"Little Brothers Watching You:" Raising Awareness of Data Leaks on Smartphones

Rebecca Balebako
Carnegie Mellon University
Pittsburgh, PA, USA
balebako@cmu.edu

Jaeyeon Jung
Microsoft Research
Redmond, WA, USA
jjung@microsoft.com

Wei Lu
Microsoft
Redmond, WA, USA
weilu@microsoft.com

Lorrie Faith Cranor
Carnegie Mellon University
Pittsburgh, PA, USA
lorrie@cmu.edu

Carolyn Nguyen
Microsoft Technology and Policy Group
Redmond, WA, USA
cnguyen@microsoft.com

ABSTRACT

Today's smartphone applications expect users to make decisions about what information they are willing to share, but fail to provide sufficient feedback about which privacy-sensitive information is leaving the phone, as well as how frequently and with which entities it is being shared. Such feedback can improve users' understanding of potential privacy leakages through apps that collect information about them in an unexpected way. Through a qualitative lab study with 19 participants, we first discuss misconceptions that smartphone users currently have with respect to two popular game applications that frequently collect the phone's current location and share it with multiple third parties. To measure the gap between users' understanding and actual privacy leakages, we use two types of interfaces that we developed: just-in-time notifications that appear the moment data is shared and a visualization that summarizes the shared data. We then report on participants' perceived benefits and concerns regarding data sharing with smartphone applications after experiencing notifications and having viewed the visualization. We conclude with a discussion on how heightened awareness of users and usable controls can mitigate some of these concerns.

Keywords

Privacy, Smartphones, Android Permissions, Just-in-Time Notifications, Data Sharing, Usable Privacy, Mobile

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: Interfaces

Symposium on Usable Privacy and Security (SOUPS) 2013, July 24–26, 2013, Newcastle, UK.

General Terms

Human Factors; Design

1. INTRODUCTION

Users are concerned about protecting their privacy on smartphones. In a telephone survey of 1,203 US adults, most were as concerned about the privacy of data on their smartphone as on their home computers. The majority of these participants oppose practices in which applications collect their contacts. Forty-six percent felt that wireless providers who collected location should not store it at all, and an additional 28% thought it should be deleted within a year [19]. A separate telephone survey of 2,254 US adults found that 57% of all smartphone app users "have either uninstalled an app over concerns about having to share their personal information, or declined to install an app in the first place for similar reasons" [3]. These surveys show users have expectations of how their privacy-sensitive data should be treated on the smartphones.

Existing interfaces typically fail to make users aware of relevant aspects of data sharing, e.g., destination, frequency, and purpose of the sharing. Without this awareness, it is difficult for users to make informed and optimal decisions about data sharing from smartphones. For example, Android requires that users make decisions about granting data access permissions before they install an application. The user is asked to agree to data sharing before she is able to evaluate the benefits of the application itself. In comparison, the iPhone interface provides a dialog asking for permission to send location data or address book the first two times an application requests that information. These interfaces do not notify users of the frequency, destination, or purpose of data sharing. None of these systems provide overviews about the information leaving the phone so that users can compare applications and types of information sent in a clear summary. In a recent field study of 20 Android users, we found that participants were often surprised by apps' data collection in the background and the level of data sharing [12].

We use a term *privacy leakages* when referring to privacy-sensitive data being transmitted off the smartphone by applications in a way that is unexpected by the user. In this paper, we present a smartphone app, *Privacy Leaks* that aims to improve users' awareness of privacy leakages as they

occur on an Android phone. The prototype is built on the TaintDroid platform [7], informing users about the frequency and destination of data being shared by an application in two different ways: (a) a visualization of the amount and types of information shared, after the data has been shared; (b) just-in-time (JIT) notifications at the moment the information is shared. Using the prototype, this work explores the following three research questions:

- What are participants' pre-existing understandings of data sharing with smartphone applications?
- Can runtime feedback via notifications and visualizations of data sharing on smartphones reduce the gap between users' understanding and actual privacy leakages without creating annoyance or confusion?
- What design guidelines can be drawn from participant feedback to improve smartphone privacy interfaces?

To create a concrete context for data sharing, we used a role-playing technique in our 19-participant lab study. Participants were asked to play two popular smartphone games and select one to recommend to a friend or family member. This simple task was performed twice: first on a regular Android phone and second on a phone running our prototype. Through a semi-structured interview, we first examine participants' misconceptions about data sharing. We then examine reactions to our interface and changes in understanding, and finally we look at desired control over data sharing.

This paper makes two contributions. First, we find that some participants have a very limited understanding of data sharing by smartphone applications, yet have a strong desire to remain anonymous or to protect their children from potential harms. Lacking any consumer education or interfaces raising their awareness of privacy risks, these users would be left vulnerable. Second, we provide design guidelines to improve users' understanding of privacy leakages through just-in-time notifications and a summary visualization on the phone. However, improved awareness is only the first step toward helping smartphone users reduce privacy risks. We identify future research efforts to provide users with control over their data.

2. RELATED WORK

We first discuss prior work that explored users' understanding—or lack of it—of privacy and security risks of smartphone applications. We then describe work that designed tools to inform users about various security and privacy issues and to provide control over their data. We highlight how these previous studies influenced the design of our study method and the Privacy Leaks prototype.

2.1 User Understanding of Privacy & Security Risks of Smartphone Applications

Several studies demonstrate a lack of user understanding of privacy and security risks associated with installing smartphone applications. An Internet survey of 308 Android users and a laboratory study of 25 Android users found that only 17% paid attention to the permissions (including ones which grant an application access to privacy-sensitive data) when installing an application. They also found that only 3% of the Internet survey respondents demonstrated full comprehension of the permissions screen [9]. Kelley *et al.* reported that Android users found it difficult to understand the terms and wording of the Android permissions [13]. Our study goes deeper into this lack of understanding and discusses users' misconceptions about data sharing with two popular game applications using a role-play technique.

To examine expectations and perceptions of smartphone security, Chin *et al.* interviewed and surveyed 60 users on how they would choose applications to install on their smartphones. They found that few participants considered the privacy policies. Referrals from friends or family, or on-line referrals were the predominant ways that users discovered new applications for their smartphones. Price, popularity, and recommendations from friends were important parts of the decision about whether to install [4]. Recognizing the value of recommendations, we designed our interview so that users were asked to make a recommendation to a friend or family member.

Lin *et al.* used crowd-sourcing to analyze users' "expectations of apps' access to phone resources." The crowd-sourced expectations were used to design a new privacy summary interface for installation, which was more easily understood and efficient than the existing interface. Additionally, the authors found that telling users the purpose of the access improved decisions and eased their concerns [17] . Our study exposed users to the sharing of location data and phone identifier by two popular games, Toss It and Angry Birds, which are unexpected uses of data according to the crowd-sourced results.

2.2 Designing Usable Privacy Notifications & Control Over Data Leaks

Informing users about privacy and security issues is not trivial work. Privacy is usually not the user's primary task. The concept of privacy is often abstract or highly individual, and may not lend itself well to icons or sounds. We now discuss several studies that aim to make privacy issues visible to users and how they are related to our Privacy Leaks prototype.

Privacy Bird is a browser agent developed in a pioneering study on usability issues of privacy notification. Privacy Bird notifies users with sounds and icons when a website's privacy policies do not match the user's preferences. A series of focus groups and user studies was used to determine the effectiveness and understandability of the icon. The authors specifically look at how to bundle the many aspects of the privacy preferences. They find that users appreciate short summaries, meaningful terminology, and the appropriate granularity of information [6]. We attempted to integrate all of these into Privacy Leaks.

Kelley *et al.* created nutrition labels for privacy policy. Standardized grids were found to be an effective way of presenting information about a website's use of data [15]. Privacy Leaks also uses grids to visualize which information has been transmitted off the phone by applications.

Wi-Fi Privacy Ticker is a tool designed to improve users' awareness of personal information exposure over unencrypted Wi-Fi networks and provide control to prevent unwanted exposure [5]. It automatically drops a user's connection when a highly-sensitive term (as defined by the user) is being sent in the clear. A notification called the 'Ticker Display' and balloon tip provide instant notification about the data leakage. Participants used the ticker for 3 weeks and had a resulting change in awareness, as found by both open-text statements and responses to specific questions [5]. Their findings inspired us to include "Just-In-Time" notifications in Privacy

Leaks.

Felt *et al.* propose a framework for asking permission to perform certain activities on smartphones. The authors state that for automatically granted permissions, auditing mechanisms such as notifications can help users become better aware of the permissions. They also discuss install-time warnings, which is currently the only mechanism used to inform users about data being accessed from their Android phones [8]. We propose runtime feedback in addition to the install-time mechanisms.

Kelley *et al.* investigated ways to display users privacy information in a 'privacy checklist' in the Google Play Store, and asking them to choose between comparable apps. They found that their privacy checklist could impact users' decisions; in several pairs of apps the participants chose the app that requested fewer permissions [14].

AppFence is an automated system for allowing privacy control, which can provide fake or filtered data to applications and prevent the sending of sensitive data [10]. The authors tested 50 popular Android Market applications using an automated testing methodology and found that AppFence reduced data leakage without side effects on two thirds of the applications. However, they found trade-offs between usability and privacy on the remaining applications.

Zhou *et al.* developed an application for Android with a privacy mode. This application, which requires modifications to the Android framework, allows users to set fine-grained privacy settings about whether an application will have access to real, spoofed, anonymous, or empty data. The authors did not address users' understandings of the settings or dialogs [20]. This work along with similar studies [10] helped us design some of interview questions on privacy control in order to determine whether the proposed interfaces match users' expectations and desires.

3. DESIGNING PRIVACY LEAKS

Our prototype is built over TaintDroid [7], which instruments the Android operating system to determine whether privacy-sensitive data is being transmitted off the phone by Android applications. Our prototype reads data transmission events generated by TaintDroid. These events both trigger a notification and are written to a database, where the event information can be accessed later for the visualizations. However, in general, not all data transmissions are deemed unexpected by users (e.g., location data being sent off to a map server when the user is using a navigation app) and there needs to be a filter that can differentiate privacy leakages from legitimate data transmissions. We discuss how such a filter can be implemented in Section 7.

We used an iterative process to design the notifications and the layout for our visualization interface. This process included iterating over several designs by testing paper mockups and Android prototypes on colleagues who are not privacy experts.

We named our app *Privacy Leaks*, which may have led to a slight user bias about the information. Although it is consistent with our previous definition of leakage, this title may have negative implications about data sharing. We realized this after the study had been completed.

3.1 Notifications

The just-in-time notifications were intended to notify users at the moment data was being sent. In our prototype, the phone both vibrated and made a water-drop sound when privacy-sensitive data had been transmitted off the phone.

While we attempted to build a unique vibration, it is not clear whether the users would have been able to distinguish our vibrations from vibrations caused by an application. However, as shown in Figure 1, our prototype also included an icon and short text notice in the notification area, so users can check out the phone's status bar to see the source of vibration (e.g., Privacy Leaks notifications vs. text message arrivals).

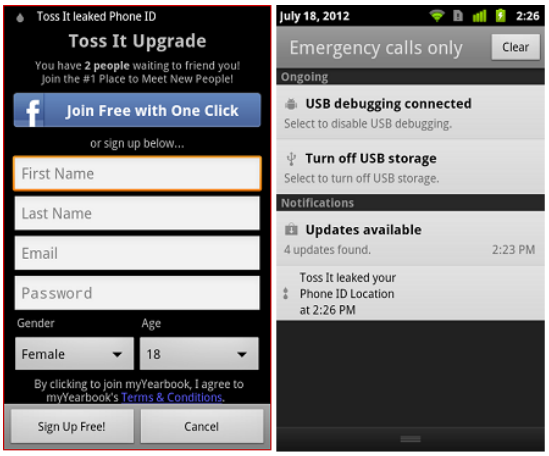

Figure 1: Notifications in the status bar (left) and in the notification drawer (right) by Privacy Leaks

3.2 Visualization

The visualization allowed users to compare, across apps, what information had been detected by TaintDroid as being shared recently. This type of visualization can be examined after an app has been used to see what was shared, and would require the user to actively open Privacy Leaks to view the information.

We focused on a simple layout that could quickly give users a sense of shared information without using jargon or requiring technical knowledge of Android. Through our iterative design process, we selected which information to show and how to display it. We used a grid layout (similar to [15]) to show an overview of the data that had been leaked. The columns showed the type of data, and the cells in each grid showed the number of times the information was sent. The cells were shown in red that became progressively brighter as the number of times increased.

The main visualization (e.g., Figure 2) shows data leaked by all applications over a period of time; this period is configurable by the user.

Our prototype included a jargon-free one-sentence description of the information: "How many times did Apps leak information about you since [timestamp]?" The rows include the application icon to help the user easily identify the application. The columns are the permission-based fields that are sent. We created a second screen, seen in Figure 3, to show the destinations of the data for the individual applications, available by clicking on the application icon.

Due to limited screen-space, we were not able to display a column for every type of data that could be shared. Therefore, we made the following design decisions to choose which

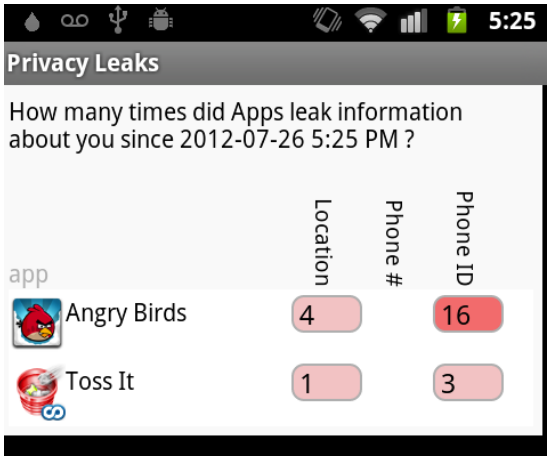

Figure 2: Main visualization screen of Privacy Leaks

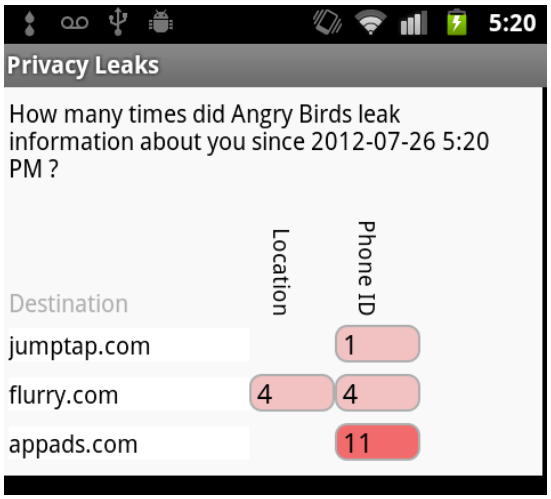

Figure 3: Application detail screen of Privacy Leaks

columns to show. Three columns are always shown, and other types of data are shown in different columns only if that data has been sent. In particular, we always show Location, Phone ID, and Phone #. Location and Phone ID are the two most frequent types of leaks [7]. We also always included the field Phone # to clarify that Phone ID is not the phone number. Phone ID can be used to uniquely identify the phone. As stated in in the Privacy Rights Clearinghouse Fact Sheet, "The privacy concern here is that information could be shared with third parties and compiled with other data to create a detailed profile about you without your knowledge or consent" [1]. Location can be used to locate the phone, and Phone # can be used to identify and to call the phone. Other privacy-sensitive columns, such as Address Book, appear if and when an application sends off that information.

During the design process, we found that users were confused by the different types of phone identifiers such as: "IMEI," "IMSI," "SIM card identifier," "Device serial number," and "Android Id." We renamed and collapsed these to a single group, "Phone ID," to avoid overwhelming jargon. Similarly, we did not distinguish between types of location data: we collapsed "Last-known Location," "NET-based Location," and "GPS Location" into "Location."

Furthermore, we did not show which location was sent, such as the exact GPS coordinates. Nor did we show a timeline of when information was sent. Our paper mockups of such visualizations were not well received, but we believe they are both feasible visualizations and we are considering them for future work.

Applications may also send parameters along with the above privacy-sensitive fields. Understanding this data often requires technical knowledge of the application. Therefore, we did not show this information out of concern that it would overwhelm or confuse users.

Our prototype was also instrumented to allow configuration for research purposes, including configuring the time frame, refreshing data, turning off notifications, and exporting data. However, participants were not expected to use these options and the usability of the configuration settings was not a part of our user study.

After the design iteration, we noticed that the grid is somewhat similar to the Wall Street Journal's visualization of data sharing[1].

4. STUDY METHODOLOGY

We conducted a lab study of 19 participants in July and August 2013 to investigate their existing understanding of potential privacy leakages while using smartphone applications and to collect initial feedback on our Privacy Leaks prototype. We interviewed each participant for up to an hour in the lab. Interviews were structured in the following order: 1) the participant plays two games without Privacy Leaks and answers questions about the games and data being sent off the phone, 2) the participant plays the same games with Privacy Leaks and answers the same questions as before, 3) the participant is interviewed about data control, the usability of Privacy Leaks, and perceptions of desired data sharing. We explain each part of the interview in further detail in the next section.

4.1 Study Procedures

The first part of the interview served as a control to gauge the participants' impression of the games and examine their knowledge of data leakage. After arriving and being briefed about the study, participants were given an Android phone that had applications pre-installed. Then, they were asked to play and compare two games: Angry Birds and Toss It. Participants were provided with copies of paper screenshots of the install process of both games, including the permissions screen. They had up to 7 minutes total to play and compare the two games. They were asked to evaluate the two games in order to recommend one to a friend or family member. This is somewhat longer than a typical session with an application of less than a minute [2]. However, participants were asked to think out loud and evaluate the applications, which typically lengthens the time to finish a task.

The participants selected a specific friend or family member, and then the researchers used that relationship (e.g., "wife" or "colleague") in all further questions to help the users create a specific and realistic scenario. Thirteen users selected a family member, such as their wife or nephew.

[1] `http://blogs.wsj.com/wtk-mobile/`

Five participants selected a minor: for example, a child or younger sibling. The participants were asked to imagine that the friend or family member would be playing the game "on the bus, at the doctor's office, or waiting to meet you somewhere."

After the first round of play, participants were asked to describe their recommendations of the games to their friends, and how they would describe the games on the app market. They were then asked about the information that was leaving the phones while they tested the games, and why the data was leaving and where it was going. This allowed us to evaluate their existing awareness and understanding of data sharing.

In the second part of the interview, participants were given a phone that was identical to what was used in the first part of the study, except it also had Privacy Leaks. The participants were told that an application was installed to notify them of data sharing, and that they had another 7 minutes to evaluate the same two games. After participants played the games, they were prompted to open Privacy Leaks to view the visualizations. Following this, participants were asked whether their recommendations changed, and were interviewed on their understanding and awareness of data leakages of the application. This allowed us to examine how users reacted to data sharing, and their revised understanding based on the information in Privacy Leaks.

The second part of the interview included the visualization-only and the JIT conditions. The interview questions about the games and the data sharing were the same across both the conditions, except for two differences. The participants who received JIT notifications were told, before the second part of the interview, "an application will inform you about information that is being shared through notifications, such as vibration and the sound of water dropping." Also, participants in the JIT condition were asked additional Likert-scale questions in the third-part of the interview on whether they found the noise and vibration annoying or interruptive.

The third part of the interview consisted only of interview questions, and did not include game-playing or application use. Participants were asked to describe what they would do if they could "completely control the data leaving the phone." Participants were asked about specific elements of Privacy Leaks notifications and visualization, which allowed us to examine the usability and likability of the application. They were also asked about desired control over data leakage, and the risks and benefits of sharing data.

The interview was structured enough to allow comparisons between participants, but open enough to allow the researcher to probe about specific comments. The interview questions and instructions were the same across all participants. There were no written components of the interview—the questions were all asked and answered orally. The interview included a combination of open-ended, yes/no, and Likert-scale questions. Participants did have access to a printed copy of the Likert-scale values to refer to when answering the Likert-scale questions.

One researcher led all the interviews. One of two additional researchers took notes in the interviews. The interview was audio-recorded. The results were coded iteratively based on both the notes of the two researchers who were present at the interview and the audio transcripts. Two researchers sought themes within the responses, and then coded the results based on the theme list. They then iteratively re-coded based on re-evaluating the responses and discussion until agreement was reached between the coders.

Our results are entirely qualitative. We include the number of participants who responded with certain themes or ideas, but we do not intend to imply statistical significance, or that this represents a larger population.

4.2 Game Features

The two games used in the study were decoys; we wanted participants' attention on the primary task of selecting the game, as opposed to thinking about privacy. Both games involved a simple flick gesture to send an object on a trajectory, aiming at either pigs or a trash can. Therefore, the games were similar in their simplicity and style. While the games themselves were not of particular importance to our study, both games had features that were important to participants' conceptions of data sharing. Neither game uses location for functionality, but both send location information to the game developers and third-parties.

Figure 4: Screenshot of Angry Birds while game is in play, with an ad

Figure 5: Screenshot of Toss It while game is in play

Angry Birds showed a banner ad, as shown in Figure 4 that several participants remarked upon. Angry Birds sometimes shows a full-screen ad as well. Since recruited participants were already familiar with Angry Birds, several commented on the possibility of viewing ads, even if none were displayed while they played the game during the lab study.

Some participants recognized that data would be shared if the game was social or allowed score sharing. As seen in Figure 5, Toss It had four buttons for social networking

id	sex	age	condition	data sent
1	m	32	notifications	33
2	m	28	notifications	28
3	f	39	notifications	29
4	f	49	visualization only	21
5	f	43	visualization only	37
6	m	52	notifications	31
7	m	44	visualization only	23
8	m	23	visualization only	21
9	f	38	visualization only	29
10	f	21	notifications	18
11	f	43	notifications	14
12	m	38	visualization only	32
13	f	37	visualization only	17
14	m	38	notifications	25
15	f	37	visualization only	29
16	m	28	visualization only	38
17	m	26	notifications	36
18	f	44	notifications	63
19	f	20	notifications	17

Table 1: Participants' demographics, condition, and number of times data was sent off the phone while they used Privacy Leaks

or social games, such as challenging another player, at the bottom of the play screen. These buttons took users to a screen asking them to log in with their Facebook account.

4.3 Participants

Ten male and nine female participants were recruited from the Seattle Metropolitan area by a service agency for our study. Our intention was to get variety and diversity, not to represent the USA population statistically. Participants were compensated with a choice of software gratuity. We screened to exclude people with a computer science degree. The average age was 35, in a range of ages from 20 to 52 years. Seven participants had a bachelor's degree, while 6 had completed high school and 6 had an advanced degree. Table 1 includes details on the participants' demographics. All participants were current Android users for at least 3 months, and had installed and played Angry Birds before participating in the study. To avoid priming the participants in advance about privacy or data leakage, the participants were told that the study was about Android games.

5. INITIAL UNDERSTANDING

Participants played both games for a total of 3-7 minutes in the first part of the interview before making a recommendation and describing the game. We then asked them what data had left the phone while they played the games. They were therefore given a specific situation in which to evaluate data sharing, that allowed us to examine the understanding of data leakage before viewing Privacy Leaks. They were asked about why, when, and what data left the phone, and whether both games shared information.

5.1 Purpose of Sharing

We found that participants' level of awareness about the data that was shared could be roughly categorized into three groups:

- Group 1: Five participants stated explicitly that they had never before thought about information leaving the phone.
- Group 2: Eight participants believed that data was shared only with application developers for the purpose of improving the application.
- Group 3: Six participants understood that data was used for marketing but were surprised by the scope of data sharing, including the frequency of data sharing and the destination of data.

While the degree of awareness was different, none of the participants entered with a complete understanding of data sharing and the scope.

Participants belonging to Group 1 had never thought about data sharing before. P4 expressed her uncertainty about whether data left the phone, "Maybe it's not [leaving]. Maybe it's all in the phone. That's a tricky question. I don't know. Does it leave it?" P3's comments represent the idea that the game is self-contained, "It was my understanding once you downloaded it to your phone, it's on your phone. It didn't need to communicate with anything." In the first part of the interview, the participants were prompted with several open-ended questions about where, when, and why data was leaving, but they were often unable to answer. Several of these participants adopted a new understanding as they pondered our questions and thought out loud. These new ways of understandings fit into the next two categories.

Participants belonging to Group 2 believed the application is a self-contained environment. For example, P5 said, "If I'm within the Rovio game I'm thinking it [data] goes to Rovio. I didn't think if I'm within the application environment [data is leaving the phone]." Some participants commented on social networking. For example, P2 said, "Toss It [would share] with online communities if I had continued to start a challenge. Other than that it's not sending anything." P19 said, "Information is useful for analyzing the product. They can customize the game based on where and how long the game is played. I think it is about knowing the market." These participants were not aware that data was shared for the purpose of marketing, and thought their level or skill was sent in order to improve the game.

Participants belonging to Group 3 were aware of targeted advertising integrated with smartphone applications. However, even those who mentioned targeted ads were still confused about the mechanisms. P6, an older participant who had seen an ad for insurance in Angry Birds, stated that data was being shared for marketing. He said, "It didn't ask for age, education, doesn't know who is playing, but it might have email. A ten-year old wouldn't receive an ad for insurance," indicating an understanding that targeted ads could exist, but not sure how they would get enough information to target him.

5.2 Additional Perspectives on Data Sharing

Seven participants referred to the existence of terms and services but were not clear on what was included in these terms. For example, P18 said, "We give them all these permissions," when referring to what data left the phone but she wasn't specific about what the terms were. P11 expressed uncertainty while correctly summarizing the situation "Does it need to ask permissions? I think it asks something when you download it. I guess you can't download it without allowing it." Only a few of the participants examined the printouts of the install screenshots that were on the table

in front of them to find out what information was being shared. This suggests that even though users are aware of the permissions requests, they rarely see them as a resource for understanding data sharing.

P8 thought that data moves in a cycle with continuously coming and going. "Data can't always be stored in memory. It is in-going and out-going." In this (incorrect) perspective, data is shared because the limited memory space on the phone pushes the data out to remote servers to use them as temporary storage.

Overall, all nineteen participants had a limited understanding of whether and how often these smartphone games may collect the user's privacy-sensitive data. Next, we analyze participants' response after repeating the same task with Privacy Leaks.

6. EARLY EXPERIENCES WITH PRIVACY LEAKS

This section discusses the participants' reactions to our Privacy Leaks prototype in the second part of the interview, after having viewed the visualization shown in Figure 2 and Figure 3. Ten participants in the JIT condition felt and heard JIT notifications in addition to seeing the visualizations. Participants were prompted with the same open-ended questions about where, when, and why data was leaving as in the first part of the interview, in order to gage the difference in understanding after using Privacy Leaks. We discuss overall reactions to our prototype implementation, including which new information violated the participants' initial understanding of data sharing.

6.1 Surprised by Actual Data Leakage

Across all groups, participants were most surprised by the frequency and destinations of the data. Usually, the information that was new, and did not fit into their previous understanding, was the most surprising.

Many participants were surprised by the frequency of data sharing, regardless of their initial perspective. However, participants belonging to Group 1 were very surprised by the frequency. They struggled to understand why the data was sent multiple times in the short time span they played the games. P3 said, "Why does it need to say what my Phone ID was more than once?"

Participants belonging to Group 2 were typically most surprised by the unrecognized destination URLs (e.g., flurry.com or admob.com). P1 expressed this concern about not knowing where the data was going: "Destination is surprising; that is a little concerning. It would be nice to have some sense of who is collecting the information." Participants were sometimes able to make assumptions about the destinations upon examining the URLs, as some have "ad" in their name. P7 said, "I've never heard of any of these companies. I assume they are using it for marketing." P10 had a similar comment while looking at the list of destination URLs, "Are those things supposed to mean anything to me? Oh. It's all advertisers."

Participants belonging to Group 3 were typically most surprised by the number of different destinations *and* the frequency of sharing. P19 expressed her anger that the game was sharing the data with many companies, "I find Toss It slime as they let other companies collect information." She continued, "My eyes have been opened today. Every time you use the phone, every time you download an application [it] is not big brother watching you, but a lot of little brothers watching you. And they want to sell something to you."

6.2 Opinions of Privacy Leaks

Playing the games for 3-7 minutes was sufficient for participants to experience notifications and build up a history of data sharing to see on the visualization. Data was shared an average of 29 times per participant. As participants played each game for different amounts of times and accessed different parts of the games, the amount and types of information shared varied.

Overall, we found that participants liked our Privacy Leaks prototype and would want to install a similar app on their phone. Sixteen participants agreed or strongly agreed that "the information provided by Privacy Leaks is useful," and fifteen agreed or strongly agreed with the statement "I am likely to install an application like Privacy Leaks." For more information on responses, a histogram of responses to these particular questions is shown in Figure 6 in Appendix B. Unfortunately, the Privacy Leaks app would only provide useful information on a special Android phone that is instrumented to run TaintDroid. Interested users can build and flash their Android phone with the TaintDroid system image following the instructions available in `http://appanalysis.org`. However, this additional step can be a substantial barrier for deployment.

A number of participants indicated their desire to install the Privacy Leaks tool immediately, and asked when it would be available on the application market. P2 said Privacy Leaks is "a great asset to have on your phone. It gives you information about where your data is going so you can choose [apps] more wisely." Most participants disagreed (3) or strongly disagreed (13) with the statement, "The information was irrelevant." P11 described the interface as "a good app for a person who is curious about data sharing", indicating that although she did not worry about data sharing, she thought it was useful for others. We are cautiously optimistic that this result indicates that there is a demand for privacy applications on Android smartphones that provide information about data sharing.

Typically, participants were able to read the text and numbers in the grid format and interpret them quickly. After using Privacy Leaks, participants were able to correctly answer questions about with whom information was shared, the type of information being shared, and which application sent the most data.

However, we did find that there were areas for improving the interface, as only 6 participants claimed they "understood what everything meant in Privacy Leaks." As discussed in Section 7, participants struggled to understand what Phone ID meant. Additionally, they did not know or recognize the different destination domains.

Three users initially failed to understand the purpose of Privacy Leaks, thinking that it was responsible for the data being shared. Similar results were found in a study on the impact of projecting excerpts from open network traffic on a wall in a public space [16]. More education would be needed about the purpose and goals of such tools to alleviate such confusion. This could be done through marketing, or providing an additional explanation of Privacy Leaks at install time. We also asked participants if Privacy Leaks was "accurate" and several stated they had no way of know-

ing. Two suggested it would take reviews in trusted media (e.g., "TechCruch" and "BusinessWeek") to convince them that Privacy Leaks was trustworthy. Others said they would trust Privacy Leaks if it were from a well-known and trusted corporation or part of the phone's operating system.

6.3 Reactions to Just-In-Time Notifications

In the Just-In-Time (JIT) condition, 10 participants felt and heard the JIT notifications in addition to the visualizations. Due to the small sample size, we did not run statistical tests between the two conditions. However, there were some general differences between the groups that we describe.

Some participants were surprised by the frequency of the notifications. For example, P18 said, "I hear drops, this is going crazy! There are a lot of bleeps!" When playing the games with JIT notifications, participants often tried to figure out why data was being sent. P2 commented, "I'm trying to figure out when it actually sends data. I don't know what it just sent, [drop sound] not entirely sure what is being sent, I'm just loading it up. Probably checking for new content or updates." Participants questioned whether data was sent when they scored or reached a new level.

On average, participants found the sounds more annoying than the vibrations. However, this depended heavily on the individual. Participants suggested that Privacy Leaks should allow them to configure whether sounds and vibrations should be enabled. This functionality was already built into Privacy Leaks, but we did not include it in the study.

We anticipated that participants would overwhelmingly find the JIT notifications annoying or interruptive. Figure 7 in Appendix B shows that participants had mixed reactions to the sounds and vibrations. For example, only 1 out of 10 found the vibrations distracting. While 5 out of 10 participants agreed that the sounds were distracting, 5 also said the sounds would allow them to keep working or playing without interruption. This may be due to the short amount of time using the notifications. Furthermore, by the time the participants received notifications, they had already been prompted with questions that they had a hard time answering, such as when data was sent. This probably increased their curiosity and therefore their appreciation of the notifications. Future work is needed on how users respond to JIT notifications over time.

Eight of the 10 participants who saw JIT notifications responded to the questions "The information provided by Privacy Leaks is accurate" affirmatively, while only 3 of the 9 who saw only visualizations were affirmative, typically saying they didn't know. This indicates that participants in the JIT condition were more likely to find Privacy Leaks accurate. It is possible that the audio, tactile, and visual feedback all combined to reinforce the information, making it seem trustworthier than the visualization alone.

6.4 Recommendations to Friends and Family

As smartphone owners rely on friends and family for app recommendations, we were curious about whether privacy leakage information would change participants' recommendations to friends and family [4]. Most (12) participants would not change their recommendation to their friend or family member about the game after using Privacy Leaks, saying the functionality was still the same. However, participants frequently said they would add that the data was being leaked. P14 said, "Angry Birds is still a fun game. I would probably inform her that they are tracking what you are doing." This indicates that game functionality was typically still more important to participants than data sharing.

However, some participants changed their recommendation. P16, who discussed recommending the game for a cousin in college said, "Yes, I would advise her not to play the Angry Birds after seeing the leaking."

All but 2 participants would add that information was being leaked if they were to write a description on the application market. For example, P11 said, "I might put a little note about Angry Birds talking to a couple companies I don't know." P18 described how he would recommend the game, "Probably say that both like to leak location and phone id. It is probably for marketing. It is important to let people know. Some people think it is helpful, some think it is invasive."

6.5 Privacy Preferences

Participants' existing privacy preferences impacted their reactions to the data sharing. Although we did not ask this directly, over the course of the interview six participants volunteered that they were not particularly privacy sensitive. They explained that data sharing was not overly concerning because they were not, for example, "paranoid" [P1] or "conspiracy theorists" [P17].

Two participants were even more sanguine about the data sharing. They were fully aware that data-sharing was a trade-off for free games, and were fine with this model. P11 said "It's not really a big deal to me. It can be a good thing. As long as they don't flash ads every second or something, I really don't mind." P7 said, "As long as that does not affect my life in negative way, I am ok to give the information away."

On the other hand, several participants had strong negative reactions to learning about data sharing after viewing Privacy Leaks. P5 said, "It really bothers me that this sort of thing happens, because I want to remain as anonymous as possible." P2 said, "This makes me an angry birdy."

6.6 Risks & Benefits of Data Sharing

In order to probe how users make decisions about data sharing, it is important to understand their concepts of risks and benefits of data sharing. We asked participants about the benefits and risks of sharing data with the questions, "Are there any benefits [risks] to you or [your friend] when the game shares information, and what are they?" We substituted the words "your friend" for the friend or family member they had selected at the beginning of the interview. We asked these questions at the end of the interview, to avoid biasing the interview, and therefore the participants had already viewed Privacy Leaks and knew how often data was shared and what the destination URLs were.

Fourteen participants thought there was no benefit overall to sharing the information with games such as Angry Birds. This is in contrast to Ur *et al.*, whose participants often recognized that there may be economic benefits to themselves and the websites from on-line behavioral advertising (OBA). This may be due to the wording of the question or the context (data sharing from this particular game versus OBA in general). Alternatively, the participants in Ur *et al.* may have been better informed because they watched an informational video on OBA at the beginning of the interview, whereas our participants were asked to provide opin-

Perceived benefits	Perceived risks
none (14)	accidental purchases
free games (2)	porn
targeted ads (3)	virus
search and rescue (2)	annoying SMS
	price discrimination
	telemarketers
	find kids' backyard
	creepy
	social networks—
	(friends can access information)
	identity theft
	data breach
	worker with access to data—
	'goes postal'

Table 2: Responses to "Are there any benefits/risks to you or [your friend or family member] when the game shares information, and what are they?".

ions without any education outside of our prototype [18].

Two of our participants mentioned that sharing location with certain applications was useful for functionality. Three participants mentioned targeted ads, but were unsure it was a benefit to them. For example, P19 said, "I guess customized ads are a benefit. It's a stretch. I don't click on ads, so I'm not sure I can make that argument." P3 mentioned targeted ads as a possible benefit, but doubted the efficacy: "phone ID, location, that doesn't help them hone it on what I like." Some of these participants then concluded that there was no real benefit.

Only two participants stated that free games or improved functionality are benefits of sharing information. Two participants also mentioned search and rescue as a benefit of providing their location; their (mis)understanding was that first responders would be able to find them since their location had been shared.

Participants were also asked about the risks to themselves or their friend or family member when a game shares information. The risks mentioned by participants spanned an array of possibilities, as shown in Table 2 including accidental in-app purchases, getting a computer virus, and receiving annoying SMS messages. Some risks do not involve provable harm, such as someone knowing their kids' location, being creepy, or price discrimination. Some participants were particularly concerned about the risks to their children of data sharing, and were concerned that bad people could get access to the information.

7. DISCUSSION & FUTURE WORK

We have built an interface that informed users about the frequency and destination URLs of data shared by smartphone applications. Most participants indicated that the application was useful and that they would like to install a similar application probably because it provided information participants could not get elsewhere in a simple layout with some explanatory text. However, future work is needed to improve the interface and validate through a field study how an improved interface can actually raise users' awareness of privacy leakages on smartphones. We first discuss near-term opportunities to improve the Privacy Leaks interface based on the suggestions from study participants. We then present a few suggestions to help users make informed decisions about data sharing with smartphone applications and control over their privacy-sensitive data.

7.1 Improving the Interface

There were three main areas in which participants frequently requested further information: phone ID, destination, and location. Participants felt that "Phone ID" was an unfamiliar term, and they were unsure about the implications of the Phone ID being shared. P18 expressed this confusion, "I don't know if it means type of phone or identifying the phone with the person." Participants were unclear if phone ID was just the model information, or if it also included their phone number and email address. Location was also confusing to some participants, who wanted a better understanding of how fine-grained "location" is. Participants suggested a roll-over for these columns that would explain these two fields more.

Participants did benefit from seeing the "Phone #" column, even if it was blank, as it helped them see that phone ID is not the phone number. Some participants asked about what other fields were possible. Despite the visual clutter of adding more columns, it may be helpful for users to see the possible privacy-sensitive types of data that could be sent, to help them distinguish what has been sent. For example, P14 became concerned about the capabilities of the phone and data sharing, wondering if the features of the phone could be combined to lead to inaccurate profiling, "Is that how they get those ads on top of the screen? They could take a picture of me and assume I'm into rap music." He wasn't clear that the application did not have control of the camera or other functionality.

Phone ID was often being sent along with additional information interpretable by the game developers. Our interface did not show the entire set of data sent with the Phone ID, as we thought it would be overwhelming or incomprehensible. However, additional information about the purpose of the data sharing might enable users to understand the frequency of data sharing.

As mentioned earlier, participants were confused about the destination domains, as the URLs were typically unfamiliar to them. Some users asked for the ability to click to open the domain in a browser. However, this may not be helpful, as many of the URLs do not have consumer-facing web sites, typically presenting toolkits aimed at application developers. While clicking on the domain may not be useful, other types of information could be provided for the user, such as brief descriptions of the company's purpose, and a link to the company's "opt-out" page, if one exists.

There are a number of ways to visualize data; we chose a simple grid format. The grid visualization highlighted the number of times different types of data were sent by applications. While this simple grid format allowed participants to quickly read the information displayed, it remains to be seen whether other visualization techniques (e.g., Wi-Fi Privacy Ticker [5]) would be more effective to improve users' understanding of privacy leakages associated with smartphone applications. Once an improved interface is developed, our plan is to run a field study to evaluate whether ongoing feedback on data leakages can be presented to users without causing too much annoyance and having users desensitized quickly over time.

7.2 Providing Usable Control

Informing users about data leakage is only a first step; users should also have control over data sharing. Today, even if users become aware of privacy issues with respect to certain smartphone applications, they have little choice but to uninstall offending applications on their phone. Although there are a few research prototypes that allow users to selectively hide privacy-sensitive information against data hoarding applications (e.g., AppFence [10]), it still remains open how this prototype can be properly configured by users to protect their privacy. For instance, our participants had a hard time considering all the implications of blocking data sharing on all their applications. This is not surprising, considering how many participants were unaware of data sharing before the study, and had not had time to consider how they would control data. However, some participants suggested particular contexts in which they could imagine wanting particular control. These included mobile banking, being in a government building, and not wanting their car insurance company to find out they were texting while driving. This suggests that users have very specific desires for controlling information that might not fit into broad categories. Two participants made analogies to their computers when suggesting protective steps they could take. They discussed installing an anti-virus and deleting cookies as options they could take to control information sharing. Exploring usable privacy control mechanisms is another future direction to pursue by our team.

8. LIMITATIONS

Qualitative studies with small sample sizes have limitations, such as lack of statistical power. However, the in-depth interviews did allow us to get qualitative insights into participants' reactions to the notifications and visualizations, as well as their understanding of the information they were seeing. While we provided the numbers of participants that made similar statements, we do not claim that these could generalize to larger populations.

A lab study has limited ecological validity. The participants were not using their own phones, playing games of their own choice, and were not in privacy-sensitive locations (e.g., at home). With a greater range of locations, applications, and situations, participants may be more sensitive to the particular context of information sharing. In real settings, users may actually be more concerned with privacy leakage than when they are in a lab and using a lab phone. We did ask participants to imagine their friend or family member playing the game "on the bus, at the doctor's office, or waiting to meet you somewhere." However, we have no indication that participants considered the privacy-sensitive nature of any of these locations.

On the other hand, lab participants may exaggerate their concern and interest in the task at hand (understanding or caring about privacy leakages) to appear to be a good participant. Also, naming our app as Privacy Leaks may have biased some participants who would not have considered privacy risks otherwise. Furthermore, they may not actually do what they claim they will do when they are in the lab. Previous work has shown that survey participants may report that they engage in privacy-protecting behavior, but behavioral studies show that these self-reports are inaccurate [11].

While a field study would address some of the concerns about a lab study, we would only be able to measure quantitative actions that the participants take. We would not have access to the initial verbal reactions or questions that we have in a lab, nor would we be able to probe them for details as they are viewing the interface.

9. CONCLUSIONS

Our qualitative interviews provide insight into users' understanding of data sharing, both before and after being informed in real-time about data sharing from two smartphone games. Overall, we have found that participants have misconceptions about data sharing occurring through smartphone applications, do care about applications that share privacy-sensitive information with third parties, and would want more information about data sharing. Thirteen out of 19 participants did not know that data would be shared for the purpose of advertising. Many had never considered it before; others believed data was only shared with the application's developers in order to improve the game or provide useful functionality; yet others understood that data was being shared for marketing purposes. Most participants were not aware and often not comfortable with the scope of data sharing done by these game applications, both in terms of amount of data shared and the destinations of the data. This is particularly troubling as the ad and app industry may be working on the assumption that users understand the trade-off between free apps and data sharing.

Moving forward, we continue to explore tools and interfaces that can improve users' awareness of privacy leakages while using smartphone applications and usable control mechanisms that can help users prevent unwanted data sharing with smartphone applications.

10. REFERENCES

[1] Fact sheet 2b: Privacy in the age of the smartphone. *Privacy Rights Clearinghouse*, Sep. 2012.

[2] M. Böhmer, B. Hecht, J. Schöning, A. Krüger, and G. Bauer. Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage. In *Proc. of MobileHCI*, 2011.

[3] J. L. Boyles, A. Smith, and M. Madden. Privacy and data management on mobile devices. *Pew Internet and American Life Project*, Aug. 2012.

[4] E. Chin, A. Felt, V. Sekar, and D. Wagner. Measuring user confidence in smartphone security and privacy. In *Proc. of SOUPS*, 2012.

[5] S. Consolvo, J. Jung, B. Greenstein, P. Powledge, G. Maganis, and D. Avrahami. The Wi-Fi privacy ticker: improving awareness & control of personal information exposure on wi-fi. In *Proc. of Ubicomp*, 2010.

[6] L. Cranor, P. Guduru, and M. Arjula. User interfaces for privacy agents. *TOCHI*, 13(2):135–178, 2006.

[7] W. Enck, P. Gilbert, B. Chun, L. Cox, J. Jung, P. McDaniel, and A. Sheth. Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones. In *Proc. of OSDI*, 2010.

[8] A. Felt, S. Egelman, M. Finifter, D. Akhawe, and D. Wagner. How to ask for permission. In *Proc. of HotSec*, 2012.

[9] A. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, and D. Wagner. Android permissions: User attention,

comprehension, and behavior. In *Proc. of SOUPS*, 2012.

[10] P. Hornyack, S. Han, J. Jung, S. Schechter, and D. Wetherall. These aren't the droids you're looking for: retrofitting android to protect data from imperious applications. In *Proc. of CCS*, 2011.

[11] C. Jensen, C. Potts, and C. Jensen. Privacy practices of internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies*, 63:203 – 227, 2005.

[12] J. Jung, S. Han, and D. Wetherall. Short paper: Enhancing mobile application permissions with runtime feedback and constraints. In *Proc. of the workshop on Security and Privacy in Smartphones and Mobile devices*, 2012.

[13] P. Kelley, S. Consolvo, L. Cranor, J. Jung, N. Sadeh, and D. Wetherall. A conundrum of permissions: Installing applications on an android smartphone. In *Proc. of USEC*, 2012.

[14] P. Kelley, L. F. Cranor, and N. Sadeh. Privacy as part of the app decision-making process. In *Proc. of CHI*, 2013.

[15] P. G. Kelley, L. Cesca, J. Bresee, and L. F. Cranor. Standardizing privacy notices: an online study of the nutrition label approach. In *Proc. of CHI*, 2010.

[16] B. Kowitz and L. Cranor. Peripheral privacy notifications for wireless networks. In *Proc. of the Workshop on Privacy in the Electronic Society*, 2005.

[17] J. Lin, S. Amini, J. Hong, N. Sadeh, J. Lindqvist, and J. Zhang. Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing. In *Proc. of UbiComp*, 2012.

[18] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *Proc. of SOUPS*, 2012.

[19] J. Urban, C. Hoofnagle, and S. Li. Mobile phones and privacy. *UC Berkeley Public Law Research Paper*, 2012.

[20] Y. Zhou, X. Zhang, X. Jiang, and V. Freeh. Taming information-stealing smartphone applications (on android). In *Proc. of TRUST*, 2011.

APPENDIX

A. INTERVIEW SCRIPT

Welcome to our study. My name is ... and this is ... who will be taking notes.

Thank you for coming. Before we begin, let me tell you some important information about the study. We will be recording what is said in this interview, but everything will be anonymous. Your name and identifying information will be stored separately from your comments.

Please think out loud as you go through the tasks. That is, tell us what you are thinking as you go. Our goal is to evaluate our tools; not you. Everything you say, including confusion and questions, is very valuable to us.

Imagine that a family member or friend has just acquired an Android. They would like your advice on which game they should install. Imagine they will be playing these games during on the bus, waiting in the doctor's office, or maybe while they wait to meet you somewhere. Please take a minute to choose someone and tell us their relationship to you.

A.1 First Part of Interview

We will be giving you an Android phone with two free games, which we just installed before this interview. We are asking you to try these two games and decide which one you recommend to your friend. One game you are already familiar with is Angry Birds. The second game is called Toss It. Have you already played Toss It?

Screenshots from the install for each game are provided. You are welcome to refer to these in addition to actually playing the games. You will have up to 7 minutes to decide which game you prefer. Remember to think aloud.

[Participants played the games for 7 minutes or less.]

- Which game would you recommend and why?
- How would you describe each game to your friend?
- What would you write about each game in the app market?
- What information do you think was leaving the phone in the past 7 minutes while you played the games?
- Who was the information being shared with?
- Why was the information leaving the phone?
- Which application was sharing the data?
- What were you doing when the data was shared?

A.2 Second Part of Interview

This second phone has the same two games freshly installed. We have also installed an application that will inform you about information that is being shared [through notifications, such as vibration and the sound of water dropping]. You will have 7 minutes to play these games again. Imagine that you are evaluating these two games for your friend to use as he is waiting, for example, at the bus stop, at the doctors, or meeting you somewhere. Remember to think aloud. The app we installed is called Privacy Leaks, and you may look at it after playing the games.

[Participants played the games for 7 minutes or less. After playing games, participants were prompted to view Privacy Leaks.]

- Have your recommendations to your friend changed and why or why not?
- Would you describe these games the same way?
- What would you write about each game in the app market?
- Was there a relationship between when data was shared and what you were doing?
- Who was the information being shared with?
- Why was information leaving the phone?
- What type of information was being sent the most?
- Which application sent the most data and what data was being sent?

A.3 Third Part of Interview

Now imagine that these two games and Privacy Leak were on your own phone.

- Imagine that you had complete control over how your data was shared. What would you do?
- What if you could... would that be ok?
  - Stop information being sent when I'm in a particular location (e.g. at work, at home).
  - Stop information from being sent to particular companies or websites.
  - Stop information from being sent when I'm doing certain things (e.g. driving, sleeping).
  - Stop information from being sent by particular apps.

– Stop certain information being sent, such as phone Id, or location, regardless of app or anything else.

The next few questions are specific to Privacy Leaks.

- Do you think the information given by Privacy Leaks was accurate?
- Would would you tell your friend about Privacy Leaks.
- What would you write in the app market about Privacy Leaks?
- What would you change about Privacy Leaks?

I'm going to ask a series of questions about the apps that you can respond to on a scale of 1-5, with one being "Strongly agree" and five being "Strongly disagree." [Interviewer places paper with likart-scale on table for reference.] Feel free to elaborate.

- The information provided by Privacy Leaks is useful.
- I understood what everything meant in Privacy Leaks.
- The sounds are distracting. [JIT condition only]
- The vibrations are distracting. [JIT condition only]
- I am likely to install an application like Privacy Leaks.
- The sounds would allow me to keep working or playing without interruption. [JIT condition only]
- The vibration would allow me to keep working or playing without interruption. [JIT condition only]
- The information was irrelevant.
- The information provided by this tool is confusing.

Final few questions

- Would you pay extra for a game that didn't send this information?
- Are there any benefits to you or your friend when the game shares information, and what are they?
- Are there any risks to you or your friend when the game shares information, and what are they?

B. HISTOGRAMS OF RESPONSES TO LIKERT-SCALE QUESTIONS

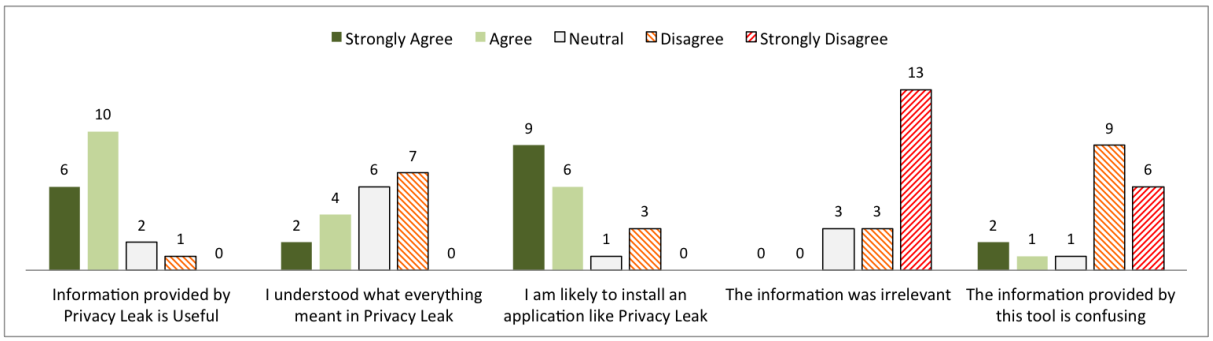

Figure 6: Responses to Likert-scale questions about Privacy Leaks

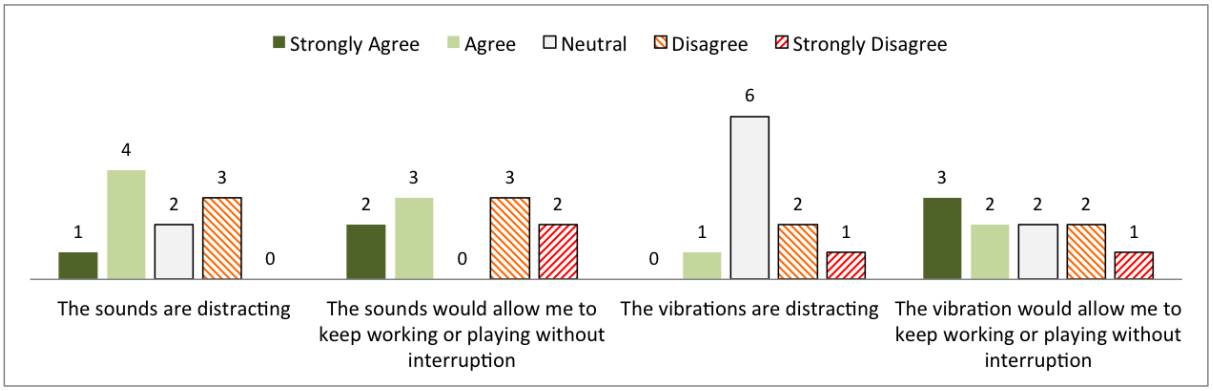

Figure 7: Responses to Likert-scale questions about Just-In-Time Notifications