of social insider attacks, and the other as perpetrators. The difference between those estimates was expected to offer some insight into how common it is for people to never learn that they were victims.

We also decided to focus on the population of US Facebook users, since the adoption rate of Facebook among US adults (all demographic groups above the age of 18) is high; according to a 2014 Pew survey, 62% use Facebook [29]. This fact would make it easy to find Facebook users among US MTurk workers, from where we were to recruit.

3.2 List Experiment Design

3.2.1 Design Considerations

One important design consideration in list experiments is the composition of the list question. Common advice on building list questions includes:

1. **Avoid floor and ceiling effects** Ceiling effects are experienced when statements common to both control and treatment are so common that respondents would truthfully agree with almost all of them. Conversely, floor effects happen when respondents would almost always truthfully disagree with statements because they are so rare. In both cases, respondents in the treatment group may fear that answering the question truthfully would reveal their true (affirmative) preference for the sensitive item [3]. Details on reducing floor and ceiling effects are provided in Section 3.2.4.

2. **Avoid lists which are too short or too long** The number of items in a list experiment is a trade-off between the variance and likelihood of ceiling or floor effects. The fewer statements in a list question increase the chances of experiencing floor or ceiling effects. If, for example, the list comprised of two statements(one control and one sensitive statement), we would guarantee a floor or ceiling effect since a participant can either agree (ceiling) or disagree (floor) with the control statement and reveal the sensitive statement with an ambiguous answer. If the number of statements in the list is very large, the variance of the mean number of statements agreed with increases, making it harder to establish a significant difference in the mean number of statements agreed with by participants responding to the control and treatment surveys. Furthermore, longer lists require more attention and take longer to complete. Based on design recommendations by [3, 21], we chose to use 4 control items for the experiment.

3. **Avoid contrasting effects** Chosen control statements must not have a sharp contrast with the sensitive statement since respondents might be able to spot the sensitive statement and worry that any non-zero answer to the list experiment indicates an affirmative response to the sensitive statement. Thus we

ensured that all candidate statements we created (Figure 3.2) were directly related to Facebook, its users habits and activities.

3.2.2 Treatment Item Phrasing

We created two treatment items: a statement that would identify participants as victims of social insider attacks, and a statement that would identify them as perpetrators. Through multiple iterations, we ultimately settled on the following wording:

- **Perpetrator** I have used a device of someone I know to access their Facebook account without permission.
- **Victim** Somebody I know has used my device to access my Facebook account without permission.

We avoided, as much as possible, using security terms, like ”perpetrator”, ”attack”, ”victim”, or ”insider”, to not bias participants, and to reduce the contrast with control items. We used ‘my device’ to imply a physical attack, ‘Someone/-Somebody I know’ to imply insider, and ’access without permission’ to refer to the attack.

3.2.3 Control Items

To select 4 control items for the list question, we ran a direct question survey with MTurk workers. Our goal was to find a combination of control items that would minimize the chances of ceiling and floor effects. In other words, we wanted to find such a set of 4 statements, for which participants would rarely agree with either all or none of the statements.

Our task advertisement asked for participants who have a Facebook account and avoided charged terms such as “privacy” or “attack”. The survey consisted of demographic questions such as age, level of education and the state of residence. We also explicitly asked participants to indicate whether or not they had a Facebook account. Following these questions, participants responded to a list of 22 checkbox items with the prompt “Please check all statements that apply to you”. We placed this question last so that the participants would not be overwhelmed by the

long list of questions. Workers were paid $0.20 for completing the survey. Only workers with location set to US were allowed to participate. At the beginning of the survey, a filter based on IP addresses further prevented participation from non-US locations.

The statements in the check-box question were 20 candidate control items, drawn from previous research on motivations for Facebook use [34] and common Facebook use cases developed by the research team in brainstorming sessions. We also included the 2 treatment items, so that we could have estimates both from direct questioning, and from the list experiment. The ordering of the statements was randomized when presented to each participant.

3.2.4 Results of Item Selection Pilot Survey

We collected 202 complete responses, and excluded 28, which either indicated that participants did not use Facebook, or were given in less than 40 seconds (based on a prior pilot with 5 native English speakers). The remaining 174 participants reported an age range from 19 to 69 (mean = 33.7, SD = 10.6, and a gender distribution of 43% male, and 57% female. Table 3.1 shows the percentage and number of respondents who checked each statement.

We selected 4 control items from the list of 20 candidate control items that would result in the fewest cases of floor and ceiling effects, if they were administered to the same sample. Statements 7, 8, 13, and 16, also shown in Figure 3.2, were thus selected.

Having included the treatment items to the check-box question, we were also able to estimate that, under direct questioning, 8.6% of participants identified as perpetrators of social insider attacks, and 9.2% as victims. Peeking at the results of the list experiment (described in the next section), the estimates obtained with direct questioning were less than half than those obtained with the list experiment.

Some limitations related to the selection of items remain. For control items, it is possible that some candidate control statements might have been perceived as sensitive by some participants, thus, subject to the same bias as the treatment statements. For example, some might consider the number of friends they have on Facebook a sensitive subject, if they feel it is correlated with their popularity.

Additionally, the wording used for the control items was crafted not only to minimize the likelihood of participants perceiving them as sensitive, but also to limit their contrast with the sensitive items. Yet, some contrast is unavoidable, which may lead to underestimation in our measurements. Finally, the treatment items are subject to participants' own interpretations, which might not be consistent across participants, or coincide with our definition of a social insider attack, despite the broadness with which we scoped the construct.

Statement		Participants	Checked
1	I have posted a message in a group on Facebook and received a reply	109	62.6%
2	Someone I know has posted content on my Facebook wall	103	57.5%
3	I have received 5 or more unsolicited messages from strangers on Facebook	58	32.4%
4	One of my relatives has sent me a friend request on Facebook	117	65.4%
5	I have posted a picture of myself on Facebook	119	66.5%
6	Someone liked one of the pictures I posted on Facebook	118	65.9%
7	I have more than 300 friends on Facebook	81	45.3%
8	I am friends with one of my parents on Facebook	78	43.6%
9	I check Facebook every day	142	79.3%
10	On average, I spend more than 30 minutes on Facebook every day	100	55.9%
11	I have changed my Facebook profile picture in the last 12 months	109	60.9%
12	In the last week, I have clicked on a link posted on my Facebook newsfeed	91	50.8%
13	I have commented or liked a post in the last month on Facebook	123	68.7%
14	I am a member of a Facebook group	136	76.0%
15	In the last week, I have checked Facebook while at work	103	57.5%
16	I have reported an account on Facebook	48	26.8%
17	I re-shared someone's post on Facebook	111	62.0%
18	I have made my birth date publicly visible on Facebook	90	50.3%
19	I have clicked on an advertisement on Facebook	105	58.7%
20	I have responded to an event invitation on Facebook	99	55.3%
21	**I have used a device of someone I know to access their Facebook account without permission**	**15**	**8.6%**
22	**Somebody I know has used my device to access my Facebook account without permission**	**16**	**9.2%**

Table 3.1: Statements in a multiple choice question, administered to 174 MTurk workers, and respective percentages and number of respondents who checked them. Statements 1 to 20 were candidate control items for a list experiment; statements 21 and 22 were treatment items.

3.3 List Experiment Study Procedure

For the list experiment study, we again recruited among US MTurk workers and limited participation only to those who were accessing our survey server from US IP addresses. As before, we avoided words like "privacy" or "attack" in the task advertisement and consent form, with participants being informed that they were filling out a survey about their behavior on social media, and that being a Facebook user was a requirement for participation.

After providing consent, participants were randomly assigned to either the *Control* group, the *Treatment-P* group (which included the treatment item indicating that participants had been perpetrators), or the *Treatment-V* group (which included the treatment item indicating they had been victims). Again following Marques et al. [21], we added an attention check statement to all versions of the list question, that we expected no participants to agree with ("I have had dinner with the founder of Facebook, Mark Zuckerberg"). The final list question can be seen in Figure 3.2.

Other than the list question, the survey contained questions on participant age, gender, level of education, US state of residence, and OSNs which participants used. Each worker was paid $0.20 for completing the survey.

Blair and Imai [3] statistically validated list experiments using sample sizes of 500, 1000 and 2000 and concluded that although a greater sample size generally leads to a greater statistical power, the region where the proposed test has zero statistical power stays the same regardless of the sample size. Thus, we chose 1500 responses as a 'safe' choice when deciding size of the experiment.

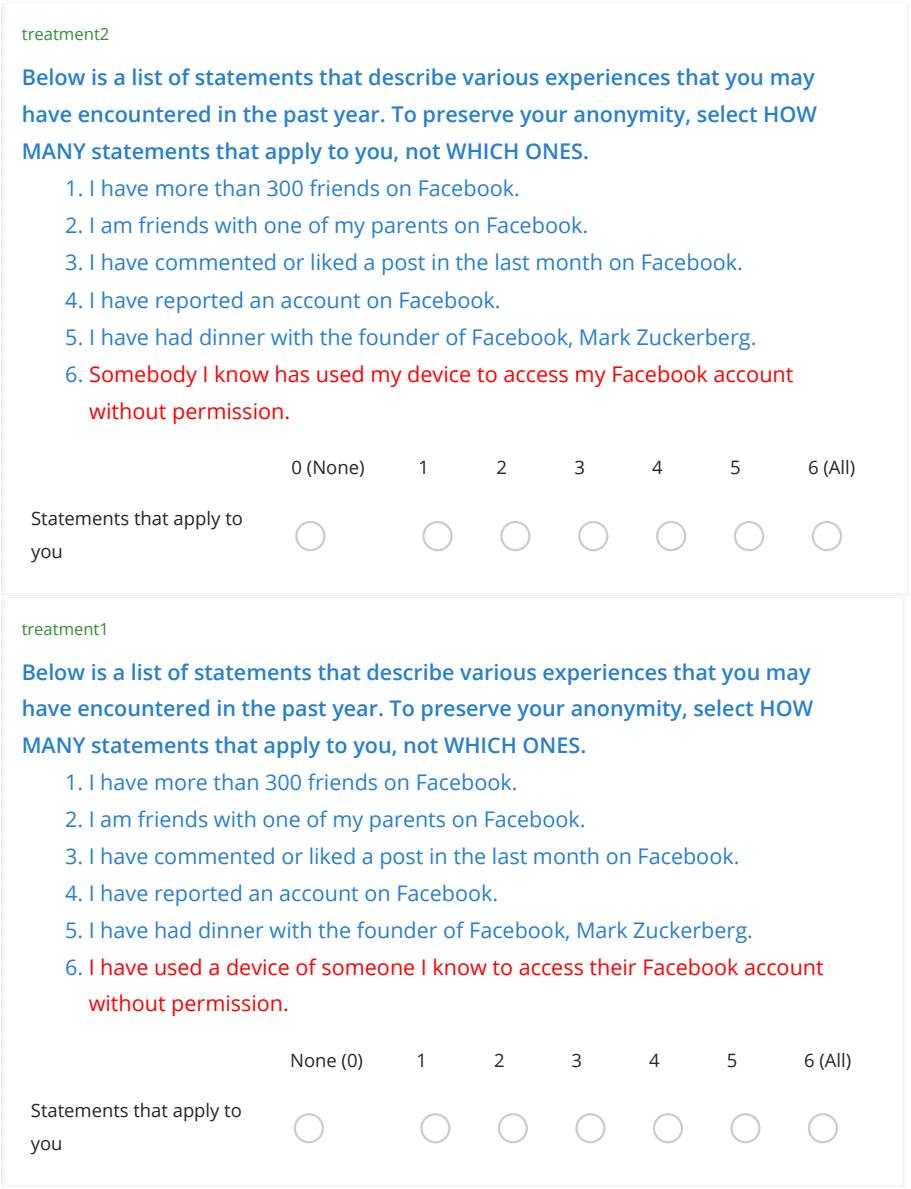

Figure 3.2: List question administered in list experiment, including 4 control items selected to minimize for ceiling and floor effects, 1 attention check item, and 2 treatment items (highlighted in red only for the manuscript), each administered to a separate treatment group. The control group did not have a treatment item.

3.4 Dataset

3.4.1 Data Clean-up

We received a total of 1,512 complete responses, and cleaned up the data by applying the following exclusion criteria:

- Responses in which participants had agreed with all statements (including the attention check one).
- Responses in which participants failed to confirm they used Facebook.
- Responses that took less than 30 seconds to complete (based on a prior pilot with 5 native English speakers).
- Responses in which the reported age was below 18.

We were thus left with 1,308 responses, on which the following analysis is based.

3.4.2 Participants

Out of the 1,308 validated participants, 440 were assigned to the control group, 423 to Treatment-P, and 445 to Treatment-V. Overall, reported ages ranged from 18 to 72, with the mean being 32.9 (SD = 10.16). Reported genders were 49% female, and 51% male. Most participants indicated being college graduates (52%), followed by those indicating being high school graduates (29%), and those indicating having post-graduate degrees (16%). Grouping reported states of residency into Census regions, the geographical distribution was 32% South, 21% West, 21% Midwest, and 18% Northeast. On average, participants reported being on 3.29 Online Social networks (OSN) (SD = 1.38), with only 9% reporting being only on Facebook. Reddit (65%), Twitter (56%), Pinterest (37%), LinkedIn (23%), Tumblr (19%) and Instagram (9%) were the most popular among participants, aside from Facebook.

To test for a priori demographic differences between the control and the treatment groups, we ran a logistical regression of group assignment per all available

	Control (n_c = 444)	Treatment-P (n_{t1} = 423)	Treatment-V (n_{t2} = 445)	Total (n = 1312)
Gender				
Male	51 %	53 %	48 %	50.8 %
Female	48.5 %	47 %	51.2 %	49 %
Other	0.5 %	0 %	0.2 %	0.2 %
Age				
18-24	16.3 %	18.6 %	20.6 %	18.5 %
25-34	47 %	51.2 %	45.9 %	48 %
35-44	21.9 %	18.8 %	19.2 %	20 %
45-54	9.5 %	4 %	5.4 %	8.6 %
55-64	4.5 %	4 %	5.4 %	4.7 %
65 +	0.9 %	0 %	0 %	0.3 %
Education				
High school	27.6 %	28.9 %	30 %	28.8 %
College	51.1 %	52.3 %	52.1 %	51.9 %
Graduate School	17.6 %	15.2 %	13.8 %	15.5 %
Other	3.6 %	3.5 %	4.1 %	3.7 %
Region				
Midwest	23.1 %	18.3 %	21.2 %	20.9 %
Northeast	18 %	16 %	22.4 %	18.9 %
South	35.9 %	36.4 %	35.3 %	35.8 %
West	22.9 %	29.2 %	21 %	24.3 %
# OSN accounts				
One	8.3 %	10.9 %	7.9 %	9 %
Two	19.8 %	21.5 %	20.9 %	20.7 %
Three	33.8 %	25.1 %	29.7 %	29.6 %
Four	22.3 %	26 %	23.1 %	23.8 %
Five+	15.8 %	16.5 %	18.4 %	16.9 %

Table 3.2: Summary of participant demographics in list experiment study

demographic variables, and then applied the stepwise procedure for variable selection. The selected model had no demographic variables, which indicates a lack of evidence for a priori demographic differences between groups.

We compared demographic variables between our survey sample and that of the target user population. Specifically, we looked at age and gender to see how close our survey sample was to the US Facebook user population in general. Comparing to the latest (July 2016) gender and age data available from Statista [35, 36], our

	Control	Treatment-P	Treatment-V
0	8 (1.8%)	8 (1.9%)	10 (2.2%)
1	87 (19.6%)	58 (13.7%)	68 (15.3%)
2	145 (32.7%)	143 (33.8%)	142 (31.9%)
3	156 (35.1%)	124 (29.4%)	136 (30.6%)
4	48 (10.8%)	77 (18.2%)	72 (16.2%)
5	-	13 (3.1%)	17 (3.8%)

Table 3.3: Number and proportion of respondents who selected each option in the list experiment item (adjusted for 4 control items). Each row represents an option indicating the number of statements agreed with by participants. Treatment-P column shows option choices made by participants that were presented with the perpetrator statement as the sensitive item. Similarly, Treatment-V column shows option choices made by participants that were presented with the victim statement as the sensitive item.

Group	Proportion of Facebook Users	Proportion of respondents
20-29	28%	46%
30-39	21%	33%
40-49	16%	12%
50-59	13%	7%
60+	13%	3%
Male	46%	51%
Female	54%	49%

Table 3.4: Comparison of age-group distributions between US Facebook and our respondents in the list experiment study. Our sample population was younger than the US Facebook population

sample was younger and slightly skewed to males, as is shown in Table 3.4.

3.5 Results

3.5.1 Prevalence Estimates

The distribution of number of statements agreed with by participants in the list experiment is shown in Table 3.3, and this served the primary source of experimental

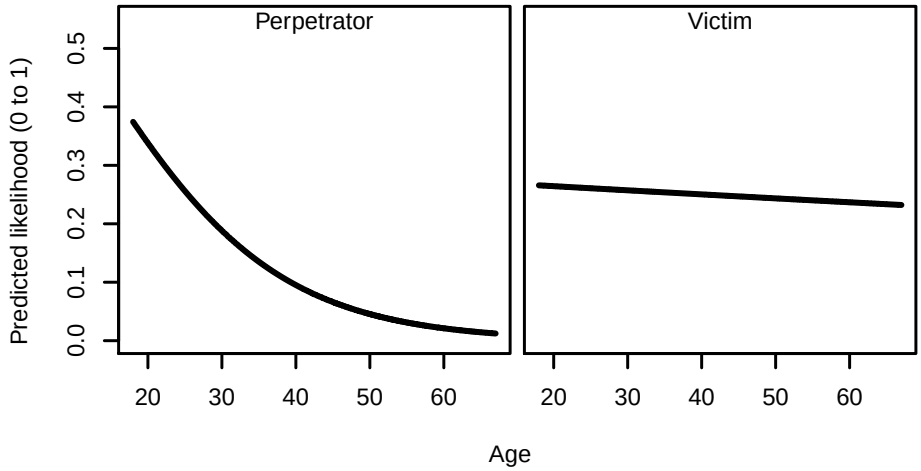

Figure 3.3: Regression model of likelihood of being a perpetrator, or a knowing victim of social insider attacks on Facebook, predicted by age of participants.

data.

The mean number of items selected was 2.334 (SE [standard error] = 0.046) in the control group, 2.574 (SE = 0.053) in Treatment-P group, and 2.546 (SE = 0.053) in Treatment-V group. The estimates of participants identifying with the treatment items, based on the differences in means technique, are:

- **Perpetrator** 24.0% (SE = 0.070)
- **Victim** 21.2% (SE = 0.070)

3.5.2 Effects of Age and OSN Participation

Marques et al. [21] found evidence that snooping on other people's mobile phones was more prevalent among younger people, and among people that had adopted the smartphone more deeply (used their own smartphones such that it would retain more private data.) To verify if similar effects existed in social insider attacks on Facebook, we ran list experiment regression models [3] on the age variable, and, lacking a specific measure of depth of adoption, on the count of OSNs that participants reported using.

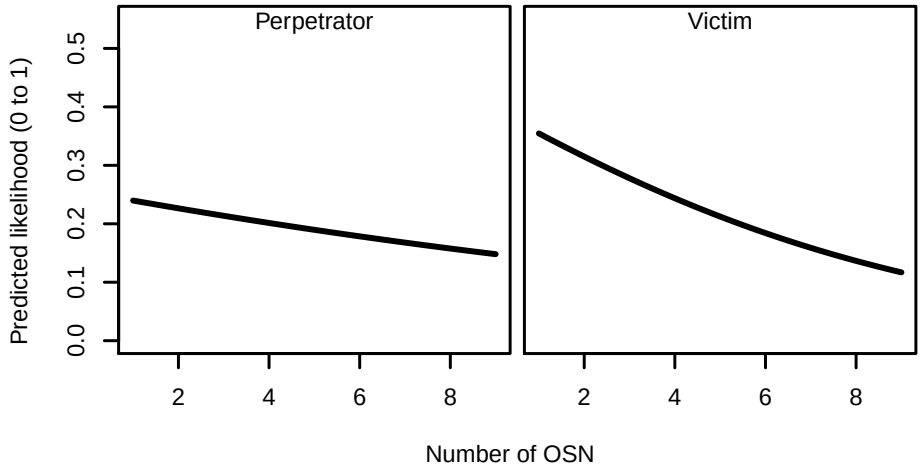

Figure 3.4: Regression model of likelihood of being a perpetrator, or a knowing victim of social insider attacks on Facebook, predicted by age number of OSNs participants used.

Figures 3.3 and 3.4 depict those regression models graphically. Regarding age, there is indeed a visible pattern of decreasing likelihood of being a perpetrator of social insider attacks as age increases. However, that age pattern is much more less pronounced, and indeed almost flat, for the likelihood of being a victim.

For the number of used OSNs, the opposite seems to be true. Using more OSNs was a weak predictor of being a perpetrator; at best, using more OSNs slightly decreases the likelihood of conducting the attacks. For being a victim, however, the pattern appears to be clearer: the more OSNs participants used, the less likely they were to be victims of such attacks.

Model Predictions

Because estimates of positive responses to the sensitive item have to be recovered from aggregates, list experiments reduce social desirability bias at the expense of statistical efficiency. List experiment regression models [3] can recover some of that efficiency and predict, for each participant, the likelihood that they have identified with the sensitive item. To obtain such predictions, we built another list experiment regression model, with age, number of OSNs used, and the interac-

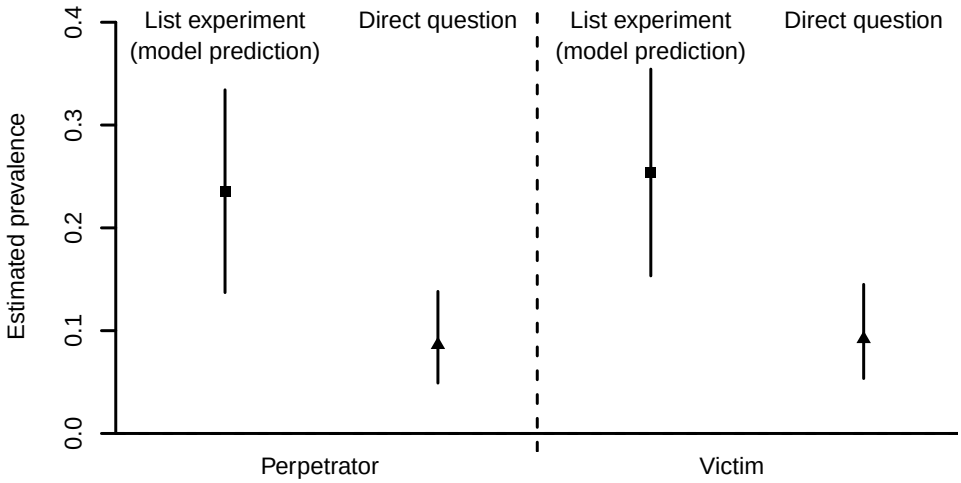

Figure 3.5: Estimated prevalences based on list experiment model predictions, and response to direct questions. Predictions, and 95% confidence interval of predictions, from a list experiment regression model of age, number of OSNs participants used, and the interaction between the two terms. Proportion of positive response to direct questions, and 95% confidence intervals, from the item selection survey (n = 174).

tion between the two variables. From the model, we obtained the predicted per-participant likelihood of being a victim or a perpetrator. Those predictions, and a 95% confidence interval of predictions, are depicted in Figure 3.5. The points represent the mean of predictions, and therefore approximate, but do not exactly match, estimates obtained with differences in groups means. For reference, the figure also depicts the proportion of participants that selected the sensitive items in the 174-participant item selection survey (see Table 3.1), and respective 95% confidence intervals. The graph illustrates that the prevalence estimates obtained with direct questions are considerably lower than the ones obtained with list experiments, which can be attributed to social desirability bias. It also illustrates the loss of statistical efficiency, reflected in wider confidence intervals for model predictions, even with much larger sample sizes (14-33% for perpetrator, with n = 863, and 15-35% for victims, with n = 885).

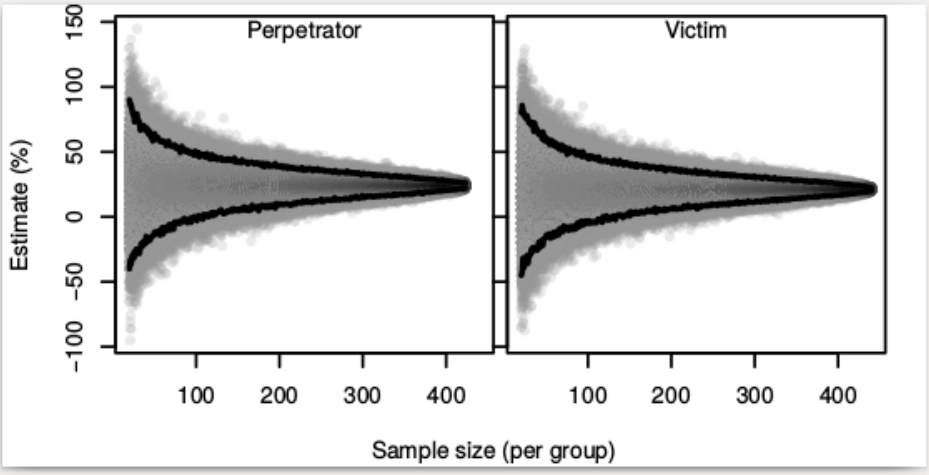

Figure 3.6: Prevalence estimates obtained with 1000 simulations at each increasing sample size. Each dot represents the estimate prevalence of being a perpetrator or victim, when only considering a random subset of responses. The black line represents the 2.5 and 97.5 percentiles of estimated prevalence at each sample size.

Simulation

We further ran simulations on the data to see how sample size varied with the result using the difference in means technique. At each round of simulation, we randomly sampled an equal number of participants from each group, and calculated the difference in means in their responses to the list question, repeating this process 1000 times. We started with 1000 simulations with 20 participants per group, and iterated until 423 participants per group, 423 being the lowest number of participants in a group (Treatment-P). Figure 3.6 depicts the results of these 403,000 simulations, with the black lines delimiting 95% of obtained estimates at each sample size. As expected, with small samples, the range of estimates was high, and, as sample sizes increase, the range decreases sharply. For instance, with samples of 150 responses, 95% of estimates for being a perpetrator were between 7% and 43%, and 95% of estimates for being a victim were between 1% and 40%. However, with samples of 400 participants, those ranges fell significantly: 20% to 28% for being a perpetrator, and 17% to 26% for being a victim. Although such ranges

should not be interpreted as confidence intervals, they are an indication that the estimates we obtained are likely to be within an acceptable range of the estimates that could be obtained with larger samples.

3.6 Discussion

The main objective of this study was to estimate how common social insider attacks on Facebook are. The results suggest that they are not uncommon, with 24% of participants estimated to have implicitly identified with the statement "I have user a device of someone I know to access their Facebook account without permission", and 21% with "somebody I know had used my device to access my Facebook account without permission".

Since the estimates for being a perpetrator and being a victim are close, we cannot conclude that victims are often unaware of attacks.

Contrasting the estimates obtained through the list experiment (24% Perpetrator, / 21% Victim) with the ones obtained through direct questioning (9% Perpetrator, / 9% Victim), possible effects of social desirability bias can be observed. This effect was expected for perpetrators, as people are generally unwilling to openly admit behaviors of this kind [21]. For victims of social insider attacks, the effect was more surprising, and could potentially be related to victims assigning themselves responsibility for intrusions, for example, being embarrassed that simple operational security oversights allowed the attack to happen. [32].

The regression models we fitted also indicate two clear trends. First, that younger people are more likely to conduct social insider attacks, mirroring prior findings on mobile phone snooping [21]. Second, that people who use more OSNs are less likely to be victims. One possible explanation for this trend is that those people tend to be more tech savvy and more aware of what private information is retained on OSNs, thus being, respectively abler, and more motivated, to protect themselves.

We acknowledge that our experiment is limited by the fact that our results cannot be generalized to the US adult Facebook population since the age of respondents in our survey was younger and serve as a approximation of the prevalence of social insider attacks in the wild. Furthermore, like all surveys with sensitive

questions, we rely on participants honesty to answer questions truthfully with the expectation that strict anonymity will encourage them to do so. The rate at which different demographic groups answer honestly may also vary, for instance, individuals in some groups (like youth) may be less likely to do so because the effect of the desirability is felt more strongly.

From a security perspective, these findings suggest that the probability of social insider attacks on Facebook is not negligible. We noted that this estimate was significantly greater than the prevalence of other common security threats as the previously mentioned 419 scams [1]. Nevertheless, as important as probability, is the severity of the threat. The study reported in the next section provides insights into this issue.

Chapter 4

Social Insider Attack Dimensions Study

We established that social insider attacks are common but we did not yet have insight into what exactly a social insider attack was. More specifically, we sought to establish what it means to conduct a social insider attack, what they looked like, why they took place, how they happened and what consequences of such attacks were. To find out, we used a qualitative approach to cast as wide a net as possible for the various dimensions that influenced, affected and pertained to social insider attacks.

4.1 Methodology

We collected qualitative data through an online survey where we asked participants to report on social insider attacks with them either being the perpetrator or the victim. This survey was deployed on Amazon Mechanical Turk. It included a consent form, and filling of qualification and demographic questions to ensure that participants were eligible for participation. The main eligibility criteria was having perpetrated or being a victim to a social insider attack on Facebook. Other requirements included being at least 15 years old, and having used Facebook in the past 12 months. As before, we chose to focus on US participants only, thus being geographically located within the US was a requirement to be able to accept the

task.

The main section of the survey was the open-ended question where participants were asked to write a story about a past experience with social insider attacks on Facebook. The prompt emphasized that the participants relate a real story that they themselves had experienced rather than a fictional or third-person account.

To minimize priming participants, we avoided using charged terms in survey advertisement and questions. Instead of labeling the phenomenon as a social insider attack, we referred to it as *an instance where either you accessed the Facebook account of someone you know without their permission, or someone accessed your Facebook account without your permission*. We also avoided language that portrayed the incident as overly negative so that participants would not be dissuaded from writing about their experience truthfully. To protect participant anonymity and avoid self-implication, we asked for no personally identifying information in any of the sections of the survey. We asked respondents to use gender neutral names: *Casey* as the person who perpetrates the social insider attack, and *Alex* as the target of the attack.

4.2 Data and Analysis

We collected and performed thematic analysis on a total of 45 stories reporting social insider attacks. Stories had min/mean/max word count of 92/263/527 from which three researchers inductively created and refined a codebook, until saturation was reached at 35 stories The final codebook had a total of 71 codes across six main themes (perpetrators and victims, premeditation, attack vector, attack variants, attack aftermath, motivation). A batch of 10 more stories was collected from which inter-rater reliability for two independent coders was calculated (Cohen's kappa k=0.95).

Participants in the study were 59% male and 41% female with a minimum, maximum and average age of 15, 56 and 32 respectively. They were geographically spread across 22 states from all four US census regions. We provided above average compensation of $4 and offered a bonus of $1 if the story was well written as an incentive.

4.3 Findings

In this section, we present our findings, structured by the main themes that emerged in the analysis. These themes depict the sequence of events of an attack, describing the circumstances before, during, and after the attack, as reported in the stories.

4.3.1 Perpetrators and Victims

The stories noted a variety of perpetrator-victim relationships. The variability in social and physical proximity had, unsurprisingly, a significant impact on the attack motivations and in some cases, the type of attack launched. Relationship types included parent-child, married couples, dating couples, ex-romantic couples, intimate friends, co-workers and acquaintances described by terms like close, in love, best friends and having worked together. Respondents gave important context as to the state of their relationship before the attack which was as important as the relationship itself and often gave probable cause for the motivations of the attacker. In some cases, they explicitly identified that their relationship was struggling:

> Casey and Alex lived together as a couple in (redacted). They were a heterosexual couple that were breaking up due to Casey's infidelity and crazy behavior. [Story 7]

Some common relationships such as that of a parent and child had an atypical relationship dynamic. In one case the parent and child roles were inverted, with the child tending to act as the parent and the parent acting irresponsibly. However, the social contract of being a parent gave the perpetrator a justification to conduct the attack:

> (Casey) would spend all hours of the day playing one game to the next. Alex had to keep making sure they were eating and drinking, and being insistent Casey get some sleep. ...(Casey)they snuck into Alex's room while they were asleep. Casey had it in their mind that they were the parent, they had full right to access Alex's personal computer and their Facebook account ...When Alex woke up, seeing their parent exhausted, slamming a very expensive mouse because they missed a rare tree, there was a long talk. [Story 2]

4.3.2 Premeditation

The reported attacks were either premeditated or opportunistic. In premeditated attacks, the perpetrator was proactive in bypassing device and account security measures. In one case, the perpetrator actively searched for the victim's password in their living space:

> Casey started snooping through Alex's belongings, Alex's wallet, desk, folders, but had no luck, maybe he kept his passwords on the computer or in his head. [Story 8]

In another case, the perpetrator installed key logging software onto a shared device to steal the password:

> I kept putting off installing a keylogger so that I could get her passwords and then go have a look around her email accounts and Facebook. [Story 39]

Opportunistic attacks were enabled by two factors: (1) victim's negligence, and (2) an activity that separated the victim from their device. For example, in one story the opportunity arose while the victim was in the shower:

> Alex left his phone on the table in front of her while he went to go take a shower. Casey knew that Alex would be taking a shower for awhile and usually took around thirty minutes. [Story 20]

We also noted that victims used poor security practices, such as not logging out of their Facebook account:

> Alex had a habit of signing into Facebook on their laptop and forgetting to log out after using the site. [Story 29]

Since the attack took place on the victim's personal device (or one they had regular control over), victims in our stories did not take measures to safeguard their account or device. Two possible explanations for this is that they did not think that unauthorized access could come from someone they knew well, or that they felt a false sense of security knowing that the particular device was under their close watch.

4.3.3 Attack Vector

The absence of device- and account-level protection was a common feature in many social insider attacks. And, in the presence of additional protection, such as biometric verification, perpetrators used creative coercive techniques:

> Alex's iphone used fingerprints for access, so Casey grabbed Alex's sleeping hand and pressed a finger up to the sensor on the iphone. [Story 6]

In some cases, the perpetrator shared passwords with the victim with the supposed mutual understanding that they would respect each others privacy, considerably lowering the bar to initiating an attack.

> I didn't have any trouble getting into the phone because, as I said, I knew the code to his and he knows the code to mine as well. [Story 24]

In several stories, we observed a mismatch between the perceived security of victims' account and how, in reality, accounts were exposed to people in the victims' social inner circles, indicating that both security measures, and how people innocently create breaches, are opening vectors for attacks to their privacy.

4.3.4 Attack Variants and Target Assets

We noted a number of attack variants in our data, including impersonation, snooping, and data destruction. Impersonation involved the perpetrator performing actions on Facebook in a way that others would believe that the actions were taken by the victim. In snooping attacks, the perpetrator silently looked for information in the victim's account. In data destruction attacks, the perpetrator deleted victim's information like messages, photos, or videos. In some cases, perpetrators actively covered their traces:

> Casey switched off notifications from the statuses and hid them from Alex's time-line, ensuring that he could not find out that they even existed! [Story 1]

Some attacks were a combination of the above attack variants. In such cases, one attack variant would follow another until the perpetrator achieved their goal:

> Casey suspected Alex of cheating and picked up the phone to see if the suspicions were correct. They ended up finding nothing at all. However that was not enough. Casey used Alex's phone to start messaging random girls that were friends asking if they wanted to have a sexual encounter. [Story 32]

Attacks focused on a variety of assets in the victims account, such as the newsfeed, liked posts, the victim's profile, photos, videos, messages, posts/comments/status-updates, and notifications. However, some attack variants targeted some of the

assets disproportionately (see below, under *"Motivation"*).

Attack variations also had a direct influence on how they were discovered by the victim. Impersonation attacks were generally the most noticeable, as they resulted in a visible action on the victim's account. Snooping attacks were the most challenging for their victims to detect, as they did not leave explicit traces. Victims were sometimes able to trace their perpetrators because there was no other possible explanation.

> Alex allowed Casey to use their phone to make phone calls on several occasions at work . . . Alex was a bit curious why FB was listed as an open program on their phone, even though they were sure that it had not been open before they had lent Casey their phone. [Story 41]

In other cases, perpetrators admitted to attacks either by stating it upfront, or by confronting the victim with information they found during the attack.

> Casey told Alex the next day that they knew that Alex was talking to their former partner. [Story 25]

4.3.5 Attack Aftermath

The stories in our dataset recorded a range of social and emotional consequences as a result of the attack for both the perpetrator and the victim. For the most part,victims were often livid with their attackers:

> When Alex found out he was furious. He had not cheated and felt their relationship could not recover from this breach of trust. [Story 32]

Many attacks led to permanent changes in the relationship between the victim and the perpetrator including ending of marriage, commitment, and friendship. Perpetrators primarily exhibited relief or regret, but some, upon further reflection of their actions, displayed a greater depth of emotion including a sense of empathy for their victim.

> Casey learned some troubling things, while peeking through Alex's facebook, things that were frightening and sad. It disturbed Casey to know that Alex was going through things and hadn't been talking about it. . . . Only now, Casey knew some things about Alex that hadn't made any sense at all. [Story 3]

From prior work, we know that people care about privacy from social insiders [19] and social insider attacks are a violation of privacy. However, we observed

a dichotomy in emotional aftermath. On the one hand, attacks perceived by the victim as privacy violations had severe impact. These used terms like 'furious' and 'mad'. On the other hand, in a few cases, attacks were simply laughed off, either because they did not perceive the attack seriously or they found a way to justify the attack to themselves irrespective of the privacy violation:

> I'm assuming he didn't do it because he didn't trust her, I just think he was bored and was looking for something to do. He told her that he had accessed her Facebook account. She wasn't upset at all. [Story 23]

Overall we noticed a great deal of variability in emotional aftermath of the attacks though, understandably, they tended to be mostly negative.

Some victims responded to the attack by changing their Facebook passwords and employing better security, such as using device auto-locking mechanisms and logging off their account after each use:

> From then on Casey always made sure to log off Facebook and made sure to change the password. [Story 36]

In one story the victim reported the attack to an authority, with significant consequences for the attacker:

> Alex had no choice but to call their boss and get Casey fired. [Story 35]

In the overall, if the attack intention was not meant to have fun or play a prank, the consequences of an attack for both parties are predominantly profound and harsh. These events are likely to affect relationships and emotions deeply. People tend to improve their security measures upon discovering an attack, which suggests that (1) they were not aware of or discredited the insider threat, and that (2) they were able to better protect their security, at the cost of convenience, once they become sensible to the possible perpetrations.

4.3.6 Motivation

We observed 5 types of motives: fun, curiosity, jealousy, animosity, and utility. Figure 4.1 shows the distribution of attacks with the aforementioned motives. We note that since this was a qualitative study, the figure does not represent actual frequency of such attacks in general. Motivations often implied other attack features, which we discuss below.

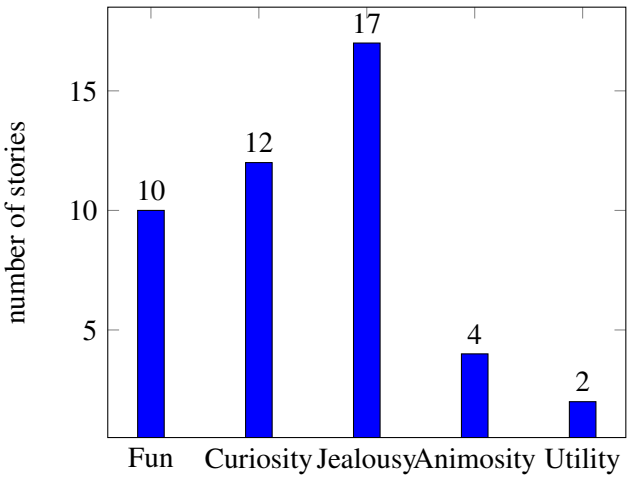

Figure 4.1: Distribution of Attack Motivations

Fun. Attacks were motivated by 'fun' if the perpetrator wanted to play a prank on the victim without a premeditated malicious intent.

In such attacks victims were either family members or friends of the perpetrator, and the attack was exclusively opportunistic. Prank attacks were short in length, and used impersonation. Perpetrators targeted highly visible parts of their victim's Facebook account such as the profile picture or status updates. They changed these to what the perpetrator perceived to be funny. How far the perpetrator went during the attack directly influenced its emotional aftermath for both parties. If the victim perceived the impersonation to be benign, they were amused.

> She posted "I smell" . . . (Alex) then told her that it was a pretty funny comment . . . [Story 4]

Some pranks had more serious consequences for the victim, who feared backlash of their Facebook account's social circle and posted apologies and explanations.

> The postings mainly inferred that Alex was coming out to his friends and was a gay person . . . Alex posted an apology and explanation on Facebook. [Story 37]

Pranks had little negative influence on the relationship. One story reported a positive outcome:

> Hence, there weren't any severe consequences except for a good laugh that probably

> ended up boosting more than hurting the friendship between Alex and Casey. [Story 22]

In terms of attack patterns, prank attacks were short in length, sometimes lasting only a few actions and were exclusively single instance access; the attacker accessed the victim's device only once.

Curiosity. Curiosity was assigned as the primary motive in cases where the perpetrator was curious about content on the victim's Facebook without a predetermined emotional foundation to the intent.

Such attacks were conducted against a range of social relations including friends, family and romantic partners. Nearly all attacks were opportunistic and perpetrators gained access to the victim's device because it had neither device-level nor Facebook account login security, e.g., already logged-in. The perpetrator simply could not resist the opportunity.

> (Alex) loved all his cousins . . . Casey was one of them . . . The account was already open so she didn't have to hack into it or anything. Being curious about any details in regards to Alex's potential relationships, she read a few of the messages and checked out the girl's FB page/pictures. [Story 31]

Attacks motivated by curiosity were exclusively snooping attacks but the relationship between the victim and perpetrator heavily influenced the targeted assets. Romantically involved individuals targeted private messages only, while family and friends snooped on the profile, photos, and public and private social interactions. Many attacks went undiscovered, but in some cases the perpetrator was caught in the act:

> Alex saw Casey hurriedly put down the ipad and remembered that his FB account was still open. He put two and two together. . . [Story 31]

Curiosity-motivated attacks had a high initial emotional impact on the victim but there were few stories that noted a long-lasting effect on their relationship. They lasted longer than prank-motivated attacks but still usually under 30 minutes long. Once again, they were exclusively single instance access, which is understandable since they were mostly opportunistic.

Jealousy. To limit the scope of a broad term, we restricted jealousy to that of an emotional nature where, for example, the perpetrator wanted to know if the victim had been emotionally involved with others.

In all the cases in this category, the victim and the perpetrator were romantically involved and often co-habitating, indicating that they were close socially and physically. Attacks motivated by jealousy were equally likely to be premeditated and opportunistic. One instance was a combination of the two:

> Casey heard a rumor from a friend that Alex is flirting someone else on Facebook. This angered Casey, however Casey could not confront Alex because there was no proof of the infidelity ... (One day) Alex walked into the home to find Casey asleep on the couch with the cell phone on the coffee table. [Story 9]

All stories noted that at least one level of security, either device or Facebook account, was bypassed trivially because the victim was already logged in. Most jealousy-motivated attacks lasted longer than 15 minutes and were of the snooping variety, targeting the victim's personal messages. This can potentially be explained by the fact that in these attacks the perpetrator is already socially close to the victim, and private messages are the only kind of information that they cannot readily access. Jealousy-motivated insider attacks had a high emotional impact for both the victim and the perpetrator and severe consequences for their relationship. Victims were often angry and felt their privacy had been violated. Perpetrators were often regretful, enough to admit to the attack, even if it had given them temporary relief.

> While Casey was relieved after checking his girlfriend's phone, he had an amazing sense of relief as well as incredible guilt . . . Casey decided later that day when he returned her phone he would tell Alex what he had done. [Story 10]

Nearly half of the stories explicitly mentioned an end to their relationship as a result of the attack. Attacks often lasted more than 30 minutes long with the perpetrator performing a large number of interactions with the victim's device. In some cases, perpetrators accessed the victim's device multiple times.

Animosity. In these attacks, the perpetrator's primary motive was to hurt the victim. This ranged from deleting the victim's data, diminishing the victim's social standing by impersonating them, and performing other disreputable actions with the victim's account that were visible to others. In these cases, the perpetrators had a spectrum of relationships with their victims, ranging from very close (ex-romantic partners), to far apart (co-workers).

Attacks with animosity as a motive used a combination of attack variants. Im-

personation was used to post mean comments about the victim's friends, destruction was used to delete victim's information, and snooping was used to gather messages, photos and videos that could be used against the victim later.

> (Casey) deleted everything on my account including pictures that only existed on Facebook. There were also mean messages sent to friends and relatives. [Story 7]
>
> Casey attacked Alex's LMGTQ friends, calling them all sorts of horrible names and even posted some very negative content. [Story 11]

The emotional aftermath was high for victims — they were angry, embarrassed, and felt that their privacy was violated.

> Casey was a horrible person. [Story 7]
>
> Casey made Alex look like a hateful person and changed how others viewed Alex in a single day. [Story 11]

Since most such stories were written from a victim perspective, there was little information about the perpetrator's emotional state. This was also the only category in which an outside authority, such as a boss, intervened. Escalation of the attack aftermath to an external authority seemed to have been a rare strategy; in most stories, the victim dealt with the attack on their own. For similar reasons, its difficult to tell how long attacks since its impossible to speculate how much time or number of actions it took the perpetrator to be perform the attack or how often they had access to the victim's device.

Utility. In utility motivated attacks, the perpetrator was not directly interested in the victim's account but wanted to use it to achieve a goal. For instance, using the account to view photos of a victim's social connection (Facebook friend):

> I only accessed it for a short period of time in order to look for attractive pictures of the aforementioned girl. [Story 44]

In another case, the perpetrator used the victim's account to play a Facebook game:

> Facebook games can be addicting You have little jobs that just keep building up, limited amount of energy to do them all in, and constantly needing friends to finish tasks. Casey was absorbed in this . . . (Casey) snuck into Alex's room while they were asleep.[Story 2]

Utility-motivated attacks were carried out exclusively against friends or family. Most attacks had little information to indicate significant negative emotional

impact for either the victim or the perpetrator; they were either benign or positive for their relationship. In Story 2 (quoted above), it acted as a pivot for positive emotional communication:

> When Alex woke up, seeing their parent exhausted, slamming a very expensive mouse because they missed a rare tree, there was a long talk. [Story 2]

Utility-motivated attacks length and number of accesses vary depending on what the the perpetrator was trying to achieve. In the quotes mentioned above one lasted a 'short period of time' while another lasted the entire night.

4.3.7 Impact

The purpose of the study was primarily exploratory; to see what different dimensions of social insider attacks are. Some trends were found across all types of attacks such as the social relation that was targeted and based on story details, perpetrators often found it very easy to access the victims device when cohabitating.

On the other hand, we discovered that there is a large variation in aspects like attack timing, targeted assets, duration of the attack most of which centered around the motivation of the attack. We find this to be a useful way to classify social insider attacks as well as inform counter-measures which could use this to detect different kind of attacks. An example that stands out is that of jealousy-motivated attacks where perpetrators were often performing the "message scroll" action; constantly scrolling up private message threads (Facebook usually paginates sets of messages so after approximately 20 messages, it performs a message set fetch) but not sending messages.

Duration of attack could be another useful tool to detect anomalous behaviour. For example, if the account owner usually logs in between 5 pm and 12 am, then an access at 3 am could be considered anomalous.

Individually, each dimension may not be useful to identify behaviour but based on our findings, they may paint a much more vivid picture when combined.

To adopt a countermeasure against social insider attack, users must perceive them as a threat. Our next study methodology informs this aspect; how severely do users perceive social insider attacks.

Chapter 5

Social Insider Attack Severity Perception Study Methodology

5.1 Background and Motivation

In the prior two experiments, we studied the prevalence of social insider attacks and several of their dimensions. However, from these studies we are unable to establish how the different kind of attacks scenarios we collected are perceived by Facebook users.

Another step to justify creating countermeasures against social insider attacks is to examine whether users perceive them to be a serious threat. If they do not, even if countermeasures were developed, users may not be inclined to use them. Furthermore, we must also find how likely users feel they could be targeted by such an attack because even if they feel the attack is severe but unlikely, they may not favor using mitigation techniques. Anecdotal evidence, surveys and studies suggest that the human factors are by far, often the weakest link in the chain of computer security. Humans often do not adequately understand security and privacy threats and sometimes lack the knowledge, desire or time to handle them properly [2].

For many, a Facebook account is to the cyber world what a users home is to the physical world; a vault of private belongings and interaction with others. In the same way an individual is concerned of protecting their house from threats like robbery, our study design aims to discover what Facebook users see as threats to

their personal and private information with respect to social insiders and suggest which aspects threat should mitigation techniques prioritize.

In this chapter, we outline methodology to perform a mixed qualitative and quantitative study to answer two primary research questions:

1. What is the perceived *severity* of social insider attacks (by Facebook users)?

2. What is the perceived *likelihood* of social insider attacks (by Facebook users)?

As of the writing of this thesis, this study is not yet complete. However, we suggest methodological procedure and recommendations based on the results of our previous studies and intend to complete it as future work.

5.2 Methodology

5.2.1 Survey Structure

We aim to perform a large-scale study on Amazon Mechanical Turk where participants will be asked to answer a 5-10 minute survey, preceded by a similar pilot. Both the survey and pilot comprises of two parts:

Demographic and Informational Section. In this section, we ask participants to fill out demographic questions of age, gender, state-level geographic location as well as an optional question regarding their relationship status. As an extra parameter, if the participant indicates that they are in a relationship, we ask them whether they share a living space with their partner. This is to identify a connection between social insider proximity and perception of risk or likelihood if one exists. Furthermore, we question the user's attitude towards computer privacy and security. To this effect, we used Security Behavior Intentions Scale (SeBIS) [13], a scale that allowed us to estimate how security-conscious respondents are by measuring their intentions to comply with computer security advice and best practices.

Primary Section. In primary section of the survey, respondents will be asked to read a story relating an incident of a social insider attack on Facebook. They will then be asked to rate it on a Likert scale ranging between 1 (Not Severe at all) to 5 (Highly Severe) to indicate the perceived severity of the attack in the story. Following the rating, they will be given space to write a qualitative response of

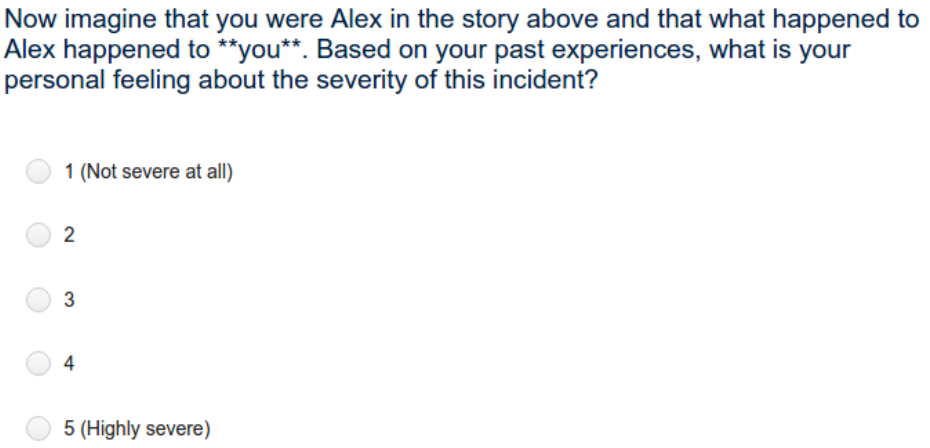

Figure 5.1: Prompt asking participants to rate the perceived severity of the social insider attack story

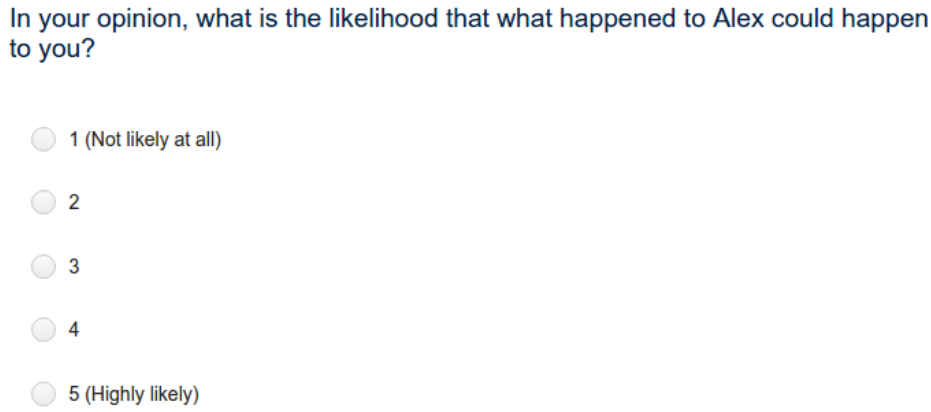

Figure 5.2: Prompt asking participants to rate the perceived likelihood of the social insider attack story happening to them

about 30 words to justify their rating. The question prompt (Figure 5.1) places an emphasis on creating empathy for the victim and asks participants to use their own past experiences as a reference when answering. Next, a similar question Likert Scale followed by qualitative response is presented where the participant will be asked to rate how likely they are to be the victim to such an attack (Figure 5.2).

5.2.2 Design Considerations and Limitations

The stories we use for this study are taken from previous social insider attack dimensions study. All stories used were reanonymized; all cases where the participants did not use the correct pronouns, tense or names have been edited so that story may adhere to the standard where characters are referred to by fictional names, 'Alex' and 'Casey' where Alex is the victim and Casey is the perpetrator. Each survey in the study will present participants with only a single story chosen at random from the pool of 45 collected in the previous study to rate for perception and likelihood. We chose use random sampling instead of using specific stories since typifying stories based on any given criteria would take away from the richness of the data observed in the stories.

When asked to perform a rating, we anticipate that participants would often find it difficult to assign a value since they may not have anything to compare it to. We could consider asking them to compare it to well known computer security threats of a similar nature such as bank fraud, email scams and others. This might be problematic for two reasons. Firstly, the comparison may not be a valid one; similar computer security threats may not be perpetrated by a social insider or may not target a victim's social or personal information. Participants may perceive threats differently based on what information is being targeted. Secondly, we would be presenting dissimilar scenario details for social insider attack and other comparable threats; the stories we present provide a lot of detail to the social insider attack, however we would be unable to present similar levels of detail for other computer security threats. This makes comparing the two an uneven comparison for participants.

To minimize bias that may adversely affect the validity of the study, participants that who had taken part the earlier insider attack dimension study will not be

allowed to access and participate in the survey to this study as they may encounter their own story.

5.2.3 Future Work

As of the writing of this thesis, this study is not yet complete. We intend to add it as a note, or extension to publication when complete.

Chapter 6

Discussion

6.1 Discussion

Our results show that social insider attacks are common and occur in a variety of circumstances. They also suggest that the typical Facebook user is likely to prioritize usability over security of their account. With the results of our studies we can now address the questions we posed in the introduction.

Attacks are common. From our prevalence estimates in 3.5.1, a sizable fraction of Facebook users seem to have been involved in instances of social insider attacks. The high prevalence of attacks demonstrates a need for effective mechanisms to detect and report these attacks to account owners. In the 45 stories we collected there were numerous instances where the perpetrator accessed the victim's account because either the device or the Facebook account was already unlocked. If users had logged out of their accounts, or locked their devices, those attacks would not have been possible. However, we know that we cannot expect users to choose security if there is a substantial usability cost [18]. Thus, existing secret-based authentication mechanisms are unlikely to be effective at countering a social insider threat.

Attacks are opportunistic and have a variety of motives. The range of collected stories reveals that the threat of social insider attacks is a phenomenon that encompasses a range of motives, with a broad set of relationships, attack vectors and variants, and with significant consequences for the parties involved. The at-

tacker's motive often, but not always, determines the attack characteristics. Most attacks are opportunistic, and multiple stories indicated an attacker struggling, and failing, to control the urge to carry out the attack. For victims, the stories highlighted a high emotional and practical toll of the attack. This hints at a mismatch between the degree to which Facebook users value privacy, and their ability (or desire) to attain this privacy.

Mitigating these attacks will require a coordinated approach. We believe that several complementary approaches are necessary to mitigate social insider attacks and inform account owners when their data may have been compromised, as using few techniques would limit their effective coverage radius and make them very narrow.

- **Education.** In many stories the victim adopted better security practices *after the attack*. Educating users about the social insider threat might motivate them to adopt more secure practices, such as signing out of their account.

- **Visible logging.** A possible technical solution is more visible logging of user activities on Facebook. This approach may be effective for snooping attacks, in which any Facebook usage would leave a non-reputable trace of activity. The account owner could access this log to verify actions that they did and did not perform. This would help victims with attack deterrence, detection, and investigation. The log, however, would be different from Facebook's current and limited 'activity log,' which only captures write events such as comments and posts. Limitations of such a technique that discovery would only be after the fact and the account owner would have to regularly check for activity. Making such a log effective in the hands of diverse users could be an interesting subject of future research.

- **Continuous authentication.** Another technical approach is continuous authentication [7, 33]. The OSN (or a trusted third party) can construct a profile of the user's actions by persistently analyzing them and if they do not match their expected behavior according to the previously generated profile, an alarm can be raised. Nonetheless, this solution does come with multiple trade-offs. Continuous authentication has the benefit of running silently and

does not impose an added usability overhead while still being able to detect unauthorized usage of the account. Furthermore, apart from context specific action such as those performed while using the OSN service, multiple alternative sources of data can be used to train anomaly detection systems as well. Some of these includes geo-location tagging and biometric interaction with input output devices such computer mouse movement, typing styles on keyboards and touch screen interactions. However, as the technology currenty stands today, making continuous authentication robust however can be challenging due to erratic or difficult-to-predict usage of the account by the legitimate account owner which can greatly increase the number of false positives reported and the false negatives unreported by the system.

6.1.1 Limitations

Our findings are not without limitations, most of which stem from our study design choices. We recruited study participants that reside in US and our findings reflect US culture. Our results may not generalize to the worldwide Facebook user population. As an extension to our study, it would be interesting to see if this 'snooping' culture exists as a world-wide phenomenon or varies between geographical or socio-economic divisions. Our prevalence results apply to a broad range of Facebook social insider attacks. But, as our second study suggests, there is substantial variation in these attacks. For example, some attacks are considered harmful while others are perceived as benign. Because our second study was qualitative, we were not able to estimate the prevalence of each kind of attack.

The reported studies are also subject to the limitations of their respective research methods. The first study was a list experiment and its results depend on the assumption that respondents were truthful. The second study uses self reporting and may have blind spots, either because the participant sample was not diverse, or because people may not be willing to report certain attack incidents.

The extent to which this research applies to other OSNs is also unclear. There is indication that accounts on other OSNs, such as Twitter, are also targets of social insider attacks [38]. The stories in our second study often noted that the attacker considered the victim's Facebook account as a *reliable* source of information. It

seems that as long as OSN accounts contain information a perpetrator would conceive as valuable and reliable, the threat of digital and social insider will exist. This suggests that our findings may not be unique to Facebook. A recent trend in online social networks is to provide added guarantees about the security of personal data. Snapchat and Cyberdust are examples of networks that either do not store private content or store it for a limited amount of time. These networks raise questions about social insider attacks – are they still feasible on such networks? And, how far would attackers go to gain access to data on such networks?

6.1.2 Ethics

Our studies were approved by our institutional research ethics board [details blinded]. We also provided a feedback form at the end of each study to allow participants to express their concerns. Two participants in the second study expressed discomfort towards recalling negative experiences. One indicated feeling *"a little anxiety from the story itself but that was expected"*, and the other said to have *"had a bad experience and dredging it up [...] bothered me"*. On the other hand, one participant reported *"actually enjoyed venting about this"*. We believe that researchers considering studies in this space should strive to further improve informed participant consent about the harm/benefits trade-off.

6.2 Conclusion

Online social networks contain a wealth of personal information. Information that may be hidden from and valuable to close contacts, such as spouses and friends. In this paper we studied the prevalence and the factors surrounding *social insider attacks* against Facebook accounts. Using the anonymous list experiment method we determined that these attacks are widespread: 24% of participants perpetrated such an attack and 21% were victims of this attack. We solicited anonymous stories describing episodes of a social insider attack and then used thematic analysis to understand the salient dimensions. We found that these attacks target a variety of victim information, have a broad range of motives, are predominantly opportunistic, and have severe emotional consequences for victims. An implication of our analysis is that the existing device and Facebook account security measures appear

to be ineffective in countering the social insider threat.

Bibliography

[1] E. R. Associates. Baseline, online probability survey of internet users regarding cyber security, 2011. URL http://www.ekospolitics.com/articles/032-11.pdf. → pages 3, 27

[2] K. Aytes and T. Conolly. A research model for investigating human behavior related to computer security. *AMCIS 2003 Proceedings*, page 260, 2003. → pages 40

[3] G. Blair and K. Imai. Statistical Analysis of List Experiments. *Political Analysis*, 20(1):47–77, Jan. 2012. ISSN 1047-1987. doi:10.1093/pan/mpr048. → pages 3, 11, 12, 17, 22, 23

[4] A. Braunstein, L. Granka, and J. Staddon. Indirect content privacy surveys: Measuring privacy without asking about it. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, SOUPS '11, pages 15:1–15:14, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0911-0. doi:10.1145/2078827.2078847. URL http://doi.acm.org/10.1145/2078827.2078847. → pages 9

[5] E. Bursztein, B. Benko, D. Margolis, T. Pietraszek, A. Archer, A. Aquino, A. Pitsillidis, and S. Savage. Handcrafted fraud and extortion: Manual account hijacking in the wild. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, IMC '14, pages 347–358, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-3213-2. doi:10.1145/2663716.2663749. URL http://doi.acm.org/10.1145/2663716.2663749. → pages 5

[6] A. Chaudhuri and T. Christofides. *Indirect Questioning in Sample Surveys*. Springer, 2013. → pages 9

[7] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann. Touch me once and i know it's you!: Implicit authentication based on touch screen

patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 987–996, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi:10.1145/2207676.2208544. URL http://doi.acm.org/10.1145/2207676.2208544. → pages 46

[8] U. Dictionary. Urban dictionary definition - frape, 2007. URL http://www.urbandictionary.com/define.php?term=Frape. Accessed: 2016-12-01. → pages 2

[9] U. Dictionary. Urban dictionary definition - facekjack, 2010. URL http://www.urbandictionary.com/define.php?term=Facejacking. Accessed: 2016-12-01. → pages 2

[10] J. A. Droitcour, E. M. Larson, U. General, et al. The three card method: Estimating sensitive survey items with permanent anonymity of response. In *in'Proceedings of the Social Statistics Section', American Statistical Association*. Citeseer, 2001. → pages 9

[11] S. Eberz, K. B. Rasmussen, V. Lenders, and I. Martinovic. Preventing lunchtime attacks: Fighting insider threats with eye movement biometrics. In *NDSS*, 2015. → pages 1

[12] J. Edwards, 2014. URL http://www.businessinsider.com/frape-facebook-rape-now-a-crime-2014-7. Accessed: 2016-12-01. → pages 2

[13] S. Egelman and E. Peer. Scaling the security wall: Developing a security behavior intentions scale (sebis). In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2873–2882. ACM, 2015. → pages 41

[14] G. Gavai, K. Sricharan, D. Gunning, R. Rolleston, J. Hanley, and M. Singhal. Detecting insider threat from enterprise social and online activity data. In *Proceedings of the 7th ACM CCS International Workshop on Managing Insider Security Threats*, MIST '15, pages 13–20, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3824-0. doi:10.1145/2808783.2808784. URL http://doi.acm.org/10.1145/2808783.2808784. → pages 6

[15] M. Gilens, P. M. Sniderman, and J. H. Kuklinski. Affirmative action and the politics of realignment. *British Journal of Political Science*, 28(01): 159–183, 1998. → pages viii, 11

[16] C. Graham and K. Mathis. Frape, Stalking and Whores: Semantics and social narrative on Facebook. In *Immersive Worlds and Transmedia Narratives 1st Global Conference*. Inter-Disciplinary.Net, 2012. URL http://www.inter-disciplinary.net/critical-issues/wp-content/uploads/2012/10/grahamtmpaper.pdf. → pages 7

[17] F. L. Greitzer and R. E. Hohimer. Modeling human behavior to anticipate insider attacks. *Journal of Strategic Security*, 4(2):25, 2011. → pages 2

[18] C. Herley. So long, and no thanks for the externalities: The rational rejection of security advice by users. In *Proceedings of the 2009 Workshop on New Security Paradigms Workshop*, NSPW '09, pages 133–144, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-845-2. doi:10.1145/1719030.1719050. URL http://doi.acm.org/10.1145/1719030.1719050. → pages 45

[19] M. Johnson, S. Egelman, and S. M. Bellovin. Facebook and privacy: It's complicated. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, pages 9:1–9:15, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1532-6. doi:10.1145/2335356.2335369. URL http://doi.acm.org/10.1145/2335356.2335369. → pages 7, 33

[20] B. Krishnamurthy and C. E. Wills. Characterizing privacy in online social networks. In *Proceedings of the First Workshop on Online Social Networks*, WOSN '08, pages 37–42, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-182-8. doi:10.1145/1397735.1397744. URL http://doi.acm.org/10.1145/1397735.1397744. → pages 1

[21] D. Marques, I. Muslukhov, T. Guerreiro, L. Carriço, and K. Beznosov. Snooping on mobile phones: Prevalence and trends. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 159–174, Denver, CO, June 2016. USENIX Association. ISBN 978-1-931971-31-7. URL https://www.usenix.org/conference/soups2016/technical-sessions/presentation/marques. → pages 7, 10, 12, 17, 22, 26

[22] A. Marwick and D. Boyd. 'It's just drama': teen perspectives on conflict and aggression in a networked era. *Journal of Youth Studies*, 17(9):1187–1204, Apr. 2014. ISSN 1367-6261. doi:10.1080/13676261.2014.901493. URL http://dx.doi.org/10.1080/13676261.2014.901493. → pages 7

[23] J. D. Miller. The nominative technique: A new method of estimating heroin prevalence. *NIDA Research Monograph*, 54:104–124, 1985. → pages 9

[24] W. Moncur, K. M. Orzech, and F. G. Neville. Fraping, social norms and online representations of self. *Computers in Human Behavior*, 63:125–131, 2016. ISSN 0747-5632. doi:http://dx.doi.org/10.1016/j.chb.2016.05.042. URL http://www.sciencedirect.com/science/article/pii/S0747563216303697. → pages 7

[25] I. Muslukhov, Y. Boshmaf, C. Kuo, J. Lester, and K. Beznosov. Know your enemy: The risk of unauthorized access in smartphones by insiders. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '13, pages 271–280, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2273-7. doi:10.1145/2493190.2493223. URL http://doi.acm.org/10.1145/2493190.2493223. → pages 7

[26] E. J. Pauwels and O. Ambekar. One class classification for anomaly detection: Support vector data description revisited. In *Proceedings of the 11th International Conference on Advances in Data Mining: Applications and Theoretical Aspects*, ICDM'11, pages 25–39, Berlin, Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-23183-4. URL http://dl.acm.org/citation.cfm?id=2033796.2033800. → pages 6

[27] E. Peer, J. Vosgerau, and A. Acquisti. Reputation as a sufficient condition for data quality on amazon mechanical turk. *Behavior Research Methods*, 46 (4):1023–1031, 2013. ISSN 1554-3528. doi:10.3758/s13428-013-0434-y. URL http://dx.doi.org/10.3758/s13428-013-0434-y. → pages 11

[28] Pew Research Center. Anonymity, Privacy, and Security Online. Report, 2013. URL http://pewinternet.org/Reports/2013/Anonymity-online.aspx. → pages 5

[29] Pew Research Center. The Demographics of Social Media Users. Report, 2015. URL http://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/. → pages 11

[30] D. Raghavarao and W. T. Federer. Block total response as an alternative to the randomized response method in surveys. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 40–45, 1979. URL https://www.jstor.org/stable/2984720. → pages 3, 10

[31] B. Rosenfeld, K. Imai, and J. N. Shapiro. An empirical validation study of popular survey methodologies for sensitive questions. *American Journal of Political Science*, 2015. → pages 10

[32] R. Shay, I. Ion, R. W. Reeder, and S. Consolvo. "my religious aunt asked why i was trying to sell her viagra": Experiences with account hijacking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 2657–2666, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1. doi:10.1145/2556288.2557330. URL http://doi.acm.org/10.1145/2556288.2557330. → pages 5, 26

[33] S. J. Shepherd. Continuous authentication by analysis of keyboard typing characteristics. In *Security and Detection, 1995., European Convention on*, pages 111–114, May 1995. doi:10.1049/cp:19950480. → pages 46

[34] T. Spiliotopoulos and I. Oakley. Understanding motivations for facebook use: Usage metrics, network structure, and privacy. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 3287–3296, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1899-0. doi:10.1145/2470654.2466449. URL http://doi.acm.org/10.1145/2470654.2466449. → pages 14

[35] Statista. Facebook: U.s. user age distribution 2016 — statistic. Report, 2016. URL https://www.statista.com/statistics/187041/us-user-age-distribution-on-facebook/. Accessed: 2016-12-01. → pages 1, 20

[36] Statista. Facebook u.s. user gender share 2016 — statistic. Report, 2016. URL https://www.statista.com/statistics/266879/facebook-users-in-the-us-by-gender/. Accessed: 2016-12-01. → pages 20

[37] R. Tourangeau and T. Yan. Sensitive questions in surveys. *Psychological bulletin*, 133(5):859, 2007. URL http://eric.ed.gov/?id=EJ774165. → pages 7

[38] L. Vaas. Wikipedia co-founder jimmy wales' twitter account hijacked - naked security, August 2016. URL https://nakedsecurity.sophos.com/2016/08/23/wikipedia-co-founder-jimmy-wales-twitter-account-hijacked/. Accessed: 2016-12-01. → pages 47

[39] B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Towards detecting anomalous user behavior in online social networks. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 223–238, San Diego, CA, Aug. 2014. USENIX Association. ISBN 978-1-931971-15-7. URL

https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/viswanath. → pages 6

Appendix A

Supporting Materials

Survey for Item Selection for the Study on the Prevalence of Physical Insider Attacks

How old are you?

[ ]

What is your gender?

- ○ Male
- ○ Female
- ○ Other

What is your highest level of education

- ○ High School
- ○ College Degree
- ○ Graduate School
- ○ Other _____

In which state do you reside?

... State choices...

Which of the following social networking sites do you use?

- ☐ Facebook
- ☐ Twitter
- ☐ Reddit
- ☐ Pinterest
- ☐ Tumblr
- ☐ LinkedIn
- ☐ Other _____

How much time do you spend on social networking sites per day?

- ○ Less than 30 minutes
- ○ 30 min - 1 hour
- ○ More 1 hour

Please check all statements that apply to you.

- ☐ I have posted a message in a group on Facebook and received a reply
- ☐ Someone I know has posted content on my Facebook wall
- ☐ I have received 5 or more unsolicited messages from strangers on Facebook
- ☐ One of my relatives has sent me a friend request on Facebook
- ☐ I have posted a picture of myself on Facebook
- ☐ Someone liked one of the pictures I posted on Facebook
- ☐ I have more than 300 friends on Facebook
- ☐ I am friends with one of my parents on Facebook
- ☐ I check Facebook every day
- ☐ On average, I spend more than 30 minutes on Facebook every day
- ☐ I have changed my Facebook profile picture in the last 12 months
- ☐ In the last week, I have clicked on a link posted on my Facebook newsfeed
- ☐ I have commented or liked a post in the last month on Facebook
- ☐ I am a member of a Facebook group
- ☐ In the last week, I have checked Facebook while at work
- ☐ I have reported an account on Facebook
- ☐ I re-shared someone's post on Facebook
- ☐ I have made my birth date publicly visible on Facebook
- ☐ I have clicked on an advertisement on Facebook
- ☐ I have responded to an event invitation on Facebook
- ☐ I have used a device of someone I know to access their Facebook account without permission
- ☐ Somebody I know has used my device to access my Facebook account without permission

List Experiment Survey for the Study on the Prevalence of Physical Insider Attacks - Control

<redacted>

Consent decision:

○ Yes

○ No

Below is a list of statements that describe various experiences that you may have encountered in the past year. To preserve your anonymity, select HOW MANY statements that apply to you, not WHICH ONES.

I have more than 300 friends on Facebook.

I am friends with one of my parents on Facebook.

I have commented or liked a post in the last month on Facebook.

I have reported an account on Facebook.

I have had dinner with the founder of Facebook, Mark Zuckerberg.

	0 (None)	1	2	3	4	5 (All)
Statements that apply to you	○	○	○	○	○	○

How old are you?

[ ]

What is your gender?

- ○ Male
- ○ Female
- ○ Other

What is your highest level of education

- ○ High School
- ○ College Degree
- ○ Graduate School
- ○ Other _____

What country do you reside in?

... choices hidden ...

Which state do you reside?

... choices hidden ...

Which of the following social networking sites do you use?

- ☐ Facebook
- ☐ Twitter
- ☐ Reddit
- ☐ Pinterest
- ☐ Tumblr
- ☐ LinkedIn
- ☐ Other _____
- ☐ None

List Experiment Survey for the Study on the Prevalence of Physical Insider Attacks – Treatment 1

<redacted> Consent decision:

○ Yes

○ No

Below is a list of statements that describe various experiences that you may have encountered in the past year. To preserve your anonymity, select HOW MANY statements that apply to you, not WHICH ONES.

I have more than 300 friends on Facebook.

I am friends with one of my parents on Facebook.

I have commented or liked a post in the last month on Facebook.

I have reported an account on Facebook.

I have had dinner with the founder of Facebook, Mark Zuckerberg.

I have used a device of someone I know to access their Facebook account without permission.

	None (0)	1	2	3	4	5	6 (All)
Statements that apply to you	○	○	○	○	○	○	○

How old are you?

What is your gender?

- ○ Male
- ○ Female
- ○ Other

What is your highest level of education

- ○ High School
- ○ College Degree
- ○ Graduate School
- ○ Other _____

What country do you reside in?

... choices hidden ...

Which state do you reside?

... choices hidden ...

Which of the following social networking sites do you use?

- ☐ Facebook
- ☐ Twitter
- ☐ Reddit
- ☐ Pinterest
- ☐ Tumblr
- ☐ LinkedIn
- ☐ Other _____
- ☐ None

List Experiment Survey for the Study on the Prevalence of Physical Insider Attacks – Treatment 2

<redacted> Consent decision:

- ○ Yes
- ○ No

Below is a list of statements that describe various experiences that you may have encountered in the past year. To preserve your anonymity, select HOW MANY statements that apply to you, not WHICH ONES.

I have more than 300 friends on Facebook.

I am friends with one of my parents on Facebook.

I have commented or liked a post in the last month on Facebook.

I have reported an account on Facebook.

Somebody I know has used my device to access my Facebook account without permission.

I have had dinner with the founder of Facebook, Mark Zuckerberg.

	0 (None)	1	2	3	4	5	6 (All)
Statements that apply to you	○	○	○	○	○	○	○

How old are you?

What is your gender?

- ○ Male
- ○ Female
- ○ Other

What is your highest level of education

- ○ High School
- ○ College Degree
- ○ Graduate School
- ○ Other _____

What country do you reside in?

... choices hidden ...

Which state do you reside?

... choices hidden ...

Which of the following social networking sites do you use?

- ☐ Facebook
- ☐ Twitter
- ☐ Reddit
- ☐ Pinterest
- ☐ Tumblr
- ☐ LinkedIn
- ☐ Other _____
- ☐ None

Survey for the Study on The Dimensions of Physical Insider Attacks

Section 1 : Consent

Consent

<redacted>

- ○ Yes
- ○ No

Section 2 : Information I

Answer yes below if you have experienced a situation that satisfied all three of the following conditions: You either accessed someone else's Facebook account without permission, or had your Facebook account accessed without permission, and You and the other party knew one another, and The Facebook account holder regularly controlled the device on which the Facebook account was accessed.

- ○ Yes
- ○ No

Section 3 : Demographics

How old are you?

[ ]

What is your gender?

- ○ Male
- ○ Female

- ○ Other

Which state do you reside in?

... 29 additional choices hidden ...

What is your highest level of education?

- ○ High School
- ○ College Degree
- ○ Graduate School
- ○ Other _____

Section 4 : Writing Task

Tell us a story from your own experience

Recall a situation where you have either used a device of someone you know to access their Facebook account without their permission, or someone you know has used your device to access your Facebook account without your permission. If you can recall more than one such experience, please consider the one that you think is less common.

Your task is to write a story describing that situation, giving enough detail so that a person who doesn't know the people involved in the story would understand it.

Do not use real names or any personally-identifiable information. Instead, use: Alex to refer to the account holder, and Casey to refer to the person who accessed the account. If there are other characters in your story, use fictional names for them as well. To maintain anonymity, use gender-neutral pronouns such as 'they' instead of 'he' or 'she' or refer to the character by their fictional name. Your story should include details, such as:

Where did the situation take place and when?

What were the relationships among the people and how well did they know each other?

How did Casey come to have access to Alex's device?

What difficulties, if any, did Casey face in gaining access?

What did Casey end up doing with access to Alex's Facebook account?

For how long did Casey have access to Alex's account?

What were Casey's motivations and objectives?