Characterizing Social Insider Attacks on Facebook

Wali Ahmed Usmani[1], **Diogo Marques**[2], **Ivan Beschastnikh**[1],
Konstantin Beznosov[3], **Tiago Guerreiro**[2], **Luís Carriço**[2]
[1] NSS, Department of Computer Science, University of British Columbia
[2] LaSIGE, Faculdade de Ciências, Universidade de Lisboa
[3]LERSSE, Department of Electrical and Computer Engineering, University of British Columbia
[wusmani, bestchai]@cs.ubc.ca, [dmarques, tjvg, lmc]@di.fc.ul.pt, beznosov@ece.ubc.ca

ABSTRACT
Facebook accounts are secured against unauthorized access through passwords and device-level security. Those defenses, however, may not be sufficient to prevent *social insider attacks*, where attackers know their victims, and gain access to a victim's account by interacting directly with their device. To characterize these attacks, we ran two MTurk studies. In the first (n = 1,308), using the list experiment method, we estimated that 24% of participants had perpetrated social insider attacks and that 21% had been victims (and knew about it). In the second study (n = 45), participants wrote stories detailing personal experiences with such attacks. Using thematic analysis, we typified attacks around five motivations (fun, curiosity, jealousy, animosity, and utility), and explored dimensions associated with each type. Our combined findings indicate that social insider attacks are common, often have serious emotional consequences, and have no simple mitigation.

ACM Classification Keywords
K.6.m Security and Protection: Miscellaneous

Author Keywords
Usable security; privacy; insider attack; Facebook

INTRODUCTION
Facebook users often share and maintain personal and potentially sensitive information on their accounts, including messages, pictures and videos [10]. This information can entice adversaries to try to obtain it without the owner's consent. Adversaries who are *social insiders* have a social relationship with the account owner and are of special concern. The proximity between the victim and a social insider makes it easier for the insider to obtain unauthorized access to the victim's device and Facebook account.

In this paper we focus on *social insider attacks* on Facebook, which is when an insider accesses the Facebook account of a victim using Facebook's end-user interfaces, like the web or a mobile application, on the victim's device, and without the victim's permission. We consider a victim's device to be one that is regularly controlled by the victim. This includes not only personal devices, but also work computers and 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' [4, 5]) is often dismissed as a prank. However, some of these pranks have been regarded as defacement and resulted in criminal prosecution [6].

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 attacks are very rare, or of little consequence? And if they are not rare, how could effective defenses be designed against the spectrum of attacks that might exist if this spectrum is not well understood?

This paper helps to bridge our gaps in knowledge by quantitatively and qualitatively characterizing social insider attacks against Facebook accounts.

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 [17, 1]. 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 U.S. adult participants who reported being Facebook users, and found that social insider attacks are common. We estimated that 24% of participants carried out Facebook social insider attacks, and that 21% were knowing victims.

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, anonymous stories about past experiences with social insider attacks, and used thematic analysis to extract salient dimensions. We report on

CHI 2017, May 6-11, 2017, Denver, CO, USA.

ACM ISBN 978-1-4503-4655-9/17/05 ...$15.00.
DOI: http://dx.doi.org/10.1145/3025453.3025901

several themes that emerged from our analysis, including the relationships between perpetrators and victims, attack vectors, and the role of premeditation. We further organize social insider attacks on Facebook accounts by the types of motivation, and discuss how attacks tend to unfold.

To summarize, this paper addresses the following questions:

- How common are social insider attacks against Facebook accounts? Who is more likely to be a victim, and who is more likely to be a perpetrator?
- What are the salient dimensions in these attacks? Why do people conduct attacks, how do they conduct them, and what are the practical and emotional repercussions?
- What are the security implications of these attacks?

Our findings suggest that attacks are common, opportunistic, and have a range of motives, including fun, curiosity, jealousy, animosity, and utility. Considering the diversity of these attacks, we believe that no single mitigation strategy would be effective.

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 [15]. 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 better understood. For instance, the main characteristics of manual "hijacking" on Google accounts have been studied [3] 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 Online Social Networks (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 [18]. Although this study did not exclude insider perpetrators, only five 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 significant emotional consequence of being a victim of an attack.

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 restricted to younger people, considered a practical joke, and even to have some positive effects, as a factor of in-group bonding [13]. However, fraping may sometimes be interpreted as a form of cyberbullying and online hate speech [7]. As in the case of younger people using the word "drama" to refer to some online interactions that adults would classify as bullying, using the word "frape" may allow for some ambiguity between serious and frivolous attacks, as a way to avoid framing incidents as instances of victimization [12].

Research on privacy perceptions of Facebook users suggests that there is particular concern with unauthorized insider access to information. In a 2012 study with 260 people, 86% of participants were either not very concerned with the threat of strangers on Facebook, or used Facebook's privacy settings to mitigate these concerns [9]. By contrast, 37% of participants were concerned about people in their friends list viewing their profile or social content. While our work also aims to understand activities of insiders, we focus on social insiders rather than Facebook friends.

A 2013 investigation on concerns over social insider attacks on smartphones found that users are aware of the insider threat [14]. The study included a survey of 724 participants, of which 12% reported having experienced unauthorized data access and 9% reported having 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 [21]. A recent list experiment study of snooping attacks (the subset of smartphone social insider attacks in which the objective is limited to snooping), found a much higher prevalence, with an estimated 30% of participants having been perpetrators in a one-year period [11]. Our research differs by focusing on Facebook instead of sensitive data on smartphones, by including all platforms by which a social insider can conduct an attack, such as desktop computers, laptops and tablets, and by broadening the scope of attacks beyond snooping.

PREVALENCE OF SOCIAL INSIDER ATTACKS

A security threat is of general interest if it is both probable to materialize, and harmful when it does materialize. In our first study, we aimed to understand how common (or uncommon) social insider attacks on Facebook are now, as a proxy for their future occurrence, all other factors remaining equal.

Obtaining such measurements reliably, however, is challenging. Asking users if they are victims or perpetrators of attacks is likely to lead to underestimation. Victims may be unaware of or unwilling to report intrusions. Perpetrators may not want to self-incriminate or may be biased towards socially-desirable answers by language associated with privacy and security [2].

To address these challenges, we used the *list experiment* technique [17]. We closely followed the method in Marques et al.'s [11] recent study of snooping on smartphones, as the type of attacks in that study was similar: both attacks involve unau-

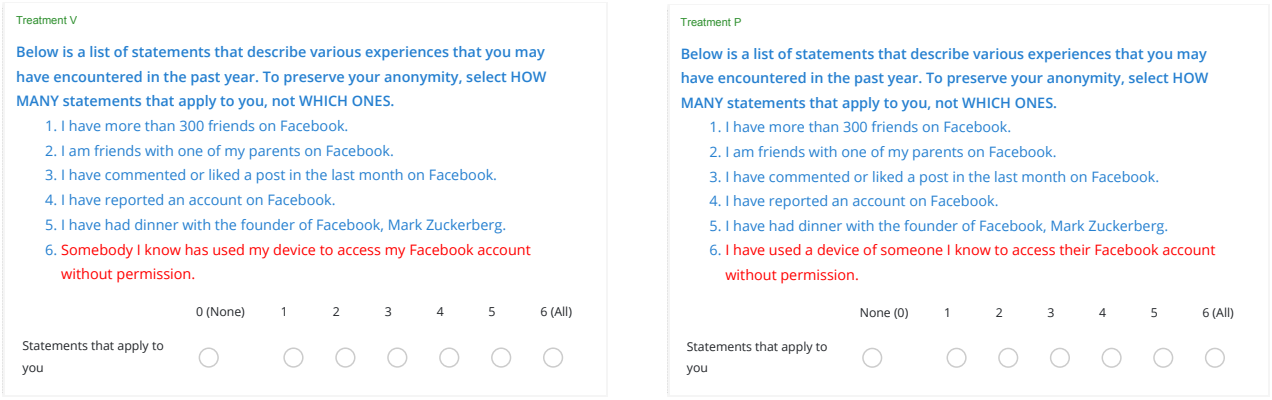

Figure 1. List question administered in list experiment, including 4 control items selected to minimize for ceiling and floor effects, 1 attention check item, and 2 treatment items (highlighted in red only for the manuscript), each administered to a separate treatment group. The control group did not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

thorized physical access to devices, and perpetrators in both attacks are likely to be social insiders.

In a list experiment, participants are randomly split into a *control* group and a *treatment* group. Participants are presented with a *list question*, a set of items, typically formulated as statements, and a prompt to indicate *how many* statements they agree with (though not which ones). The list questions presented to the control and treatment groups are similar, both containing a set of *control items*, statements that are of no interest to the research question. However, 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 can be used to estimate the proportion of participants who selected the treatment item.

Unlike Marques et al., in our work we used two treatment groups. One group was shown a treatment item which identified the participant as having been a victim of a social insider attack, while the other identified the participant as a perpetrator. The difference between those estimates was expected to offer some insight into how common it is for people to be unaware they been victims of a social insider attack.

Our study targeted the U.S. Facebook users population, since the adoption rate of Facebook among U.S. adults is high; according to a 2014 Pew survey, 62% use Facebook [16]. This made it easy to find Facebook users among U.S. MTurk workers.

Item selection

An important design consideration in list experiments is the composition of the list question. Common recommendations when building list questions include: avoiding floor and ceiling effects (many participants identifying with none or all statements in the list), avoiding lists that are too long or too short, and avoiding items that stand out in relation to the others [11]. We used four control items with statements related Facebook usage, and two neutrally-worded treatment items.

Treatment items

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. After multiple iterations we 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 avoid biasing 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.

Control items

To select four 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 four statements for which participants would rarely agree with 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 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 check-box items with the prompt “Please check all statements that apply to you”. Workers were paid $0.20 for completing the survey. Only workers with location set to U.S. were allowed to participate. At the beginning of the survey, a filter based on IP addresses further prevented participation from non-U.S. locations.

The statements in the check-box question were twenty candidate control items, drawn from previous research on motivations for Facebook use [19] and common Facebook use cases developed by the research team in brainstorming sessions. We

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

Table 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.

also included the two 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.

Results of Item Selection Survey

We collected 202 complete responses, and excluded 28 that either indicated that participants did not use Facebook, or were given in less than 40 seconds (based on a prior pilot with five 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 1 shows the percentage and number of respondents who checked each statement.

To select the control items, we computed all possible combinations of four statements 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 1, 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 through direct questioning were less than half of 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 and 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, for example 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.

List Experiment Study Procedure

For the list experiment study, we again recruited among U.S. MTurk workers and limited participation to those who were accessing our survey server from U.S. IP addresses. As before, we avoided words like "privacy" or "attack" in the task advertisement and consent form, informing participants 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). 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"). Figure 1 lists the finalized list question. Other than the list experiment, the survey contained questions on participant age, gender, level of education, U.S. state of residence, and OSNs which participants used. Each worker was paid $0.20 for completing the survey.

Dataset

Data clean-up

We received 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).
- 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.

This left us with 1,308 responses, on which the following analysis is based.

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 OSNs (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 OSNs 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 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 age and gender variables between our survey sample and the Facebook population. Despite our sample being demographically similar to MTurk samples in other studies, it was younger and slightly skewed to males as compared to the U.S. Facebook user population in general [20].

Results

Prevalence Estimates

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, are thus:

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

Effects of Age and OSN Participation

Marques et al. [11] found evidence that snooping on mobile phones was more prevalent among younger people, and among those who adopted smartphones more deeply (used their own phone in a way that retained more private data). To verify if similar effects exist in social insider attacks on Facebook, we ran list experiment regression models [1] on the age variable, and lacking a specific measure of depth of adoption, on the number of OSNs that participants reported using.

Figures 2 and 3 depict those regression models graphically. Regarding age, there is a visible pattern of decreasing likelihood of being a perpetrator of social insider attacks as age increases. However, for the likelihood of being a victim, the dependency on age is less pronounced and nearly flat.

For the number of used OSNs, the opposite seems to be true. Using more OSNs was a weak predictor of being a perpetrator. Rather, using more OSNs at best 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.

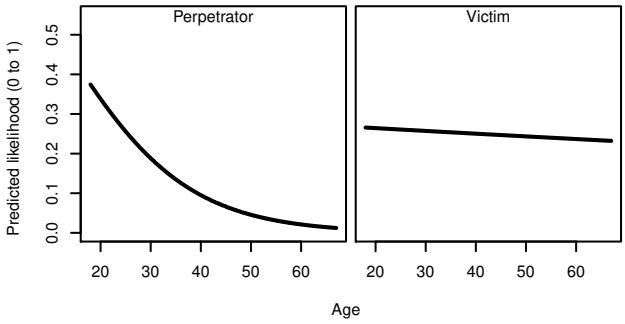

Figure 2. Regression model of likelihood of being a perpetrator, or a knowing victim of social insider attacks on Facebook, predicted by age of participants.

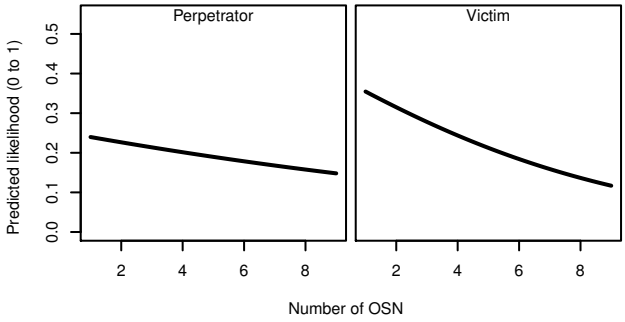

Figure 3. Regression model of likelihood of being a perpetrator, or a knowing victim of social insider attacks on Facebook, predicted by age number of OSNs participants used.

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 [1] 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 interaction 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 4. 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 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).

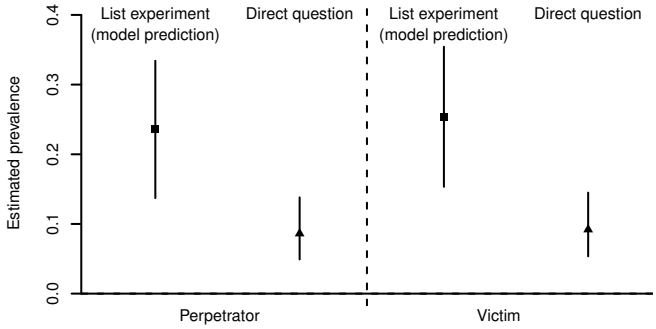

Figure 4. Estimated prevalences based on list experiment model predictions and response to direct questions. Predictions, and 95% confidence interval of predictions, from a list experiment regression model of age, number of OSNs participants used, and the interaction between the two terms. Proportion of positive response to direct questions, and 95% confidence intervals, from the item selection survey (n = 174).

Discussion

The main objective of this study was to estimate the prevalence of social insider attacks on Facebook. The results suggest that they are not uncommon, with 24% of participants estimated to have implicitly identified with the statement "I have used 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".

Contrasting the estimates obtained through the list experiment (24%/21%) with the ones obtained through direct questioning (9%/9%), 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 [11]. For victims of social insider attacks, the effect was more surprising, and could potentially be related to victims assigning themselves responsibility for intrusions [18].

The regression models we fit also indicate two clear trends. First, younger people are more likely to conduct social insider attacks, mirroring prior findings on mobile phone snooping [11]. Second, people who use more OSNs are less likely to be victims. One possible explanation for this trend is that those who use more OSNs tend to be more tech savvy and more aware of what private information is retained on OSNs, and thus are more able and motivated to protect themselves.

From a security perspective, these findings suggest that the probability of social insider attacks on Facebook is not negligible. While understanding attack probability is important, so is understanding the severity of the threat. The study reported in the next section provides insight into this issue.

DIMENSIONS OF SOCIAL INSIDER ATTACKS

We established that social insider attacks are common but what exactly constitutes a social insider attack? In the next study we sought to establish what it means to conduct a social insider attack, what the attacks looked like, why they took place, how they happened, and what the 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 influence, affect, and pertain to social insider attacks.

Methodology

We collected qualitative data through an online survey in which we asked participants to report on social insider attacks in which they were either the perpetrator or the victim. This survey was deployed on Amazon Mechanical Turk. It included a consent form and qualification and demographic questions to ensure that participants were eligible for participation. The main eligibility criteria was having perpetrated or been 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 twelve months. As before, we chose to focus on U.S. participants only; thus being geographically located within the U.S. was required in order to accept the task. Following the consent form and opening questions was an open-ended question asking participants to write a story about a past experience with a social insider attack on Facebook.

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 perpetrated the social insider attack, and *Alex* as the target of the attack.

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,and motivation). A batch of ten 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 U.S. census regions. We provided above average compensation of $4 and offered a bonus of $1 if the story was well-written as an incentive.

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.

Perpetrators and victims

The stories noted a variety of perpetrator-victim relationships. The variability in social and social 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. 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 had it in their mind that they were the parent, they had full right to access Alex's personal computer and their Facebook account. [Story 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.

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 victims' perceptions of the security of their accounts and the true degree to which their accounts were exposed to those in their social inner circles, indicating that both security measures and how people innocently create breaches are opening vectors for attacks on their privacy.

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 news-feed, 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"*). We also categorized attacks in terms of how long they lasted, the time of day they took place, and the number of sessions used in the attack. Attack length was diverse with some attacks lasting a few seconds, to others lasting many hours. Most attacks occurred in a single instance of account access, but some were spread over multiple sessions.

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 challeng-

ing 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]

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. 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 [9] 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, some 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.

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]

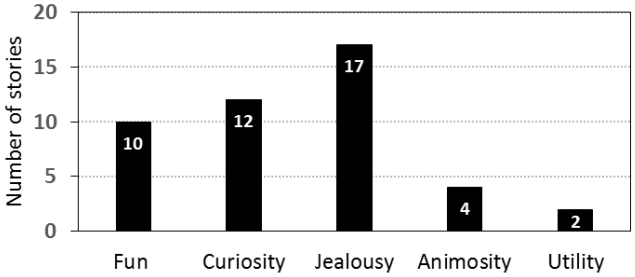

Figure 5. Distribution of attacks by their motivation category.

Overall, the consequences of an attack for both parties were predominantly profound and harsh. These events are likely to affect relationships and emotions deeply. People also tend to improve their security measures upon discovering an attack, suggesting that (1) they were not aware of or discredited the insider threat, and that (2) they were willing to improve their security, at the cost of convenience, once they became aware of the insider threat.

Motivation

We observed five types of motives: fun, curiosity, jealousy, animosity, and utility. Figure 5 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.

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]

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 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.

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.

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. Impersonation 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.

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]

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. 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 [8]. 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 attacker'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.

No obvious mitigation. We believe that there is no single mitigation to social insider attacks. Defending against an attacker who has social standing with the victim and who gains physical access to a device will require a combination of technical defenses, legal deterrents, and user education. For example, in many stories the victim adopted better security practices after the attack. This indicates that educating users about the social insider threat might motivate them to adopt more secure practices. On the technical side, Facebook has little support for logging passive account activity, such as browsing of message history. Many attacks in our stories could be easily identified with such an append-only log. The existence of such a log would also act as a deterrent, as many attackers in our stories noted that they carried through with the attack because it would not be discovered.

Limitations

Our findings are not without limitations, most of which stem from our study design choices. We recruited study participants that reside in U.S. and our findings reflect U.S. culture and may not generalize to the worldwide Facebook user population.

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.

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 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 [22]. The stories in our second study often noted that the attacker considered the victim's Facebook account as a *reliable* source of information. This suggests that our findings may not be unique to Facebook.

Ethics

Our studies were approved by our institutional research ethics board Behavioural Research Ethics Board of the University of British Columbia. We 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 at recalling negative experiences; one noted *"a little anxiety from the story itself but that was expected"*, while the other described having *"had a bad experience and dredging it up [...] bothered me"*. We believe that researchers considering studies in this space should strive to further improve informed participant consent about the harm/benefits trade-off.

CONCLUSION

Online social networks contain a wealth of personal information.This information 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 at times 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.

Reproducibility

Materials, analysis code, and limited data available at:
https://bestchai.github.io/social-insider-study/

Acknowledgments

This project has been funded by the Office of the Privacy Commissioner of Canada (OPC); the views expressed herein are those of the authors and do not necessarily reflect those of the OPC. This work was partially supported by FCT through funding of a PhD studentship SFRH/BD/98527/2013, and of the LaSIGE Research Unit, ref. UID/CEC/00408/2013.

REFERENCES

1. Graeme Blair and Kosuke Imai. 2012. Statistical Analysis of List Experiments. *Political Analysis* 20, 1 (Jan. 2012), 47–77. DOI:http://dx.doi.org/10.1093/pan/mpr048
2. Alex Braunstein, Laura Granka, and Jessica Staddon. 2011. Indirect Content Privacy Surveys: Measuring Privacy Without Asking About It. In *Proceedings of the Seventh Symposium on Usable Privacy and Security (SOUPS)*. ACM, New York, NY, USA, Article 15, 14 pages. DOI:http://dx.doi.org/10.1145/2078827.2078847
3. Elie Bursztein, Borbala Benko, Daniel Margolis, Tadek Pietraszek, Andy Archer, Allan Aquino, Andreas Pitsillidis, and Stefan Savage. 2014. Handcrafted Fraud and Extortion: Manual Account Hijacking in the Wild. In *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC)*. ACM, New York, NY, USA, 347–358. DOI: http://dx.doi.org/10.1145/2663716.2663749
4. Urban Dictionary. 2007. Frape. (2007). http://www.urbandictionary.com/define.php?term=Frape
5. Urban Dictionary. 2010. Facejacking. (2010). http://www.urbandictionary.com/define.php?term=Facejacking
6. Jim Edwards. 2014. (2014). http://www.businessinsider.com/frape-facebook-rape-now-a-crime-2014-7
7. Caroline Graham and Krista Mathis. 2012. Frape, Stalking and Whores: Semantics and social narrative on Facebook. In *Immersive Worlds and Transmedia Narratives 1st Global Conference*. Inter-Disciplinary.Net. http://www.inter-disciplinary.net/critical-issues/wp-content/uploads/2012/10/grahamtmpaper.pdf
8. Cormac Herley. 2009. 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)*. ACM, New York, NY, USA, 133–144. DOI: http://dx.doi.org/10.1145/1719030.1719050
9. Maritza Johnson, Serge Egelman, and Steven M. Bellovin. 2012. Facebook and Privacy: It's Complicated. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS)*. ACM, New York, NY, USA, Article 9, 15 pages. DOI: http://dx.doi.org/10.1145/2335356.2335369
10. Balachander Krishnamurthy and Craig E. Wills. 2008. Characterizing Privacy in Online Social Networks. In *Proceedings of the First Workshop on Online Social Networks (WOSN '08)*. ACM, New York, NY, USA, 37–42. DOI:http://dx.doi.org/10.1145/1397735.1397744
11. Diogo Marques, Ildar Muslukhov, Tiago Guerreiro, Luís Carriço, and Konstantin Beznosov. 2016. Snooping on Mobile Phones: Prevalence and Trends. In *Twelfth Symposium on Usable Privacy and Security (SOUPS)*. USENIX Association, Denver, CO, 159–174. https://www.usenix.org/conference/soups2016/technical-sessions/presentation/marques
12. Alice Marwick and Danah Boyd. 2014. 'It's just drama': teen perspectives on conflict and aggression in a networked era. *Journal of Youth Studies* 17, 9 (April 2014), 1187–1204. DOI: http://dx.doi.org/10.1080/13676261.2014.901493
13. Wendy Moncur, Kathryn M Orzech, and Fergus G Neville. 2016. Fraping, social norms and online representations of self. *Computers in Human Behavior* 63 (2016), 125–131. DOI: http://dx.doi.org/10.1016/j.chb.2016.05.042
14. Ildar Muslukhov, Yazan Boshmaf, Cynthia Kuo, Jonathan Lester, and Konstantin Beznosov. 2013. 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)*. ACM, New York, NY, USA, 271–280. DOI: http://dx.doi.org/10.1145/2493190.2493223
15. Pew Research Center. 2013. Anonymity, Privacy, and Security Online. Report. (2013). http://pewinternet.org/Reports/2013/Anonymity-online.aspx
16. Pew Research Center. 2015. The Demographics of Social Media Users. Report. (2015). http://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/
17. Damaraju Raghavarao and Walter T Federer. 1979. Block total response as an alternative to the randomized response method in surveys. *Journal of the Royal Statistical Society. Series B (Methodological)* (1979), 40–45. https://www.jstor.org/stable/2984720
18. Richard Shay, Iulia Ion, Robert W. Reeder, and Sunny Consolvo. 2014. "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)*. ACM, New York, NY, USA, 2657–2666. DOI: http://dx.doi.org/10.1145/2556288.2557330
19. Tasos Spiliotopoulos and Ian Oakley. 2013. Understanding Motivations for Facebook Use: Usage Metrics, Network Structure, and Privacy. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM, New York, NY, USA, 3287–3296. DOI: http://dx.doi.org/10.1145/2470654.2466449
20. Statista. 2016. Facebook: U.S. user age distribution 2016 | Statistic. Report. (2016). https://www.statista.com/statistics/187041/us-user-age-distribution-on-facebook/
21. Roger Tourangeau and Ting Yan. 2007. Sensitive questions in surveys. *Psychological bulletin* 133, 5 (2007), 859. http://eric.ed.gov/?id=EJ774165
22. Lisa Vaas. 2016. Wikipedia co-founder Jimmy Wales' Twitter account hijacked - Naked Security. (August 2016). https://nakedsecurity.sophos.com/2016/08/23/wikipedia-co-founder-jimmy-wales-twitter-account-hijacked/