N Heads Are Better Than One

Morris Hopkins, Mauricio Castaneda, Swapneel Sheth, Gail Kaiser,
Department of Computer Science
Columbia University
New York, NY, USA
{mah2250, mc3683}@columbia.edu, {swapneel, kaiser}@cs.columbia.edu

ABSTRACT

Social network platforms have transformed how people communicate and share information. However, as these platforms have evolved, the ability for users to control how and with whom information is being shared introduces challenges concerning the configuration and comprehension of privacy settings. To address these concerns, our crowd sourced approach simplifies the understanding of privacy settings by using data collected from 512 users over a 17 month period to generate visualizations that allow users to compare their personal settings to an arbitrary subset of individuals of their choosing. To validate our approach we conducted an online survey with closed and open questions and collected 59 valid responses after which we conducted follow-up interviews with 10 respondents. Our results showed that 70% of respondents found visualizations using crowd sourced data useful for understanding privacy settings, and 80% preferred a crowd sourced tool for configuring their privacy settings over current privacy controls.

Categories and Subject Descriptors

K.4.1 [**Public Policy Issues**]: Privacy; H.5.2 [**User Interfaces**]: User-centered design; H.3.3 [**Information Search and Retrieval**]: Information filtering

General Terms

Human Factors

Keywords

privacy, empirical studies, crowdsourcing, data visualization, social networking, user study

1. INTRODUCTION

As social network platforms, such as Facebook, have increased the amount of personal data they contain, the ease with which individual users can understand and control their own data is a growing concern [21]. Current privacy settings are often complicated and difficult to navigate [8]. For example, Facebook obfuscates privacy settings by splitting privacy controls across several seemingly unrelated portions of the application. The problem is made worse by occasional unannounced changes that are instituted without giving users the option to opt-out before the change takes effect [6].

Previous research efforts have focused on how well users understand their privacy. Some of this research has shown a disparity between what users intend when configuring their privacy settings and the reality of what their settings represent [12, 13]. Other research has introduced techniques aimed at making privacy more understandable and easier to control. There have been several approaches, including introducing separate third-party software to store and manage a social network user's content independently of the social network service [19], providing access controls that enable social network users to segregate risky connections [4], and creating automated tools for classifying every connection into labeled groups and then applying label-based privileges to these groups [5]. However, none of these techniques allowed users to see and understand their privacy settings alongside the privacy settings from other individuals or peer groups. Because social networks are designed for sharing, a contextualized view of privacy (seeing one's privacy settings alongside the settings of others) is easier to understand than an isolated view. This is our approach. We use crowd sourced data to give users more information with which to make decisions on how they would like to set their privacy.

This paper aims to introduce a crowd sourced technique for improving social network users' understanding of privacy. To this end we conducted a study which included an online survey for which we received 59 valid responses. As part of the survey, we presented users with three alternative privacy control options: 1. a 3-option model (easy/medium/hard), 2. a survey based model, and 3. a crowd sourced model. After the survey, we performed follow-up interviews with 10 respondents in order to confirm survey responses with additional qualitative feedback, and to determine if user understanding of privacy could be improved with data visualizations. The visualizations used were produced with data collected from 512 Facebook users over a 17 month period.

The results of the survey show that users tended to prefer the 3-option and survey-based models. During the follow-up interviews users were presented with mock-ups of the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...$15.00.

three models and asked to rate them. The results show that users preferred all of the alternative models over the current controls, and in particular, the survey based and crowd sourced models were the most popular options. We believe the change-of-preference reflects the lack of user familiarity with crowd sourced models. Additionally, we found that 70% of users found data visualizations useful when configuring privacy settings.

The remainder of the paper is organized as follows: Section 2 describes our approach, section 3 describes the design of our study. Section 4-6 highlight the key results. Section 7 covers the implications of the results and limitations of the study. Related work is covered in section 8 and we conclude the paper in section 9.

2. APPROACH

The motivation for this study was to improve user understanding of privacy by addressing issues that currently exist with social network's privacy controls: persistence (how long a user-defined setting remains constant), validation (how easily a user is able to confirm their settings), and comparison (the ability for users to see their privacy settings alongside the settings of other users and groups). In this section, we outline the motivating research questions and the approach we selected to address these questions.

2.1 Research Questions

The goal of this study is to examine methods for improving user understanding of privacy settings. In particular, we want to determine if crowd sourced data can be used to simplify comprehension of privacy settings. The Oxford English Dictionary defines privacy as the "Absence or avoidance of publicity or display; secrecy, concealment, discretion; protection from public knowledge or availability [1]." We use the term *privacy* specifically as it relates to data privacy for information stored on social networks. We use the term *privacy settings* as a reference to the end-user configurable settings managing the accessibility of data stored on social networks. By *understand*, we mean the knowledge a user has about what information they are and are not sharing on a social network. We are not interested in whether or not users are aware of potential consequences that sharing information may entail.

We focused on the following research questions:

- **RQ 1:** Do users understand their current privacy settings? (Section 4)
- **RQ 2:** Can user understanding of privacy be improved using crowd sourced data? (Section 5)
- **RQ 3:** Are there tools for configuring privacy that are preferred over the currently provided tools? (Section 6)

2.2 Crowdsourcing Privacy

In order to improve user understanding of privacy through a crowd sourced approach, we created a set of visualizations that summarize a user's privacy settings and provide a comparative view with the settings from other users. These visualizations can be split into two different categories: Individual Visualizations, containing only data for a single user, and Contextualized Visualizations, showing a single users data in the context of the data of multiple users. The following subsections describe the dataset that was used and the visualizations that were created.

2.2.1 Dataset

Our dataset contains the information of 512 Facebook users, collected since May 1st 2012. The data was collected using an automated tool that retrieved users' privacy information from the Facebook Graph API. On average, there were a total of 154 scans made per user. The number of scans varied slightly due to API outages and incomplete or badly formatted API responses. Each scan has data representing the amount of information in each category that a user is sharing, such as number of mutual friends, number of check-ins, number of photos, etc., for a total of 41 different categories. By collecting data across all categories for a single user, we have a measure of how much information that particular user is sharing. Similarly, by combining the data collected across users, we have a measure of how much each category is being shared.

2.2.2 Individual Visualizations

The first visualization (Figure 1) in this group, is a line graph displaying the data from each category for an individual user across time. This view is valuable to a user because it allows them to confirm their settings have not changed unexpectedly (persistence), and get an overview of their settings (validation). The y-axis of the visualization represents the magnitude of the returned data. For example, the number of friends a user has. The x-axis represents time, with each increment marking a single scan of the user's data. Each line in the visualization represents a different category. The categories are set apart with different colors and distinct shapes for the anchors. When data was unavailable for a category the value was mapped to 0. When the Facebook Graph API returned an error, the value was mapped to -10. This visualization is helpful for identifying changes over time for the different categories, and provides users with a history and current snapshot of what information is being shared.

The second visualization (Figure 2) in this group is a word cloud. The word cloud provides users with a quick and easily understood view of what categories they are sharing (validation and persistence). The size of the font for each category represents the the magnitude of the shared content (how many friends a user has). Categories for which a user shares no data have the smallest font. If an API error was returned the text for that category is striked out. An example of this is shown with the "inbox" in Figure 2. While the word cloud visualization provides users with an easily understood display of their individual privacy settings, it could also be used to display the data of a group of users. In this case, the size of the font would represent the number of users sharing a category, and the strike out would indicate that no users share the category. This would allow users to quickly understand how common it is for an category to be shared (comparison).

2.2.3 Contextualized Visualizations

The first visualization (Figure 3) in this group is a series of donut charts. This visualization shows an individual users content in the context of a larger group of users, thereby allowing them to easily see how their settings compare to the settings of the members of the larger group. Each donut represents all of the data collected from the 512 Facebook users for single category. The blue portion of the donut represents the percentage of users that share data for the category, the orange represents the percentage that do not have data for the category, and the gray represents the percentage for which the Facebook Graph API returned an error. API errors typically indicated the category had been restricted as policy or for which the user had restricted sharing. For the individual using this visualization, we indicated whether or not they shared category by fading out the category if it was not shared, and additionally adding a color-coded dot at the top left of the donut. This visualization is useful when determining what are the trending privacy settings for a specific group (in this case, the entire set of data that was collected from the 512 over the 17 period).

The second visualization (Figure 4) in this group displays data about how the settings for a single category have changed over time, for multiple users, and highlights the individual user within this data. The y-axis represents the magnitude of the data being shared (e.g., the number of friends), and the x-axis represents the scan number (e.g., 50 represents the 50th scan). Each line represents a single users data. The highlighted line represents the individuals users data among the larger group. Again, for cases where there was no data returned the value was mapped to 0, and for cases where an API error occurred the value was mapped to -10. This visualization provides the individual with a good measure of how much information they are sharing compared to the larger group (comparison). Additionally, this visualization is useful for detecting changes in privacy, or, in a broader context, "global" changes (persistence and validation). As can be seen in Figure 4, if the visualization shows abrupt changes to many users data simultaneously, it may indicate a policy change, widespread error, or other noteworthy event.

2.3 Alternative Tools

In order to determine if there are tools that are preferred over the currently available privacy configuration tools, we selected a list of three alternatives. These tools attempt to simplify user configuration of privacy by minimizing the amount of user input required to achieve the desired settings.

The first mechanism is a 3-option (easy/medium/hard) system, where users can preset a system wide default level of privacy from a list with only three options. The user could, for example, set the default level of privacy to "completely private," "friends only," or "completely public". This mechanism enables users to quickly define a privacy level. The second system consists of a short survey where users are indirectly asked about their personal preferences, and the tool would later determine which are the best privacy settings for that user based on the responses. The third mechanism is a crowd sourced tool, where users can set their privacy level based off a particular individual or a group of their choosing.

For the three tools, in order to fully satisfy a user's preferences, the settings determined would serve as a baseline and users were given the option to later tweak the settings using the currently available mechanisms.

3. STUDY DESIGN

In this section we will describe the methods used to validate our approach.

3.1 Research Methods

Our study is designed to both qualitatively and quantitatively explore user understanding of privacy settings and preferences for privacy tools. As a quantitative approach, we created an online survey that consisted of 26 questions. Out of these, 16 were closed questions and respondents had to choose and answer from a given list of options. These questions consisted of 3-point and 5-point semantic scale questions as well as multiple choice questions. We chose to use semantic scale questions because they allow for quantitative measurement of subjective ratings while also allowing for some flexibility in interpretation [16]. For example, one of the questions was: "Would you agree or disagree that you are concerned about online privacy?" and the answer options were: "strongly agree," "moderately agree," "neutral," "moderately disagree," and "strongly disagree." We used the 3-point scale for the majority of the semantic scale questions, as we did not need respondents to have the more granular choices as provided by the 5-point scale. It has been shown that 3-point scales do not significantly reduce reliability or validity [10]. Some of the multiple choice questions allowed respondents to enter a personalized response into a field labeled other. For example, one of the questions was: "If you would be willing to give up certain services or features in exchange for privacy, what services or features would you be willing to discontinue?" and the answer options were: "Photo Sharing",' "Seeing friends of friends," "Geotagging," "Search availability (other people can find you on the social network)," "Messaging," "Lists or Groups," and "Other." In total, the survey took 5-10 minutes to answer.

In addition to the survey, we also performed follow-up interviews with 10 survey respondents to gather additional quantitative as well as qualitative data. The follow-up interview consisted of 12 questions. For these questions, we used 3-point and 5-point semantic scales as well as yes-no questions in cases where binary responses were all that was needed. For all of these questions, we used the "'think aloud" technique, where the interviewee was told to speak freely in their response to allow us to gather deeper insights about user perspectives on privacy. Data visualizations were used to help describe the proposed privacy tools to the interviewees. The follow-up interviews were performed in person or over Skype and took from 25-45 minutes to complete depending on the extent to which the interviewee expanded on the interview question responses.

To increase reliability of the study [16], we took the following measures:

- Random order of answers: The answer options for the closed questions were randomly ordered. This ensures that the answer order does not influence the response.

- Validation questions: To ensure that respondents did not fill out the answers arbitrarily, we included two validation questions [3]. For example, one of the validation questions was: "What is the result of 5+2?" Respondents who did not answer these questions correctly were not included in the final set of valid responses.

3.2 Survey Respondents

We did not have any restrictions on who could fill out the survey. Because we wanted a diverse set of respondents, we distributed our survey through a variety of channels includ-

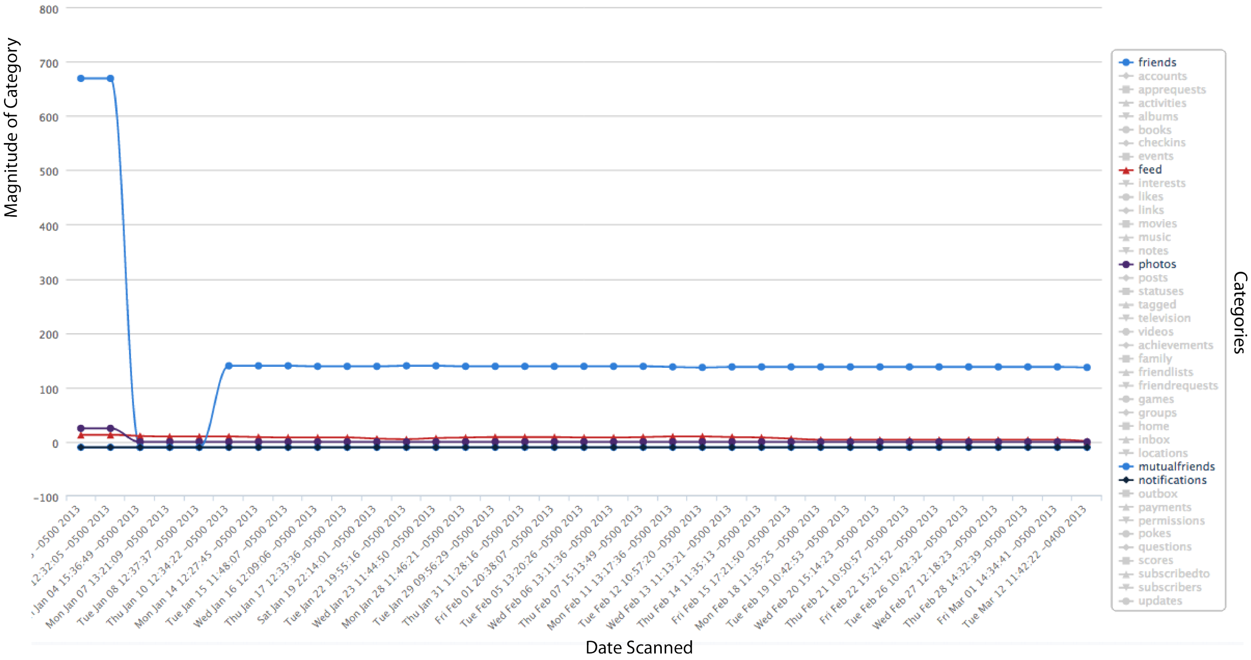

Figure 1: Individual visualization: Visualization of a single user's data from all categories presented over time.

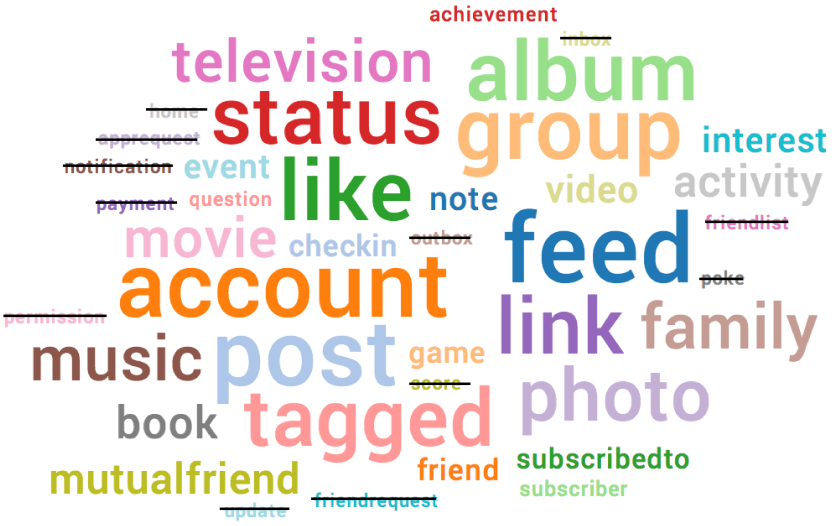

Figure 2: Individual visualization: Visualization of a single user's data from all categories as a snapshot in time

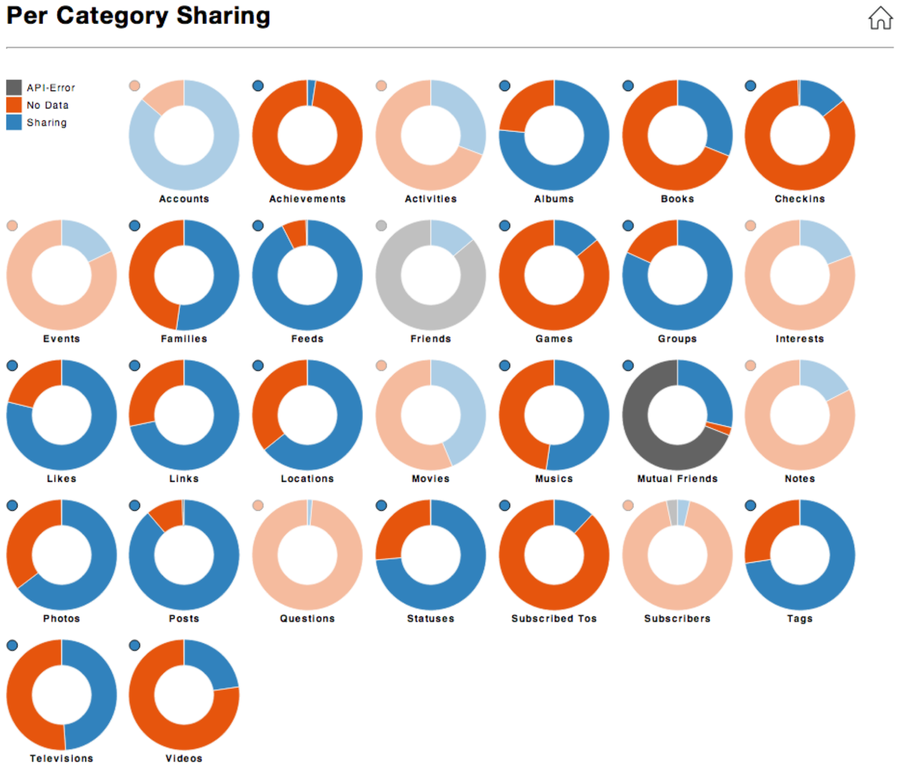

Figure 3: Contextualized visualization: A single user's data from as a snapshot in time (small dot top left corner) presented in the context of the same data collected from 512 users over 17 months (the larger donut)

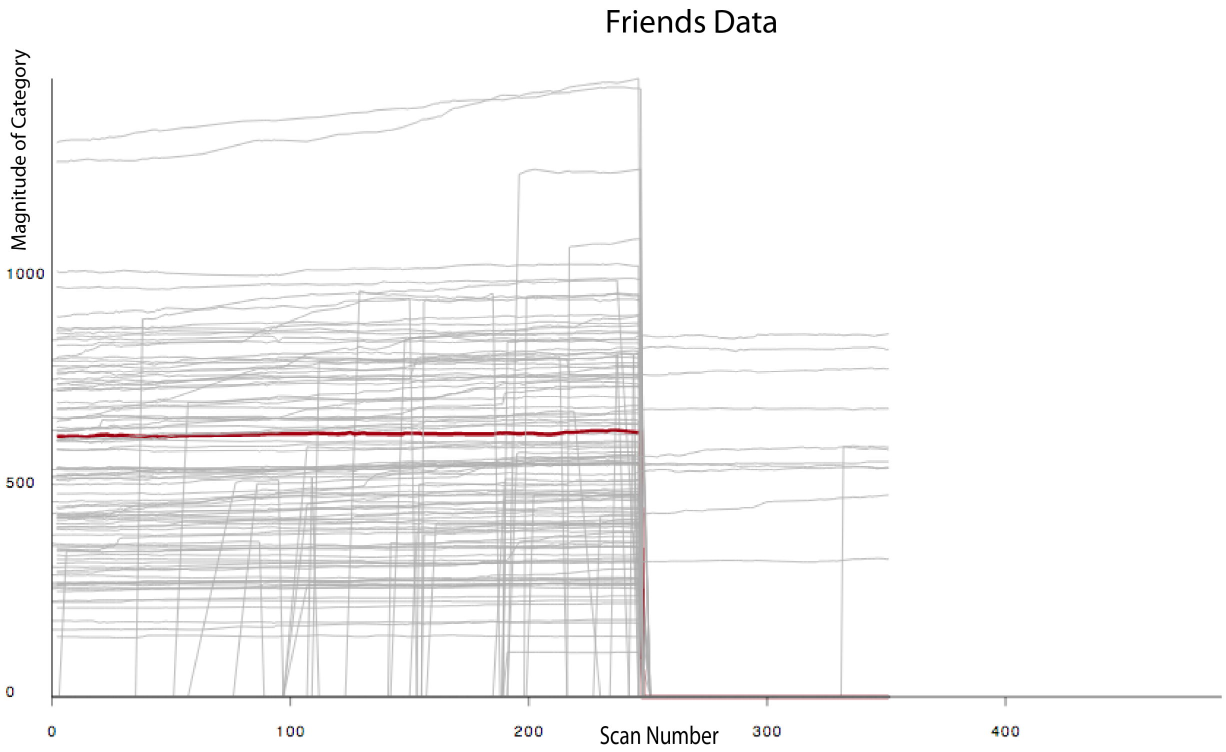

Figure 4: Contextualized visualization: a single user's data, for a single category presented over time, in the context of the same data collected from 512 users over a 17 month period

ing social networks like Facebook, Twitter, and LinkedIn, various mailing lists, and to personal and professional colleagues. We also enlisted the help of colleagues to further widen the distribution of the survey.

In total 63 respondents filled out our survey between 2 August 2013 and 6 October 2013. Filtering out the incomplete and invalid responses resulted in 59 valid responses (93.6% completion rate). The survey along with the raw data and summary information are available on our website.

3.3 Follow-up Interview Participants

After completing the survey, users were asked if they would be willing to participate in a follow-up interview. Of the 59 valid survey respondents 19 agreed to participate. From these 19, we selected 10 respondents based on schedule availability and geographic diversity. The only restriction for someone being able to participate in the follow-up interview, was that they had to be an active Facebook user. The follow-up interview included a privacy analysis that would scan the respondent's profile, and use this data to generate a visualization of their privacy setup. The visualizations that were generated are mentioned in section 2.2.

4. AWARENESS AND UNDERSTANDING OF PRIVACY

As mentioned earlier, one of the main goals of this study is to understand what is the current understanding of privacy settings and attitudes towards privacy for a social network user (RQ1). In this section, the findings of this study will be discussed.

4.1 Privacy in Social Networks

From a total of 59 users, 85% agreed that they are concerned about their privacy, while only 15% were neutral or not concerned about this issue. During the follow-up interview, it was shown that even though most of the privacy settings on Facebook are customizable with only a few clicks, most from the respondents feel that this platform is "constantly changing the settings, making it hard to customize privacy". This result corroborates what others have studied [7].

Most respondents have configured their privacy settings, with 88% clearing their cookies, 85% clearing their web browser cache, and 95% deleting their web browser history. Additionally, 73% of respondents are concerned that some social networks track their online activity through their web browsing history. These numbers show that most users are concerned about their online privacy on social networks, and thus modify what information is made available to the social network.

When users were asked to rate themselves on their understanding of how to customize their privacy settings using a scale of 1 to 5, where 1 is the least understanding and 5 is the most understanding, 80% of respondents rated themselves above a 3. Using the same scale, 50% of respondents rated themselves above a 3 on their understanding of what they are and are not sharing on social networks.

However, when users were asked to change the visibility of a specific setting on Facebook (Contact Settings: Website or Contact Settings: e-mail), only half of the respondents were able to perform the task in under 3 minutes. If the respondent could not perform the task in under 3 minutes, the task was considered timed out, and there was a binary classification of "task complete" and "task incomplete." It is interesting to note the disparity between the score respondents gave themselves, as opposed to the actual knowledge they had on how to configure their privacy settings.

4.2 Cost of Privacy

Users were asked to answer a few questions designed to help researchers understand how much users value their online privacy. These questions refer to different cost dimensions such as monetary cost, green (ecological) cost, and service cost.

Monetary cost refers to the money people are willing to spend to protect their online privacy. Even though users from different countries were asked to answer the survey, US dollars were used as a standard metric for the survey question. A total of 22% of the respondents said they would be willing to spend money in order to have privacy in their social networks, and 13% of all respondents said they would be willing to spend $1 - $10 per year on privacy. For example, sites like LinkedIn have a "premium fee" or extra charge that provide users with additional privacy features in exchange for a monetary cost. It is interesting to note that although respondents are concerned about their online privacy, few are willing to pay to protect it. These results are consistent with has been researched by others [2].

Performing any operation on a computer requires energy. According to Google, serving a single user for one month emits about 8 grams of carbon per day, which is similar to driving a car for a mile [9]. We wanted to determine if users would be willing to incur in an environmental cost associated with increased privacy. From the online survey responses, 19% of the users would be willing to give up certain services in exchange for a reduced environmental cost. The services they were more likely to give up were: friends of friends visibility (15%), geotagging (12%), and individual photo sharing (8%).

When users were asked if they would be willing to give up certain services in exchange for increased privacy, 59% said they would be willing to give up some services. Some of the most common services that users were willing to give up were: geotagging (49%), friends of friends visibility (42%), and search availability (32%).

5. IMPROVING USER UNDERSTANDING OF PRIVACY

In order to determine if user understanding of privacy can be improved through a crowd sourced approach (RQ2), users were asked to evaluate the visualizations from section 2. A semantic scale of 1 to 5 was used, where 1 is the least useful and 5 is the most useful. Additionally, users were asked if they would use a tool that included these visualizations.

Using the semantic scale, 70% of users rated the crowd sourced visualization tool above a 3. User appreciated the fact that they could "easily see any changes in their settings". Users were additionally asked if they would use such a tool. 80% of the respondents said they would use it at least once, arguing that the visualizations are a good way to verify if what they configured actually represents what they meant.

When asked how often would they like to be notified about their privacy settings, most respondents (90%) said they would prefer to be notified any time there is a change in their settings, 20% would like to have a monthly report, and 10% would like to have a daily report. Users were also asked how often they would like their information to be scanned (e.g., granularity for the graphs), and 50% of respondents preferred daily granularity, 30% a monthly granularity, and 10% preferred a weekly scan.

60% of respondents would prefer to have this tool built into the social network which was accessible at any time, as opposed to 40% of the respondents that would like to be notified through email, and only 10% would like to have the service presented as a Facebook application. There were no respondents who preferred a standalone application or a browser plug-in.

6. PRIVACY MANAGEMENT TOOLS

In order to identify if there are tools for configuring privacy that are preferred over the currently provided tools (RQ3), users were presented 3 different tools for customizing their privacy settings. The tools were present in both the survey and the follow-up interview.

From the survey, 56% of the respondents said they would like to have the 3 option system for configuring privacy. 54% of respondents said they would like to have the short survey mechanism. The crowdsourcing mechanism was the least popular, with only 22% of respondents wanting to have this mechanism available for configuring their their privacy. The two popular options are mechanisms where there is no additional information about what other users are sharing, or metrics about what is more or less common to share in a particular social network.

During the follow-up interview, users were asked to rank each of these mechanisms, as well as the current mechanisms on Facebook. For this purpose, a semantic scale of 1 to 5 where 1 is the least suitable and 5 is the most suitable was used. A summary of the results is presented in Figure 5.

Most of the respondents, 80%, ranked the current privacy mechanisms below a 3. Users justified the low score by arguing that Facebook's current privacy mechanisms are "difficult to understand and configure". For the 3 option system, most users 50% ranked it above a 3. This system had mixed reviews. According to the respondents, one of the major benefits is that they have a "known level of privacy" which they can then later tweak to meet their specific requirements.

The survey based approach was rated either above a 3 by 70% of the participants. According to the respondents, this system's main benefit is that it doesn't have predefined standards as the first mechanism (3 option), but instead it relies on inferring the ideal settings given a user's privacy requirements. Respondents argue that some networks have similar systems in place where a tour of the settings is provided, yet respondents feel they don't want to "spend useful time" answering these questions, or viewing the tours in order to get a finer tuned privacy setup. However, if the survey were short enough, they would like to define their default settings using this mechanism.

The final mechanism that was presented to the respondents consisted of using crowd sourced data to provide users with additional information when configuring their privacy. This was the most polarized system, and 60% of the respondents rated this tool above a 3. Users gave positive feedback to the fact that they were able to view their privacy setup compared to another group of users information thus enabling them to mimic the settings for that particular group. However, the respondents that did not like this system claimed that "they would not trust the privacy configuration skills from other users" in the network.

It is interesting to note that, during the survey, the crowd sourced system was the least popular option, but, once users had the opportunity to ask questions and look at a mock-up of the system, it had the greatest number of respondents ranking it a 5.

7. DISCUSSION

By using crowd sourced data, our approach introduces a new technique to end-users for understanding and configuring their privacy. As a first attempt, our research shows that, given the opportunity to ask questions and explore different configuration options in detail during follow-up interviews, users preferred the crowd sourced model. However, during the survey stage of research, before the subjects were able to ask questions, they rated the crowd sourced model lower than the other available options.

7.1 Implications of Results

Our research shows that part of the problem with current privacy controls is that users find them difficult to navigate. For example, Facebook obfuscates privacy settings by having category specific panels for certain settings instead of a single centralized set of controls. This makes it difficult for users to develop a comprehensive understanding of their individual privacy settings. When we asked users to change the website field from their Contact Settings from its current setting to some other setting (section 4.1), most users immediately navigated to the general privacy settings panel which does not provide access to this field. From the 10 interviewees only half were able to successfully change the setting in the time given. This suggests having a centralized location where all privacy settings could be configured would be beneficial to users.

Some other problems with current privacy controls concern persistence, validation, and comparison. Many of the interviewees disclaimed that although they understood how they had configured their privacy, their current settings may not reflect that configuration because of how often Facebook makes changes. This shows a lack of confidence in the persistence of user configured settings and, as shown by Mashima et al. [15], this doesn't translate to user action addressing these concerns. Furthermore, this highlights the inability of users to validate how they believe their settings are configured. Finally, current settings do not provide any context to users allowing them to compare their own settings to the settings of other users. By not providing context, there is an implicit suggestion that users already know what they do and do not want to share, and that there is some inherent "correct" way to sharing that is obvious to users. Instead, we argue that users can benefit by comparing to others and our results corroborate this. Our approach addresses these problems as follows:

- **Validate:** By using crowd sourced data, our technique can provide users with an overview highlighting what information is and is not being shared.

- **Persistence:** By collecting snapshots of a user's settings over time, our approach can monitor and alert users to unanticipated changes.

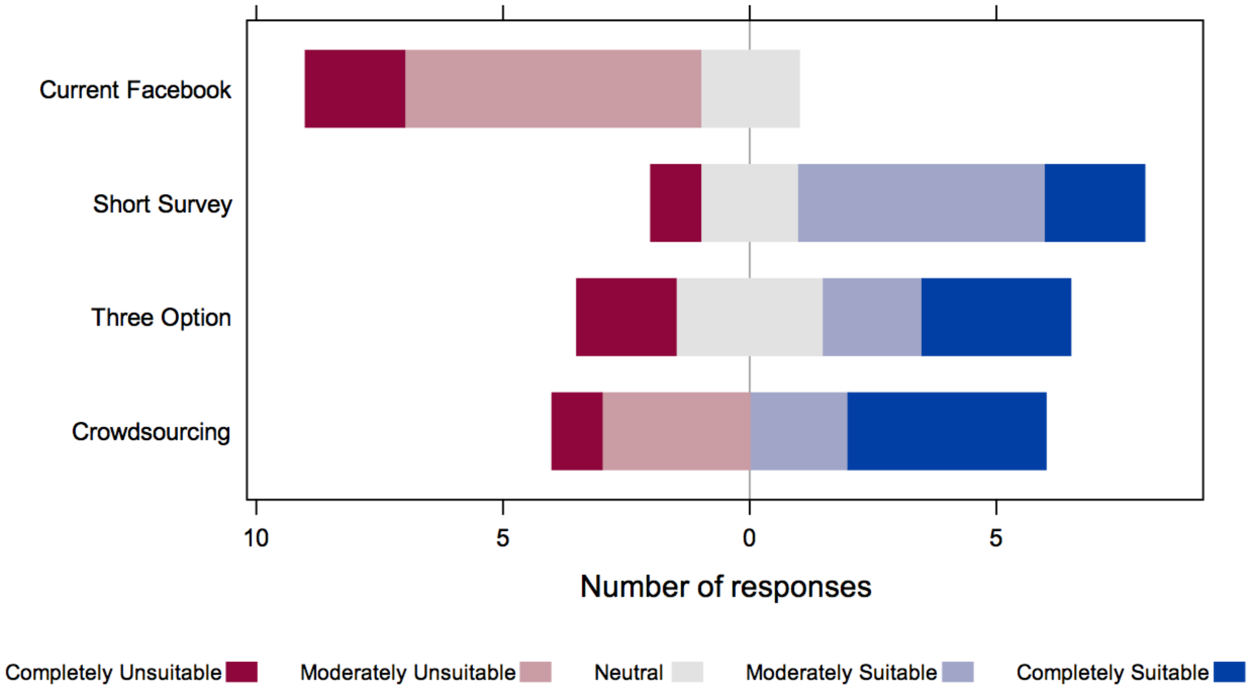

Figure 5: Distribution of of ratings for privacy management tools from follow-up interview

- **Comparison:** The contextualized view provided with crowd sourced data allows users to quickly identify irregularities, or areas where they may find they want to adjust their settings.

Our results suggests that many users do not know about all of the categories exist in their profiles. When asked to change the website field visibility from their contact settings most users said they did not know this was a part of their profile. We believe context is especially helpful when users are configuring privacy for for categories they are unfamiliar with.

As part of the follow-up interview process, we wanted to find out not only if users preferred crowd sourced data and found it helpful, but also how they would like to be presented with the crowd sourced tools. We provided users with the following options: "Stand alone application," "Browser plug-in," "e-mail service," "Facebook app," and "Other." Almost all users said they would prefer that the tool be a built in to Facebook. This suggests getting user adoption of crowd sourced tools without the participation of social network providers would be challenging.

7.2 Threats to Validity

There are several limitations to the validity of our study. First, there is a selection bias because the sample of individuals that answered our survey were self-selected. This implies that our results are only applicable to the volunteering population, which does not represent a complete sample of all the social network users. Also, the sample comprises individuals from different countries and cultures, and results may vary if the study is repeated within a specific group. However, the sample is broad enough to allow us to identify statistically significant trends and relationships.

Regarding internal validity, the fact that users filled out a survey implies that only certain concerns can be identified. Also, the questions and the format in which they are presented might constrain the limits of such concerns. It is possible that the study overlooked certain concerns and preferred methodologies due to the design of the survey. However, we attempted to mitigate this validity issue by including open-ended questions where respondents could further expand their answers, and by having "think aloud" questions during the follow-up interview.

An inherent property of an online survey is that respondents may not fully understand each of the questions and they might also select arbitrary responses. To address these concerns, we included validation questions and only report statistics about respondents that correctly answered these questions.

Despite these limitations, we managed to get a diverse and large set of respondents. This increases our confidence about the overall trends that reported in this paper.

8. RELATED WORK

Extensive research has been focused on improving user understanding of privacy and on simplifying privacy configuration. In the following paragraphs, we highlight important examples of this research.

Many studies have focused on the divide between what information users believe they are sharing and what is actually being shared and with whom. Stutzman et al. [18] show that as Facebook users increase the ammount of information they share privately they unknowingly disclosed information to so called "silent listeners." Madjeski et al. [14] find that of the 65 participants in their study, the majority either share (94%) or hide (85%) information unintentionally. This is further supported in the findings of Liu et al. [13] showing that just 37% of users' settings coincided with their expectations and "almost always expose content to more users than expected." Johnson et al. [11] conclude that user understanding of privacy is especially at risk when considering what information is shared with "members of the friend network who dynamically become inappropriate audiences based on the context of a post." Young and Quan-Haase [21] investigate factors that influence university students to disclose personal information on Facebook. They also study the different strategies students develop to protect themselves against privacy threats. These studies all highlight the challenges confronting users when trying to manage their privacy settings on social networks but stop short of making proposals for how these challenges can be addressed. Our approach compliments the results obtained from these studies and uses this information as a basis for proposing a crowd sourced system that improves user understanding.

There are other researchers investigating and developing different tools and techniques to manage privacy in online social networks. For example, an automatic technique proposed by Fang and LeFevre [5] configures a user's privacy settings in social networking sites by creating a machine learning model that requires limited user input. Tootoonchian et al. [19] propose a system called Lockr that improves the privacy of centralized and decentralized online content sharing systems. Squicciarini et al. [17] model the problem of collaborative enforcement of privacy policies on shared data by using game theory. By extending the notion of content ownership, they propose a solution that offers automated ways to share images. Another tool, developed by Toubiana et al. [20], was designed as a geo-location aided system that allows users to declare their photo tagging preferences at the time a picture is taken. This system enforces users tagging preferences without revealing their identity. Lipford et al. [12] investigate mechanisms for socially appropriate privacy management in online social networks. They study the role of interface usability in the configuration of privacy settings and develop a first prototype where profile information is presented in an audience-oriented view. A tool called PrivAware, developed by Becker and Chen [4], detects and reports unintended information loss in online social networks. Additionally, this tool recommends action to take to mitigate privacy risk. Although these studies propose alternative tools and mechanisms for configuring privacy, to the best of our knowledge, there have been no tools that present a contextualized view of privacy during user configuration.

9. CONCLUSION

In this paper, we conducted a study presenting a crowd sourced approach for simplifying the configuration and understanding of social network privacy settings. The study was divided into two sections: a survey for which we collected 59 valid responses and a follow-up interview for which 10 survey respondents participated. While only a small portion (22%) of the survey responses indicated they would prefer a crowd sourced tool for configuring privacy, during follow-up interviews 60% of participants said they would prefer such a tool over the current settings after having the opportunity to explore our approach more thoroughly.

Currently there are several obstacles that complicate user understanding and configuration of privacy. These are obfuscatory privacy control mechanisms instituted by social networks, frequent changes to privacy policy, and default settings intended to provide access to a users data. Our approach addresses these problems by simplifying comprehension of privacy through the use of data visualizations which provide an overview of the settings for each category in a user's profile.

We also provided users with data visualizations showing an individuals privacy setup and history in a contextualized view. For the individual visualizations, these are helpful as a quick summary of which categories are being shared and a quick way of determining in what ways their privacy setup has changed over time. These would allow a user to identify expected and unexpected changes in privacy by reviewing how their settings change. While some participants in the follow-up interviews did not prefer our crowd sourced approach, 70% said they found the data visualizations useful and 80% said they would use them at least once.

Finally, our approach could be used to influence the development of future privacy control mechanisms. Although our mock-up and visualizations only serve as a template for what information such a system may include, our results show that a crowd sourced approach does improve user understanding of privacy.

10. ACKNOWLEDGMENTS

We would like to thank all the participants that made this research possible by agreeing to participate in our survey and follow-up interviews. We would also like to thank Priyank Singhal and Alex Sigaras for helping with the collection of surveys. The authors are members of the Programming Systems Laboratory funded in part by NSF CCF-1302269, CCF-1161079, NSF CNS-0905246, and NIH U54 CA12185.

11. REFERENCES

[1] Oxford English Dictionary Online. `http://www.oed.com/`, September 2013.

[2] A. Acquisti, L. John, and G. Loewenstein. What is privacy worth. In *Workshop on Information Systems and Economics (WISE)*, 2009.

[3] T. Anderson and H. Kanuka. E-research: Methods, strategies, and issues. 2003.

[4] J. L. Becker and H. Chen. *Measuring privacy risk in online social networks*. PhD thesis, University of California, Davis, 2009.

[5] L. Fang and K. LeFevre. Privacy wizards for social networking sites. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 351–360, New York, NY, USA, 2010. ACM.

[6] D. Fletcher. How facebook is redefining privacy, 2010.

[7] G. Gates. Facebook privacy: A bewildering tangle of options. *The New York Times*, Year.

[8] V. Goel. Facebook to update privacy policy, but adjusting settings is no easier. *The New York Times*, August 2013. Accessed October 01, 2013.

[9] Google Inc. A better web. better for the environment. `http://www.google.com/green/bigpicture/references.html`, 2012. Accessed September 30, 2013.

[10] J. Jacoby and M. S. Matell. Three-point likert scales are good enough. *Journal of Marketing Research*, 8(4):pp. 495–500, 1971.

[11] M. Johnson, S. Egelman, and S. M. Bellovin. Facebook and privacy: it's complicated. In *Proceedings of the Eighth Symposium on Usable Privacy and*

Security, SOUPS '12, pages 9:1–9:15, New York, NY, USA, 2012. ACM.

[12] H. R. Lipford, A. Besmer, and J. Watson. Understanding privacy settings in facebook with an audience view. *UPSEC*, 8:1–8, 2008.

[13] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, IMC '11, pages 61–70, New York, NY, USA, 2011. ACM.

[14] M. Madejski, M. Johnson, and S. Bellovin. A study of privacy settings errors in an online social network. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, pages 340–345, 2012.

[15] D. Mashima, E. Shi, R. Chow, P. Sarkar, C. Li, and D. Song. Privacy settings from contextual attributes: A case study using google buzz. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, pages 257–262, 2011.

[16] R. Rosnow. *Beginning behavioral research : a conceptual primer*. Pearson/Prentice Hall, Upper Saddle River, N.J, 2008.

[17] A. C. Squicciarini, M. Shehab, and F. Paci. Collective privacy management in social networks. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 521–530, New York, NY, USA, 2009. ACM.

[18] F. Stutzman, R. Gross, and A. Acquisti. Silent listeners: The evolution of privacy and disclosure on facebook. *Journal of Privacy and Confidentiality*, 4(2):2, 2013.

[19] A. Tootoonchian, S. Saroiu, Y. Ganjali, and A. Wolman. Lockr: better privacy for social networks. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, CoNEXT '09, pages 169–180, New York, NY, USA, 2009. ACM.

[20] V. Toubiana, V. Verdot, B. Christophe, and M. Boussard. Photo-tape: user privacy preferences in photo tagging. In *Proceedings of the 21st international conference companion on World Wide Web*, WWW '12 Companion, pages 617–618, New York, NY, USA, 2012. ACM.

[21] A. L. Young and A. Quan-Haase. Information revelation and internet privacy concerns on social network sites: a case study of facebook. In *Proceedings of the fourth international conference on Communities and technologies*, C&#38;T '09, pages 265–274, New York, NY, USA, 2009. ACM.