Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy through Crowdsourcing

Jialiu Lin[1] Jialiul@cs.cmu.edu
Shahriyar Amini[1] samini@ece.cmu.edu
Jason I. Hong[1] jasonh@cs.cmu.edu
Norman Sadeh[1] sadeh@cs.cmu.edu
Janne Lindqvist[2] janne@winlab.rutgers.edu
Joy Zhang[1] joy.zhang@sv.cmu.edu
[1]Carnegie Mellon University [2]Rutgers University

ABSTRACT
Smartphone security research has produced many useful tools to analyze the privacy-related behaviors of mobile apps. However, these automated tools cannot assess people's perceptions of whether a given action is legitimate, or how that action makes them feel with respect to privacy. For example, automated tools might detect that a blackjack game and a map app both use one's location information, but people would likely view the map's use of that data as more legitimate than the game. Our work introduces a new model for privacy, namely *privacy as expectations*. We report on the results of using crowdsourcing to capture users' expectations of what sensitive resources mobile apps use. We also report on a new privacy summary interface that prioritizes and highlights places where mobile apps break people's expectations. We conclude with a discussion of implications for employing crowdsourcing as a privacy evaluation technique.

Author Keywords
Mental model, Privacy as expectations, Privacy summary, Crowdsourcing, Android permissions, Mobile app.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design, Human Factors.

INTRODUCTION
The number of smartphone apps has undergone tremendous growth since the inception of app markets. As of June 2012, the Android Market offered 460,000 apps with more than 10 billion downloads since the Market's launch; the Apple App Store offered more than 650,000 apps with over 30 billion downloads since its launch. These mobile apps can make use of a smartphone's numerous capabilities (such as users' current location, call logs, and other information), providing users with more pertinent services and attractive features. However, access to these capabilities also opens the door to new kinds of security and privacy intrusions. Malware is an obvious problem[17], but a more prevalent problem is that a good number of legitimate apps gather sensitive personal information without users' full awareness. For example, Facebook and Path, were found uploading users' contact lists to their servers, which greatly surprised their users and made them feel very uncomfortable [21, 34].

A number of research projects have looked at protecting mobile users' privacy and security by leveraging application analysis [10, 13-15, 19], or proposing security extensions that provide app-specific privacy controls to users [6, 22, 39]. These systems are useful for capturing and analyzing an app's usage of sensitive resources. However, no purely automated technique today (and perhaps not ever) can assess people's perceptions of whether an action is reasonable, or how that action makes users feel with respect to their privacy. For example, is a given app's use of one's location solely for the purpose of supporting its core functionality? It all depends on the context: for a blackjack game, probably not, but for a map application, very likely so. However, currently, users have very little support in making good trust decisions regarding what apps to install.

In this paper, we frame mobile privacy in the form of people's *expectations* about what an app does and does not do, focusing on where an app breaks people's expectations. There has been a lot of discussion about expectations being an important aspect of privacy [33]. We framed our inquiry on the psychological notion of *mental models* that first introduced by Craik [11] and later mentioned in other domains [29]. All people have a simplified model that describes what people think an object does and how it works (in our case, the object is an app). Ideally, if a person's mental model aligns with what the app actually does, then there would be fewer privacy problems since that person is fully informed as to the app's behavior. However, in practice, a person's mental model is never perfect. We argue that by allowing people to see the most common misconceptions about an app, we can rectify people's mental models and help them make better trust decisions regarding that app.

UbiComp' 12, Sep 5 – Sep 8, 2012, Pittsburgh, USA.

We believe that this notion of privacy as expectations can be operationalized by combining two ideas. The first is to use crowdsourcing to capture people's mental models of an app's privacy-related behaviors in a scalable manner. This requires some knowledge of an app's actual behaviors, which can be obtained with app analysis tools such as TaintDroid. The second is to convey these expectations to users through better privacy summaries that emphasize the surprises that the crowd had about a given app.

Our long term goal is to build a system that leverages crowdsourcing and traditional security approaches to evaluate the privacy-related behaviors of mobile apps. This paper presents the first step to understand the design space and the feasibility of our ideas.

We make the following research contributions:

- We demonstrate a way of capturing people's expectations using crowdsourcing. More specifically, we conducted user studies on Amazon Mechanical Turk (AMT) with 179 Android users, surveying their expectations and subjective feelings about different apps accessing sensitive resources (such as location, contact lists, and unique ID) in different conditions.
- We identify two key factors that affect people's mental model of a mobile app, namely expectation and purpose, and show how they impact users' subjective feelings.
- We present an analysis which indicates that informing users of why a given resource is being used can allay their privacy concerns, since most users have difficulty figuring out these purposes.
- We present the design and evaluation of a new privacy summary that emphasizes behaviors that did not match the crowd's expectations. Our results suggest that our interface significantly increases users' privacy awareness and is easier to comprehend than Android's current permission interface.

RELATED WORK

We have organized related work into three sections: an overview of the Android permission system; research on mobile app analysis and security extensions; and relevant work in mental model analysis and design for privacy-related user interfaces.

Android Permissions

The Android permission framework is intended to serve two purposes in protecting users: (1) to limit mobile apps' access to sensitive resources, and (2) to assist users in making trust decisions before installing apps. Android apps can only access sensitive resources if they declare permissions in their manifest files and get approved by users during the installation time. On the official Android Market, before installing an app, users are shown a permission screen listing the resources an app will access. Users can choose to either install the app with all the requested permissions or not to install the app at all. Once granted, permissions cannot be revoked unless users uninstall the app.

There have also been several user studies looking at usability issues of permission systems in warning users before downloading apps. Kelley et al. [26] conducted semi-structured interviews with Android users, and found that users paid limited attention to permission screens, and had poor understanding of what these permissions imply. Permission screens generally lack adequate explanation and definitions. Felt et al. [18] found similar results from Internet surveys and lab studies that current Android permission warnings do not help most users make correct security decisions.

Our work leverages this past work investigating Android's permissions. We extend their ideas in two new ways. The first is using crowdsourcing as a way of measuring people's expectations regarding an app's behavior, rather than relying solely on automated techniques. This allows us to capture a new aspect of mobile app privacy that past work has not. The second is the design and evaluation of a new privacy summary interface that emphasizes access to sensitive resources that people did not expect.

Mobile Application Analysis and Security Extensions

Researchers have also developed many useful techniques and tools to detect the sensitive information leakage in mobile apps [3, 10, 12-16, 19, 35, 36], by using permission analysis (e.g. [3, 16]), static code analysis (e.g. [12]), network analysis (e.g. [35]), or dynamic flow analysis (e.g. [14]). Their results identified the strong penetration of ads and analytics libraries, and other prevailing privacy violations including excessively accessing sensitive information. We used TaintDroid [14] in our work to investigate the ground truth of the top 100 popular Android apps on how and for what purpose sensitive resources were used. Amini et al. [2] offered an vision of an cloud-based service that leverages crowdsourcing and traditional security approaches to analyze mobile applications. Our work follows this vision and demonstrates the feasibility of incorporating crowdsourcing in application analysis.

Many security extensions have been developed to harden privacy and security. MockDroid [6], TISSA [39] and AppFence [22] substitute fake information into API calls made by apps, such that apps could still function but with zero disclosure of users' private information. Nauman et al. [28] proposed Apex which provided more fine-grained control over the resources usage based on context and runtime constraints. To enable wide deployment, Jeon et al. proposed an alternative solution that rewrote the bytecode of mobile apps to enforce more privacy controls [24] instead of modifying the Android system as the previous solutions.

Though app analysis provides us with a better understanding of apps' behaviors, it cannot infer people's perceptions of privacy or distinguish between behaviors which are necessary for an app's functionality versus behaviors which are privacy-intrusive. Similarly, while the security extensions above provide users with more control over their private data, it is unclear if lay users can correctly configure these settings to reflect their real preferences. Our work complements this past work by suggesting an alternative way of looking at mobile privacy from the users' perspective. We study users' mental models of mobile privacy, aiming to identify the most pertinent information to help users make better privacy-related trust decisions.

Expectations of Privacy, Mental Model Studies and Privacy Interface Design

The notion of expectations is fairly common in discussions of privacy [33]. For example, in Katz v. United States, Supreme Court put forward "reasonable expectation of privacy" to test reasonableness of legal privacy protections under the Fourth Amendment [1]. Palen and Dourish [30] and Barth et al. [4] discussed how expectations are governed by norms, past experiences, and technologies. Our notion of *privacy as expectations* is a narrower construct, focusing primarily on people's mental models of what they think an app does and does not do. Our core contribution is in operationalizing privacy in this manner, in terms of using crowdsourcing to capture people's expectations as well as reflecting the crowd's expectations directly in a privacy summary to emphasize places where an app's behavior did not match people's expectations.

Past work has looked at understanding people's mental models regarding computer security. For example, Camp [9] discussed five different high-level metaphors for how people think about computer security. Wash [38] identified eight mental models ('folk models') of security threats that users perceived and how these models can justify why users ignored security advice. Bravo-Lillo et al. [8] conducted studies to explore the psychological processes of users involving perceiving and responding to computer alerts. Sadeh et al. also studied the complexity of people's location sharing privacy preferences [5, 32]. This past research has a similar flavor as ours in terms of trying to understand the mental models people used to make trust decision. Our work extends this past work to a new domain, namely mobile app privacy.

Kelley et al. proposed simple visualizations called "privacy nutrition labels" [25] to inform user how their personal information is collected, used and shared by a web site. Our new proposed mobile privacy summary interface is inspired by their work. Our work differs in how we acquire privacy-related information. In their work, the expectation is that a 'nutrition label' would be generated by the owner of the web site. In our case, information is gathered through both crowdsourcing users' mental models and profiling mobile apps using dynamic taint analysis (e.g. using TaintDroid).

CROWDSOURCING USERS' MENTAL MODELS

In this section, we present the design and results of our study using crowdsourcing to capture users' mental models about a mobile app's behavior.

Taking a step back, there are four reasons why crowdsourcing is a compelling technique for examining privacy. Past work has shown that few people read End-User License Agreements (EULAs) [20] or web privacy policies [23], because (a) there is an overriding desire to install the app or use the web site, (b) reading these policies is not part of the user's main task (which is to use the app or web site), (c) the complexity of reading these policies, and (d) a clear cost (i.e. time) with unclear benefit. Crowdsourcing nicely addresses these problems. It dissociates the act of examining permissions from the act of installing apps. By paying participants, we make reading these policies part of the main task and also offer clear monetary benefit. Lastly, we can reduce the complexity of reading Android permissions by having participants examine just one permission at a time rather than all of the permissions, and by offering clearer explanations of what the permission means.

Study Design

We recruited participants using Amazon's Mechanical Turk (AMT). We designed each Human Intelligence Task (HIT) as a short set of questions about a specific Android app and resource pair (see Figure 1). Participants were asked to read the provided screenshots and description of an app, as retrieved from the official Android market. Then they were asked if they have used this app before and what category this app belongs to. The categorization questions were designed as an easy check to detect if participants were gaming our system (e.g., clicking through HITs without answering questions).

After these two questions, participants were shown one of two sets of follow-up questions. One of the conditions (referred to as *the expectation condition*) was designed to capture users' perceptions of whether they expected a given app to access a sensitive resource and why they thought the app used this resource. Participants were also asked to specify how comfortable they felt letting this app access the resource, using a 4-point Likert scale ranging from very comfortable (+2) to very uncomfortable (-2). In the other condition (referred to as *the purpose condition*), we wanted to see how people felt when offered more fine-grained information. Participants were told that a certain resource would be accessed by this app and given specific reasons, e.g. user's location information is accessed for target advertising. We identified these reasons by examining TaintDroid logs and using knowledge about ad

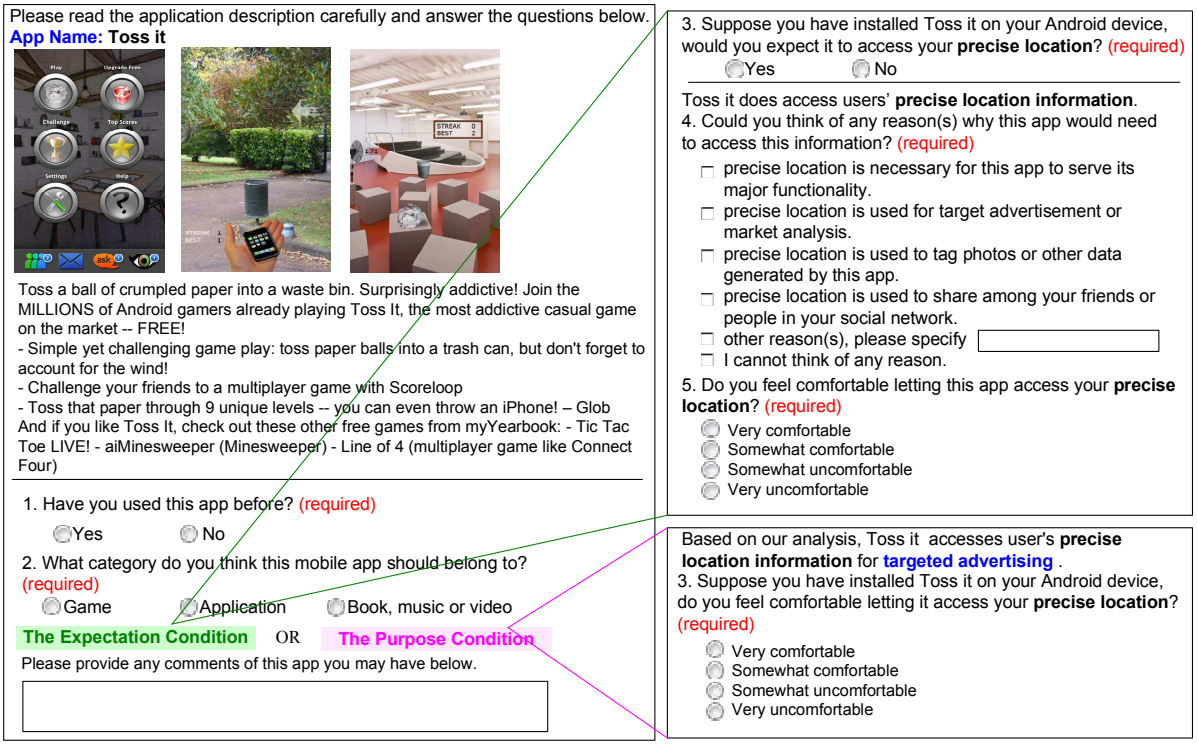

Figure 1. Sample questions in our study to capture users' mental models. Participants were randomly assigned to one of the conditions. In the *expectation condition*, participants' were asked to specify their expectations and speculate the purpose for this resource access. In *the purpose condition*, the purpose of resource access was given to participants. In both conditions, participants were asked to rate how comfortable they felt having the targeted app access their resources.

networks. Participants were then asked to provide their comfort ratings as in the expectation condition. Finally, participants from both conditions were encouraged to provide optional comments on the apps in general. The separation of the two conditions let us compare users' perceptions and subjective feelings when different information was provided.

We focused our data collection on four types of sensitive resources (as suggested by AppFence [22]): unique device ID, contact list, network location, and GPS location. We also restricted the pool of apps to the Top 100 most downloaded mobile apps on the Android market. Overall, 56 of these apps requested access to unique phone ID, 25 to the contact list, 24 to GPS location, and 29 to Network Location. This resulted in 134 app and resource pairs, i.e. 134 distinct HITs. For each HIT, we recruited 40 unique participants to answer our questions (20 per condition).

We used the following qualification test to limit our participants to Android users, as well as to filter out people who were not serious. Crowd participants were asked to provide the Android OS version of their device, with instructions on where to find this information on their Android devices. When reviewing participants' qualification requests, we also randomly assigned qualified participants to different conditions by giving them different qualification scores. In this way, we could ensure a between-subject design where a participant would only be exposed to one condition.

To prevent other confounding factors such as cultural or language issues, we restricted our participants to those who were located within the U.S. To guarantee the quality of our data, we also required participants to have a lifetime approval rate higher than 75% (i.e. the rate of successfully completing previous tasks).

All the HITs of this study were completed over the course of six days. We collected a total of 5684 responses. 211 were discarded due to incomplete answers, and 113 were discarded due to failing the quality control question, yielding 5360 valid responses. There were 179 verified Android users in our study, with an average lifetime approval rate of 97% (SD=8.79%). The distribution of Android versions our participants used was very close to Google's official numbers [37]. On average, participants spent about one minute per HIT (M=61.27, SD=29.03), and were paid at the rate of $0.12 per HIT.

The Most Unexpected and the Most Uncomfortable

Our first analysis looked at what sensitive resource usages were least expected by users based on data from the expectation condition. For each app and resource pair, we aggregated the data by calculating the percentage of participants who expected the resources to be accessed, and averaging the self-reported comfort ratings (ranging

from very comfortable +2.0 to very uncomfortable -2.0). Table 1 summarizes the resource usages that less than 20% of participants said that they expected. For example, only 5% of participants expected the Brightest Flashlight app would access users' network location information, and overall, participants felt uncomfortable about this resource usage (M= -1.25, SD=0.39). Similarly, only 10% of participants expected the Talking Tom app would access users' device ID, and 20% of people expected Pandora to access their contact list.

Generally speaking, when participants were surprised by an access to a sensitive resource, they also found hard to explain why this resource were needed. Note that in the expectation condition, participants were only informed about which resources were accessed; they were not informed about the purpose of why these resources were accessed. This is similar to what the existing Android permission list conveys to users. In this condition, we observed a very strong correlation (r= 0.91) between the percentage of expectations and the average comfort ratings. In other words, the perceived necessity of the resource access was directly linked to their subjective feelings, thus guiding the way users make trust decisions on mobile apps. As many participants also mentioned in their comments, these surprises prompted them to take different actions. For example, participant W27 said about Brightest Flashlight app, "Why does a flashlight need to know my location? I love this app, but now I know it access my location, I may delete it." W92 said, "I didn't know Pandora can read my phone book. But why? Can I turn it off? I'll search for other internet radio app." Similarly, W56 showed a similar concern (for the Toss It game), "I do not feel that games should ever need access to your location. I will never download this game."

Lay Users Have a Hard Time Identifying the Reason an App Accesses a Resource

Another way to look at the expectation condition is that it presented users with information comparable to what is provided by the Android permission system, namely what resources may be accessed. We wanted to see to what extent people understand the behaviors of apps in this optimal case, where they were paid to read the privacy summaries. Based on our results, even if users were fully aware of which resources were used, they still had a hard time understanding why these resources were needed.

We used TaintDroid [14] to analyze all the mobile apps in our study to identify the actions that triggered the sensitive resource access and where the sensitive information was sent to. We then manually categorized each app and resource pair into three categories: (1) for major functionality, (2) for sharing and tagging (or supporting other minor functions), (3) for target advertising or market analysis. Many resource usages fell into more than one category. For example, the WeatherBug application uses location for retrieving local weather information as well as for targeted advertising.

Resource	App name	% Expected	Avg Comfort
Network Location	Brightest Flashlight	5%	-1.25
	Toss It	10%	-1.15
	Angry Birds	10%	-0.43
	Air Control Lite	20%	-0.55
	Horoscope	20%	-1.05
GPS Location	Brightest Flashlight	10%	-0.95
	Toss It	5%	-0.95
	Shazam	20%	-0.05
Device ID	Brightest Flashlight	5%	-1.35
	TalkingTom Free	10%	-0.78
	Mouse Trap	15%	-0.85
	Dictionary	15%	-0.69
	Ant Smasher	20%	-1.13
	Horoscope	20%	-1.03
Contact List	Backgrounds HD Wallpapers	10%	-1.35
	Pandora	20%	-0.70
	GO Launcher EX	20%	-0.75

Table 1. The most unexpected resource usages identified in the expectation condition, i.e. resource usage expected by no more than 20% of participants. Users felt uncomfortable with these unexpected app behaviors. For each app and resource pair, 20 participants were surveyed. The comfort rating was ranging from -2.0 (very uncomfortable to +2.0 (very comfortable). For all the apps we surveyed, there was a strong correlation (r=0.91) between people's expectation and their subjective feelings.

We compared the reasons our participants provided in the expectation condition against the ground truth from our analysis as shown in Table 2. In most cases, the majority of participants could not correctly state why a given app requested access to a given resource. When the resources were accessed for functionality purposes, participants generally had better answers; however, the accuracy never exceeded 80%. When sensitive resources were used for multiple purposes, the accuracies tended to be much lower. We also note that, participants had slightly better answers of why their location information was needed compared to the other two types of sensitive resources.

Note that, these results are for the situation where participants were paid to carefully read the description. Many of them had even already used some of these apps before. We believe for general Android users, their ability to guess would be even worse. This also indicates that simply informing users of what resources are used (as today's Android permission screen does) is not enough for users to make informed decision.

Clarifying the Purpose May Ease Worries

Given the lack of clarity of why their resources are accessed, users have to deal with significant uncertainties when making trust decisions regarding installing and

Resource Type	Resource used for [1] Major functionality [2] Tagging or sharing [3]Advertising or market analysis	cnt	% of accurate guess	% of no idea
Contact List (25)	[1]	20	56%	8%
	[2]	2	28%	35%
	[1]+[2]	2	19%	16%
	[1]+[2]+[3]	1	27%	14%
GPS Location (24)	[1]	14	74%	11%
	[2]	4	80%	10%
	[3]	2	35%	55%
	[1]+[3]	3	15%	27%
	[2]+[3]	1	15%	40%
Network Location (29)	[1]	15	77%	8%
	[2]	2	55%	10%
	[3]	7	29%	63%
	[1]+[3]	3	15%	22%
	[2]+[3]	2	13%	25%
Device ID (56)	[1]	1	51%	29%
	[3]	30	22%	58%
	[1]+[3]	12	7%	55%

Table 2. Participants had a difficult time speculating on the purposes of their sensitive resource usages. The first column shows the type of resource accessed and the total number of apps accessing that resource. The second column shows the ground truth of why the resource is accessed, the third column shows the number of apps in each category (e.g. 20 apps access contact list for reason [1]). The third column shows the percentage of participants stated the purpose correctly. The last column shows the percentages of participants who had no idea why the resource is accessed.

using a given mobile app. We wanted to see if providing users with more fine-grained information, especially the purposes of resource access, would have any influence on users' privacy-related subjective feelings. To answer this question, we compared the average comfort ratings from both conditions, for each mobile app and resource pair.

We observed that for all four types of sensitive resources (i.e. device ID, contact list, network location, and GPS location), participants felt more comfortable when they were informed of the purposes of a resource access (see Table 3). The differences between the comfort ratings were statistically significant in t-tests. For example, with regard to accessing the device ID, the average comfort rating in the purpose condition was 0.3 higher than in the expectation condition ($t(55)=7.42$, $p<0.0001$). For some apps, informing people of the purpose led to totally different feelings. For example, participants felt uneasy when told the Dictionary app accessed their network location ($M_{comfort}$= -0.83, SD=0.41). However, when they were informed that the location was only used to search for trending words that people nearby are looking up, they felt much less concerned ($M_{comfort}$=0.80, SD=0.29). Similarly, Air Control Lite, eBuddy, Shazam, Antivirus, and other 7 apps all demonstrate a significant increase

Resource Type	comfort rating w/ purpose	comfort rating w/o purpose	df	T	p
Device ID	0.47(0.30)	-0.10(0.41)	55	7.42	0.0001
Contact List	0.66(0.22)	0.16(0.54)	24	4.47	0.0002
Network Location	0.90(0.53)	0.65(0.55)	28	3.14	0.004
GPS Location	0.72(0.62)	0.35(0.73)	23	3.60	0.001

Table 3. Comparison of comfort ratings between the expectation condition (2nd column) and the purpose condition (3rd column). Standard deviations are shown between parentheses. When participants were informed of the purpose of resource access, they generally felt more comfortable. The differences were statistically significant for all four types of resources. The comfort ratings were ranging from -2.0 (very uncomfortable to +2.0 (very comfortable).

($\delta>1.0$) in comfort rating when the purpose of a resource access was explained.

This finding suggests that providing users with the reasons why their resources are used not only gives them more information to make better trust decisions, but can also ease their concerns caused by uncertainties. Note that informing users about the "purpose" for collecting their information is a common expectation in many legal and regulatory privacy frameworks. Our results confirm the importance of this information. This finding also provides us with strong rationale for including the purpose(s) of resource access in our new design of privacy summary interface.

Impact of Previously Using an App

We also wanted to see how previous experiences with an app impacted participants' expectations and level of comfort. To answer this question, we compared the responses between participants who had and hadn't used the app before. The ratio of people who had and had not used the apps in our study varied greatly. Some apps (such as Facebook and Twitter) saw high usage among our participants, while others (such as Kakao Talk Messenger and Horoscope) had fairly low usage. To make the comparison fair, we only examined apps that had at least 5 responses in both the used and not used categories. In our data, the differences between participants who had and had not used these apps before were not statistically significant with respect to their expectation of sensitive resource access. Regarding their comfort level, the only significant difference we observed is the average comfort ratings for accessing the contact list. Participants who used an app before felt more comfortable letting that app access their contact list ($t(20)=2.68$, $p=0.015$). For the other three types of resources, the experiences with apps didn't cause any statistically significant differences in participants' subjective feelings.

This finding suggests that people who use an app do not necessarily have a better understanding of what the app is actually doing, in terms of accessing their sensitive resources. It also suggests that, if we use crowdsourcing to capture users' mental models of certain apps, we do not have to restrict our participants to people who are already familiar with these apps, allowing us access to a potentially larger crowd.

NEW PRIVACY SUMMARY INTERFACE

In the previous section, we had identified that purpose and expectation are two key factors that impact users' subjective feelings. Based on this finding, we present the design of a new privacy summary interface highlighting the purposes of sensitive resource usage and people's perceptions about app's behaviors.

Design Rationale

Privacy summary interfaces, such as the permission screen in current Android, are designed for users to review before downloading mobile apps. By that time, users have limited information to form their mental model of the targeted mobile app since they haven't had any interaction with it. In contrast with our crowdsourcing study, we cannot rely on general users to carefully examine an app's description or screenshots to understand how this app works in reality. In our new design, we directly leverage other users' mental models. The underlying rationale is similar to the idea of Patil et al. [31] in the sense of incorporating others' opinions in making privacy decisions. Our work differs from their work by aggregating users' subject feedback from crowds instead of from one's social circle and highlighting users' surprises. By presenting the most common misconceptions about an app, we can rectify people's mental models and help them make better trust decisions. We consider users' *expectations* and the *purposes of resource access* as the two key points that we want to convey to users in our new summary interface.

Previous research has discussed several problems with the existing Android permission screens [18, 26], including:

- The wording of the permission list contains too much technical jargon for lay users.
- They offer little explanations and insight into the potential privacy risk.
- A long list of permissions make users experience warning fatigue.

With these problems in mind, in addition to the two identified key features, we proposed several principles for our own design:

- Using simple terms to describe the relevant resources; e.g., instead of using "coarse (Network) location", we use the term "approximate location".
- Only displaying the resources that have greater impact on users' privacy, such as location, device ID, storage, contact list etc. Users could choose to check out other low-risk resources by clicking "See all".
- Sorting the list based on expectation as captured through crowdsourcing. We order the list so that the more surprising resource usages are shown first.
- Highlighting important information. We bold the sensitive resources mentioned in text, and use warning sign and striking color to highlight the suspicious resource usages, i.e. when the surprise value exceeds a certain threshold.

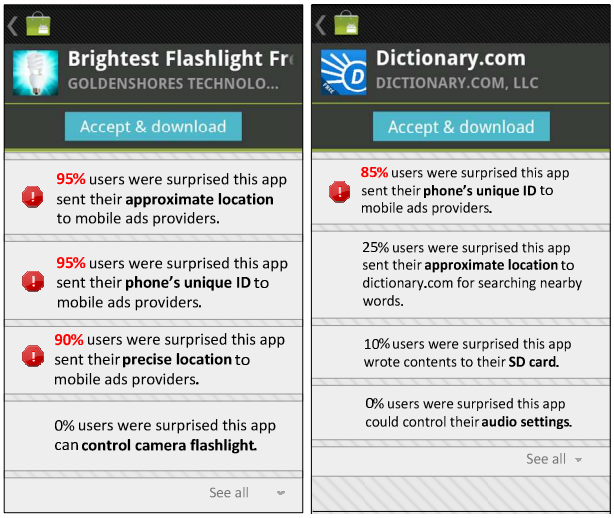

Figure 2: A mockup interface of our newly proposed privacy summary screen, taking the Brightest FlashLight and the Dictionary app as examples. The new interface provides extra information of why certain sensitive resources are needed and how other users feel about the resource usages. Warning sign will appear if more than half of the previous users were surprised about this resource access.

Figure 2 shows two examples of our new privacy summary interface. To make the comparison more symmetric, our design uses the same background color and pattern are used in the current Android permission screen. The surprise numbers (i.e. "n% of users were surprised") used in these mockups were obtained from our crowdsourcing study where possible. The surprise numbers for other resources (such as camera flashlight, SD card) were reasonable estimates made by our team.

Evaluation

We used AMT to conduct a between-subject user study to evaluate our new privacy summary interface. Participants were randomly assigned to one of the two conditions in the same way as our previous study. In *the permission condition*, participants were shown the permission screen that the current Android Market uses; in the other condition (referred as *the new interface condition*), participants were shown our new interfaces. We used the data we collected in our previously described crowdsourcing study to mock up the privacy summary

* p <0.05 ** p<0.005	# of People Mentioning Privacy Concerns (out of 20)		Accuracy (max=1.0)			Time spent (sec)		
App Name	Permission	New Interface	Permission	New Interface	p	Permission	New Interface	p
Brightest Flashlight	4	6	0.58	0.86	**	74.59	65.11	
Dictionary	1	3	0.73	0.91	**	68.21	43.92	**
Horoscope	3	7	0.75	0.95	*	68.41	48.72	*
Pandora	3	3	0.68	0.94	**	76.86	76.82	
Toss it	4	13	0.61	0.88	**	67.43	57.10	

Table 4. Comparisons between the existing Android permission screen (permission condition) and our newly proposed privacy summary (new interface condition). Our new interface makes users more aware of the privacy implications and is easier to understand. Users in general spent less time on these newly proposed interfaces but got more fine-grained information.

interfaces for five mobile apps, namely Brightest Flashlight, Dictionary, Horoscope, Pandora, and Toss it.

In both conditions, the app's name, screenshots, description and the quality control question were presented the same way as in previous study. The privacy summary was then shown (either the current permission screen or our newly proposed interface). Participants were asked whether they would recommend this app to a friend who might be interested in it, and why (or why not). We used JavaScript to keep track of the time participants spent on reading the privacy summary before making their recommendation choices. After this question, privacy summary screens were covered by grey rectangles. Participants could recheck the privacy summaries by moving their mice over the grey rectangles. In this way, we could accurately record the additional time participants spent on viewing privacy summary screens by monitoring the mouse hovering events. We then added up all these time fragments to compute the total time participants spent on reading the privacy summary. Participants were tested on their understanding of the presented privacy summary screen by specifying the resource(s) usages suggested by the privacy summary.

For each condition per app, 20 unique participants were recruited. Participants could evaluate multiple apps within the same condition. A total of 237 responses were submitted, 19 of which were discarded due to incompletion and 18 of which were discarded due to failing the quality control question. Sixty-seven Android users participated in this study with an average lifetime approval rate of 96.31% (SD=6.27%). Thirty-five participants were assigned to the permission condition, and thirty-two were assigned to the new interface condition. Participants on average spent 2 min and 41.4 sec (SD=77.3 sec) in completing each evaluation task, and were paid at the rate of $0.20/HIT.

We evaluated the new privacy summary interface from three perspectives to test its effectiveness and usability. The first is *privacy awareness*, i.e. whether users are more aware of the privacy implications. This is measured by counting the number of participants who mentioned privacy concerns when justifying their recommendation decisions. The second is *comprehensibility*, i.e. how well users understood the privacy summary. This is measured by the accuracy in answering questions about the app's behavior. The third is *efficiency*, i.e. how long it took participants to understand the privacy summary, measured by the number of seconds they spent on reading the privacy summary screens.

The comparisons between the two conditions are summarized in Table 4. Generally speaking, participants in the new interface condition weighted their privacy more when they made decisions about whether the app was worth recommending. More people in this condition mentioned privacy-related concerns when they were justifying their choices. When we asked people in both conditions to specify the resources used by the target apps of the target apps, people in the new interface condition also demonstrated a significantly higher accuracy compared to their counterparts. Furthermore, except for the Pandora app, participants in the new interface condition on average spent less time reading the privacy summaries on average, though the time difference was not always statistically significant. This finding suggests that we can provide more useful information without requiring users to spend more time to understand it.

In our future work, we plan to conduct lab studies to evaluate our new privacy summary interface in depth. We will focus on the effectiveness of the new interface when users only look at it briefly (e.g. for 5-10 secs), since in reality general users are not likely to devote a lot of time to reading.

DISCUSSION

In this section, we discuss the potential implications of our work and how it fit into our vision of leveraging crowdsourcing for application analysis.

Implications for Privacy Analysis

A Potential Win-Win A major finding of our work is that users feel more comfortable when they are informed of the reasons why their sensitive resources are needed. In some cases, it might be again tied to users' expectations. For example, the "trending, popular and nearby search" functionality provided by the Dictionary app uses location information to retrieve the words that people nearby are looking up. It is a relatively minor function of this app and may not be expected even for users who are familiar

with this app. Therefore, when we asked participants to state the reasons for accessing location information, most of them thought it was for targeted advertising purpose, hence rating the comfort level much lower than they were informed about the actual reason. We also observed several cases (e.g. the Weather Channel, GasBuddy, Compass) where participants had correct answers as to why the app was using one's location, but still felt less comfortable when compared to the condition where participants were directly given the purpose. It suggests that when dealing with uncertainties, users tend to be more concerned or even paranoid about their privacy. Our results provide evidence that properly informing users with the purposes of resource usage can actually ease their worries. In other words, it would potentially benefit all parties, including app developers, market owners, and advertisers.

Currently, the default Android permission screen doesn't contain any explanations. One possible approach for getting this information is to scale up our crowdsourcing approach, but there is the potential for errors, as we saw in Table 2. Another approach is to require app developers to include a rationale, but this is an optimistic approach assuming that developers won't lie. This also suggests that better tools are still needed for analyzing apps' behaviors in a more scalable and automated manner, as envisioned by Amini et al. [2].

Privacy Concerns of Mobile Advertising We observed that mobile advertising services were a consistent privacy concern for the most participants. For all four types of resources, users felt the least comfortable when they were used for advertising or market analysis. We understand that many developers rely on ads for income. However, there is still space for app developers and ad networks to improve the user experience, such as by providing users with more informed consent and more explanations on how and why their personal information is used. Other potential ways include tweaking the sensitive resource usage to a coarser level, or using hashing or other methods to conceal users' identities. These technical methods can address users' privacy concerns without sacrificing too much on the ads' quality.

Leveraging Crowd for Application Analysis

The long term vision of our work is to design a scalable privacy evaluation system for mobile apps by combining automated application analysis with crowdsourcing techniques. The automated techniques are meant to capture an app's behaviors involving sensitive resources, whereas the crowdsourcing techniques capture people's perceptions and expectations about an app's behaviors.

One important contribution of this paper is to demonstrate the feasibility of using crowdsourcing to capture users' perceptions, and to identify the strength and weakness of the crowd in evaluating privacy. Based on our data, users were not very good at speculating on the purpose of resource access, which is not surprising and might be compensated by leveraging existing mobile app analysis techniques. However, specifying their expectations is a relatively easy job for most people but cannot be addressed by existing app analysis tools.

As the first work of this kind, we simplified the problem by focusing only on privacy, although we realize that users may weigh utility over privacy when making decisions about installing an app. Future research will need to take utility into account in understanding how people make trust decisions.

We also only captured people's perceptions at a coarse granularity and with limited types of sensitive resources. We will extend our work to finer-grained interactions, e.g. whether users expect the Yelp app to send their location to yelp.com when they press 'Search nearby restaurant' button. We envision that this level of analysis could provide us more detailed information for evaluating mobile apps, and could possibly lead to better results when asking the crowd why an app accesses a given resource.

In our crowdsourcing study, it cost us $2.40 and about 20-25 minutes (deducted from the effective hourly rate reported by AMT) to examine one app and resource pair with input from 20 participants. There is ample room to improve the crowdsourcing efficiency. Examples include extending the participant pool to all smartphone users, minimizing the number of questions, and so on. There are also several techniques suggested by previous crowdsourcing work [7, 27] that we can leverage to improve the overall efficiency, e.g. dynamically publishing HITs, adaptively adjusting the compensation rate and the number of required responses. Given that it only took about one minute for our participants to complete a crowdsourcing task, we believe this method would scale well, though formal scalability analysis is still an open issue and will be included in our future work.

Alternatively, crowdsourcing users' perceptions could be achieved in conjunction with the exiting app rating mechanism. When users rate a mobile app, they can also optionally specify their expectations of one aspect of the target app. As the number of rating grows, the aggregated perceptions will be more representative.

CONCLUSION & FUTURE WORK

A great deal of past work in mobile security and privacy research has focused on providing tools for automated analysis. However, there is still no easy way to distinguish whether accessing certain sensitive resource is necessary, or how that action makes users feel with respect to their privacy. Our work demonstrates a new way for evaluating mobile app's privacy. We explore users' mental models of mobile privacy by crowdsourcing

users' expectations of mobile apps' sensitive resource usage. Our results suggest that both users' expectation and the purpose of why sensitive resources are used have a major impact on users' subjective feelings and their trust decisions. Another major finding is that properly informing users of the purpose of resource access can ease users' privacy concerns to some extent. Based on our findings, we proposed a new privacy summary interface that highlights common misconceptions that other users have and the purpose of a resource access. Compared to the existing Android permission screen, our interface is much easier to understand and provides users with more pertinent information for users to make better trust decision.

ACKNOWLEDGEMENT

This research was supported by CyLab at Carnegie Mellon under grants DAAD19-02-1-0389 and W911NF-09-1-0273 from the Army Research Office and by Google. Support was also provided by the National Science Foundation under Grants CNS-1012763 and CNS-0905562.

REFERENCES

[1]"Katz v United States 389 U.S. 347." Available: http://en.wikipedia.org/wiki/Katz_v._United_States

[2]S. Amini, *et al.*, "Towards Scalable Evaluation of Mobile Applications through Crowdsourcing and Automation," CMU-CyLab-12-006, Carnegie Mellon University, 2012.

[3]D. Barrera, *et al.*, "A methodology for empirical analysis of permission-based security models and its application to android," In Proc. *CCS*, 2010.

[4]A. Barth, *et al.*, "Privacy and Contextual Integrity: Framework and Applications," In Proc. *IEEE Symposium on Security and Privacy*, 2006.

[5]M. Benisch, *et al.*, "Capturing location-privacy preferences: quantifying accuracy and user-burden tradeoffs," *Personal and Ubiquitous Computing,* 2010.

[6]A. Beresford, *et al.*, "MockDroid: trading privacy for application functionality on smartphones," In Proc. *HotMobile*, 2011.

[7]M. S. Bernstein, *et al.*, "Soylent: a word processor with a crowd inside," In Proc. *UIST*, 2010.

[8]C. Bravo-Lillo, *et al.*, "Bridging the gap in computer security warnings: a mental model approach," *IEEE Security & Privacy Magazine,* 2010.

[9]L. J. Camp, "Mental models of privacy and security," *Technology and Society Magazine, IEEE,* vol. 28, 2009.

[10]E. Chin, *et al.*, "Analyzing inter-application communication in Android," In Proc. *MobiSys*, 2011.

[11]K. Craik, *the nature of explanation,* Cambridge University Press, 1943.

[12]M. Egele, *et al.*, "PiOS: Detecting Privacy Leaks in iOS Applications," In Proc. *NDSS*, 2011.

[13]W. Enck, "Defending Users against Smartphone Apps: Techniques and Future Directions," in *LNCS*. vol. 7093, ed, 2011.

[14]W. Enck, *et al.*, "TaintDroid: An Information-Flow Tracking System for Realtime Privacy Monitoring on Smartphones," In Proc. *OSDI* 2010.

[15]W. Enck, *et al.*, "A Study of Android Application Security," In Proc. *USENIX Security Symposium*, 2011.

[16]A. P. Felt, *et al.*, "Android permissions demystified," In Proc. *CCS*, 2011.

[17]A. P. Felt, *et al.*, "A survey of mobile malware in the wild," In Proc. *SPSM*, 2011.

[18]A. P. Felt, *et al.*, "Android Permissions: User Attention, Comprehension, and Behavior," UCB/EECS-2012-26, University of California, Berkeley, 2012.

[19]A. P. Felt, *et al.*, "Permission re-delegation: attacks and defenses," In Proc. *USENIX conference on Security*, 2011.

[20]N. Good, *et al.*, "Stopping spyware at the gate: a user study of privacy, notice and spyware," In Proc. *SOUPS*, 2005.

[21]S. Grobart. "The Facebook Scare That Wasn't." Available: http://gadgetwise.blogs.nytimes.com/2011/08/10/the-facebook-scare-that-wasnt/

[22]P. Hornyack, *et al.*, "These aren't the droids you're looking for: retrofitting android to protect data from imperious applications," In Proc. *CCS*, 2011.

[23]C. Jensen and C. Potts, "Privacy policies as decision-making tools: an evaluation of online privacy notices," In Proc. *CHI*, 2004.

[24]J. Jeon, *et al.*, "Dr. Android and Mr. Hide: Fine-grained security policies on unmodified Android," 2012.

[25]P. G. Kelley, *et al.*, "A "nutrition label" for privacy," In Proc. *SOUPS*, 2009.

[26]P. G. Kelley, *et al.*, "A Conundrum of permissions: Installing Applications on an Android Smartphone," In Proc. *USEC*, 2012.

[27]G. Liu, *et al.*, "Smartening the crowds: computational techniques for improving human verification to fight phishing scams," In Proc. *SOUPS*, 2011.

[28]M. Nauman, *et al.*, "Apex: extending Android permission model and enforcement with user-defined runtime constraints," In Proc. *ASIACCS*, 2010.

[29]D. Norman, *The design of everyday things*: Basic Books, 2002.

[30]L. Palen and P. Dourish, "Unpacking "privacy" for a networked world," In Proc. *CHI*, 2003.

[31]S. Patil, *et al.*, "With a little help from my friends: can social navigation inform interpersonal privacy preferences?," In Proc. *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 2011.

[32]N. Sadeh, *et al.*, "Understanding and Capturing People's Privacy Policies in a Mobile Social Networking Application," *The Journal of Personal and Ubiquitous Computing,* 2009.

[33]D. J. Solove, "A Taxonomy of Privacy," *University of Pennsylvania Law Review, Vol. 154, No. 3, January 2006.*

[34]A. Thampi. "Path uploads your entire iPhone address book to its servers." Available: http://mclov.in/2012/02/08/path-uploads-your-entire-address-book-to-their-servers.html

[35]S. Thurm and Y. I. Kane, "Your Apps are Watching You," *WSJ,* 2011.

[36]T. Vidas, *et al.*, "Curbing android permission creep," *Proceedings of the Web,* vol. 2, 2011.

[37]A. Wagner. "Google Posts Refreshed Android Distribution Numbers." Available: http://www.twylah.com/surfingislander/tweets/177040176181288960

[38]R. Wash, "Folk models of home computer security," In Proc. *SOUPS*, 2010.

[39]Y. Zhou, *et al.*, "Taming Information-Stealing Smartphone Applications (on Android)," In Proc. *TRUST*, 2011.