A Study of Privacy Settings Errors in an Online Social Network

Michelle Madejski*
michelle.madejski@gmail.com

Maritza Johnson, Steven M. Bellovin
Columbia University
{*maritzaj,smb*}*@cs.columbia.edu*

***Abstract*—Access control policies are notoriously difficult to configure correctly, even people who are professionally trained system administrators experience difficulty with the task. With the increasing popularity of online social networks (OSN) users of all levels are sharing an unprecedented amount of personal information on the Internet. Most OSNs give users the ability to specify what they share with whom, but the difficulty of the task raises the question of whether users' privacy settings match their sharing intentions. We present the results of a study that measures sharing intentions to identify potential violations in users' real Facebook privacy settings. Our results indicate a serious mismatch between intentions and reality: every one of the 65 participants in our study had at least one confirmed sharing violation. In other words, OSN users' are unable to correctly manage their privacy settings. Furthermore, a majority of users cannot or will not fix such errors.**

***Keywords*-privacy; social networking**

I. Introduction

It is increasingly common for end-users to contribute content to the Internet. The ease of use and availability of social networking websites, photo-sharing websites, and blogging software enable people with minimal technical skills to share information quickly and easily. This trend leads to many questions related to the privacy of online data and the usability of existing access control mechanisms. Are Internet users concerned about online privacy? If so, does their behavior reflect their concerns?

For many users of online social networking websites, there are two ways for them to protect their data. The first, of course, is to refrain from making the item available online. This is not a viable option, given that the purpose of online social networks (OSN) is to share information and communicate with others. The second option is to use the privacy controls provided to manage who can see which items. While the second option appears viable, both formal studies and anecdotal evidence suggest that configuring privacy policies and managing access control policies is a difficult task for most users. Do OSN users manage their privacy settings correctly? More precisely, do their privacy settings match their intentions?

We conducted an empirical evaluation of the actual preferences and behavior of Facebook users. We wished to measure whether OSN users' actual privacy settings match their sharing intentions. We chose Facebook because of its overwhelming popularity. The company itself claims over 800 million active users and more than 900 million shared objects that users interact with [1]. In this paper we describe an empirical study with three parts: a survey to measure privacy attitudes, a questionnaire to gather sharing intentions, and a results phase where participants indicate whether potential violations represent an inconsistency between their sharing intentions and privacy settings. Privacy attitudes have previously been measured in various settings (see Section II), and some laboratory studies have been conducted on the usability of privacy settings. However, our empirical study is the first to identify violations by comparing sharing intentions against users' actual privacy settings in a real OSN.

Our results show that overwhelmingly, privacy settings do not match sharing intentions. That is, OSN users are sharing and hiding information incorrectly as judged by their beliefs. Furthermore, a majority of participants indicated that they could not or would not fix the problems. The prevalence of such errors — every participant had at least one incorrect setting — suggests the current approach to privacy controls is deeply and fundamentally flawed and cannot be fixed. A completely different approach is needed.

II. Related Work

Our work draws upon many themes including research on OSN users' privacy concerns and their use of privacy features, research on how users interact via OSNs, and research on the usability of access control mechanisms.

One of the earlier investigations of Facebook users' privacy settings was conducted in 2006. Acquisti and Gross surveyed 209 Facebook users on their knowledge of the visibility of their profile, crawled the website to collect the profile data of the university's network members, then compared the survey answers against the available profile data. Some participants (8%) were sharing more that they thought they were and some (11%) were sharing less than they thought, but in general most (77%) participants were aware of what they were sharing [2]. This study is similar in nature to our study, except it only measured users' awareness of the publicness of their profile, it did not measure users' sharing intentions.

The study was a follow-up to an earlier study that passively measured information disclosure on Facebook [3]. In

*Work done while at Columbia University

2005, Gross and Acquisti analyzed 4,540 Facebook profiles to measure the information that was available and found that the majority of users shared a large amount of personal information, yet fewer users chose to limit access to their profile to just friends (0.06%).

Facebook has made major changes to the website since 2006; it is now open to anyone, not just to students as it was in the past, and many new features and privacy options have been introduced such as third party applications, the news-feed, photos, videos, status updates, notes, and the ability to tag other users in posts. The proportion of users who utilize the available privacy settings is also much different since at least 2008. Krishnamurthy and Wills measured the number of public profiles in 20 regional networks and found 53-84% of profiles were public [4]. This number is quite a bit smaller than the 99.9% that were public in 2006.

A subsequent study used a methodology similar to ours and reports results that corroborate our findings. In May 2011, Liu *et al.* asked Facebook users to report their ideal audience for ten photos and measured the correctness of their privacy settings based on the actual settings for the photos [5]. More than half (63%) of the photos had privacy settings that were inconsistent with users' desired settigns. Rather than limit the evaluation to photos, our methodology considered all textual content associated with the participant's profile.

An investigation of privacy settings is incomplete without understanding how users want to share and their goals for using an OSN. Along these lines, prior research has found that many OSN users primarily interact with people they know offline. In a study of motivations for using Facebook, Joinson found that most users utilize Facebook for "keeping in touch" with people with whom they have an offline relationship with, this includes looking up information about friends and communicating with friends [6]. Lampe *et al.* also researched how users interact with Facebook and reported similar results [7]. Joinson also found that users' privacy settings varied based on their reason for using Facebook. This point is critical to our evaluation – OSNs serve a purpose for users which is usually to facilitate communication with other users.

We liken the management of OSN privacy settings to managing an access control policy and note that the correct management of access control policies is critical to security; yet, even systems administrators and experienced technical users have trouble correctly configuring access control settings [8]. User studies have found that users have a difficult time completing basic access control management tasks, including determining who has access to which resources, and making changes to an existing policy [9].

III. Methodology

In our study we investigated whether users' privacy settings match their sharing intentions. We implemented the study as a Facebook application which allowed us to conduct the study remotely. Each participant completed the study in two sessions. Prior to installing the study application, the participant read a consent form that explained the study and they reviewed the requested privileges in the application installation dialog.[1] We collected data in late 2010.

Stage 1: Survey The study began with a survey to measure the user's privacy priorities, confidence in existing settings, Facebook usage, history of privacy violations, and exposure to privacy-related media coverage. We present the questions alongside the results in Section IV-A.

Stage 2: Collection of Intentions We asked participants to report their sharing intentions using a table where the columns displayed profile groups and the rows displayed information categories. In each cell, the participant indicated their attitude toward sharing the information category with the group. The choices were show, apathetic, and hide.

For the profile groups, our study focused on the default groups that are currently used in Facebook privacy settings: friends, friends of friends, network members, or everyone. Privacy settings can also be configured using custom friend lists though we chose not to measure this.

We collected sharing intentions based on information categories instead of data types (e.g., photos, notes, links, events, and status updates) which is how Facebook privacy settings are currently organized. Users can also configure settings on a per post basis, which we did not study. The information categories were based on textual content, rather than data type, and spanned all data types. We collected sharing intentions to assist in the identification of potential violations. For this reason, we chose categories that users would likely have a strong opinion about (the information categories are listed in Figure 4).

Stage 3: Identification of Potential Violations The application identified potential sharing violations by comparing the participant's sharing intentions with their privacy settings. First, the application compiled a list of the information categories where the participant indicated a show or hide intention (apathetic intentions were ignored since they cannot produce a violation). Then the application classified the participant's profile data using our information categories. Next, the application iterated over the classified items and checked the privacy settings for the four profiles groups. The application recorded the identifier and type of violation when there was an inconsistency between the participant's intention and privacy settings. Stage 3 produced two lists: a list of the posts where the participant intended the category to be shown but the post was hidden, and another that included the posts where the participant intended the category to be hidden but the post was visible.

In order to classify the participant's posts using our categories, the application inspected all textual data associated

[1] Columbia University Protocol IRB-AAAF1543

with the participant's profile and activity. To execute this, the participant needed to grant the application permission to access their profile data including all posts that the participant had shared on their own profile, the posts the participant had made on their friends posts, and the posts the participant's friends contributed to their profile. The application classified the posts using sets of keywords. We created the sets of keywords manually, prior to recruiting, by collecting unique words that were common to each category. We did this by consulting sources such as existing Facebook data, terminology lists, and tags on related online content.

In order to check the privacy settings for each post, we created four profiles to represent the default profile groups. We created the profiles such that they were mutually exclusive. The *friend* profile had a single friend which was the profile used to check the privacy settings for *friend of a friend*, we sent a friend request from the *friend* profile to the participant before the study began. *Stranger* did not have any friends and was not a member of any networks. The *network member* was a member of the Columbia University network and did not have any friends. Only *network member* was a member of the Columbia University network.

We define a *hide violation* to be the case where the participant's intent was to hide the information category from the profile group, but one or more objects in the category was accessible. We define a *show violation* to be the case where the participant's intent was to show the information category to profile group, but one or more objects in the category was not accessible.

Stage 4: Confirmation of Violations In the final stage, we asked the participant to review the potential violations and confirm which were actual violations. In this stage, the participant proceeded through twelve screens: one screen per information category that was divided into four sections, one section per profile group. In the case the application had identified a potential violation for the profile group and information category, the application presented the potential violation to the participant and asked the participant whether it was an actual violation of their sharing intentions.

Our algorithm for identifying potential violations was designed to liberally assign categories to increase the chance of identifying actual violations. For potential violations, the application retrieved the object in question and displayed it to the participant. The justification (i.e. matching keywords) for the potential violation was shown in boldface to provide the participant with context. Within each section the potential violations were grouped based on the source (whether the data was posted by the participant or a friend) and on the data type (photo comment, group, event, status update, etc.). We asked the participant to confirm the potential violations. This is a key step that is novel in our study design, previous studies have only guessed at potential violations; it is not possible to distinguish an actual violation from a potential violation without knowing the user's sharing intentions.

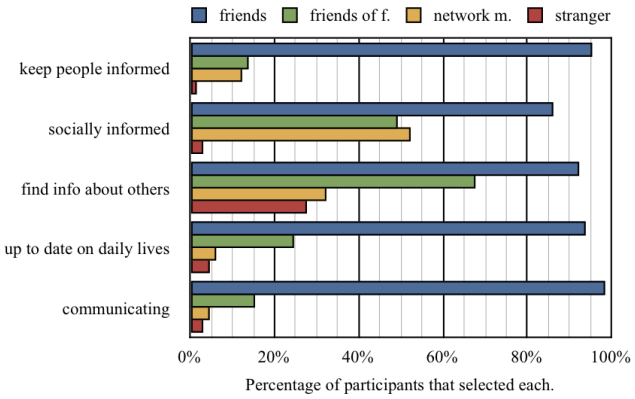

Figure 1. Participant responses to, "Why do you use Facebook?"

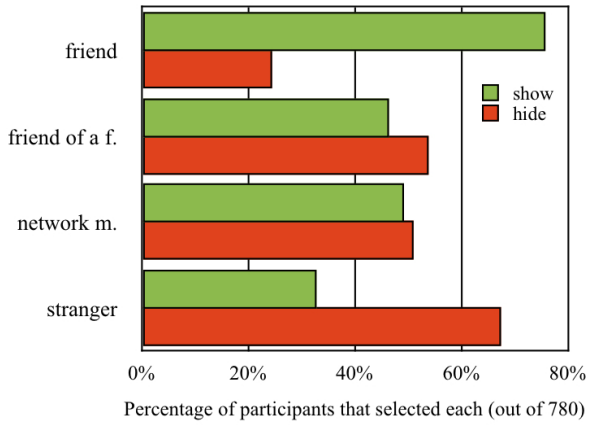

Figure 2. The participants' sharing intentions for each profile group. Each participant reported a total of 48 sharing intentions.

Finally, we asked the participant whether they would attempt to correct the violation for each confirmed violation.

A. Participants

Recruitment methods were targeted at the Columbia University community and included flyers, broadcasts to Facebook groups, broadcasts on mailing lists, and a paid advertisement on a campus blog. The final sample was a convenience sample of students who responded to the advertisements. A total of 65 people completed the study (38% male). The average age was 21.3 years (*S.D.* = 1.90). We compensated the participant $10 for their time.

IV. Results

In this section we present the results from each stage of the user study. The data confirm that users are concerned with OSN privacy, and show that even though their privacy settings are not aligned with their sharing intentions they do not intend to take action to correct their privacy settings.

A. Survey of Privacy Attitudes

Here we present the survey questions alongside the results (the full survey is reported in a technical report [10]).

First we asked, "What is the most important reason for online privacy?" Half (49%) the participants selected reputation security – to hide information to protect social reputation The next most poplar answer was economic security (39%) – to prevent identity theft and protect browsing habits from advertisers. The least important reason (12%) was physical security – to ensure physical safety, by hiding your face, location, and/or contact information from strangers.

We asked how often they untag photos and described a few scenarios when a user might untag a photo. Most participants (94%) had untagged a photo because "I didn't like the photo of me (it was unattractive or unflattering)" and most (94%) had untagged a photo because "the photo displayed behavior I did not want to be associated with (something that could be embarrassing if others saw it)."

We asked whether they engage in five activities with the four default groups (presented as a table of 20 checkboxes): "keep people informed about my life," "finding information about people ," "finding information on people's daily lives (e.g. newsfeed)," "personal communication (e.g. messages, walls)," and "being socially informed (e.g. events, groups)." Participants reported to interact with 'friends' the most and 'strangers' the least (see Figure 1).

We asked, "Do you feel your Facebook settings reflect your attitude related to privacy?" Nearly every participant (95%) responded affirmatively ($CI_{.05}$ = 5.3). We asked, "Have you ever had an accidental leak of information on Facebook that had a negative impact?" Most participants (91%) responded that they had "never had an accidental leak of information on Facebook."

We asked, "Have you heard anything regarding Facebook and privacy lately in the news lately?" Most participants (85%) had heard something from a general news source. We also asked participants, "Has the media coverage affected your behavior on Facebook?" Some (29%) replied the media had not affected their behavior at all. Those who answered yes (n = 46) could select more than one of the options listed: nearly all of them (83% of the 46) "became more selective about the information I post on Facebook," some (22%) deleted a Facebook friend, and most (91%) claimed to have modified their privacy settings to be more private.

B. Sharing Intentions

We asked the participant to state their sharing intentions across twelve data categories for four groups, then, for analysis, we combined show and apathetic intentions (Figure 2). Participants were willing to share most categories with a 'friend' (76%). Less than one-third of the categories were selected to be shared with a 'stranger' (33%). A few categories drew a large number of hide intentions for all groups like sexual, negative, drug, and alcohol.

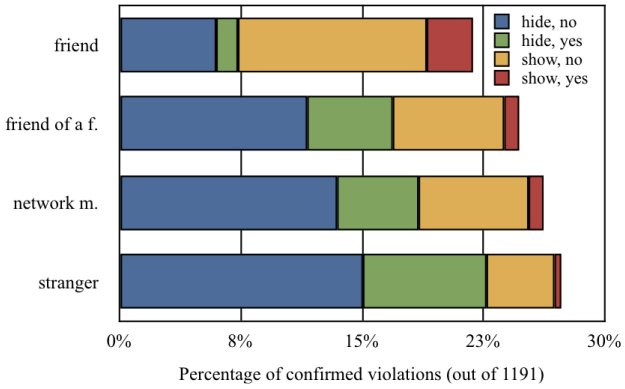

Figure 3. The percentage of confirmed violations presented by group. Each bar is divided into hide and show violations, then further divided to show the proportion of violations that elicited action.

Female participants selected more categories to share with friends and less to share with strangers. We computed a contingency table chi-square test on the frequency of show intentions for male and female participants. The difference in the number of sharing intentions between male and female participants is significant ($\chi^2(7) = 51.2$, $p < .0005$).

C. Confirmed Violations

Every single participant confirmed at least one sharing violation: 94% of participants confirmed a hide violation – they were sharing something they intended to be hidden ($CI_{.05}$ = 5.77), and 85% of participants had at least one show violation – they were hiding something they intended to be shared. We recorded a total of 1191 confirmed violations across the sample (M = 18 per participant, $S.D.$ = 10.5). More than half of the violations we recorded were hide violations (778 total, M = 12 per participant, $S.D.$ = 9.0). Show violations represented 35% of the confirmed violations (413 total, M = 6 per participant, $S.D.$ = 5.7).

For each confirmed violation, we asked the participant whether they would take action based on the violation, then estimated the perceived severity of the violation using their response. Even though every participant confirmed at least one sharing violation, only 58% of participants reported they would take action in response to at least one. Nearly all participants (97%) had at least one confirmed violation that they did not plan to address.

In Figure 3, we present the confirmed hide and show violations per profile group, each bar is further divided based on the reaction to the violation. Overall, the distribution of violations across the four profile groups is nearly balanced, however, the composition of the violations differs by group. For example, 'friend' had the most show violations and 'stranger' had the most hide violations. Hide violations were more likely to motivate action (30% of 778 hide violations), especially for the non-friend groups (stranger

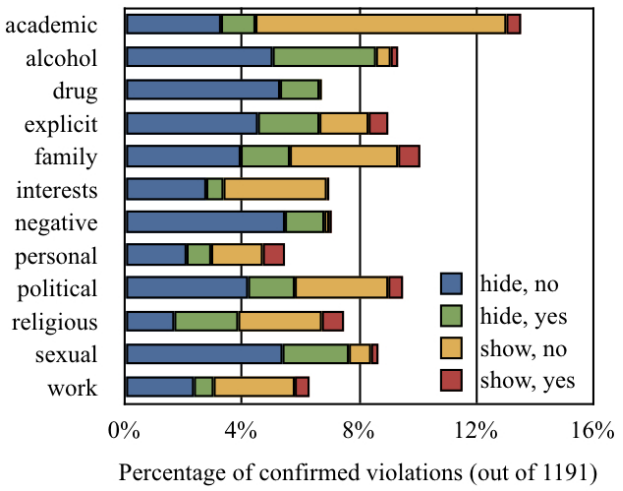

Figure 4. The percentage of confirmed violations divided by information category. Each bar is divided into hide and show violations, then further divided to show the proportion of violations that elicited action.

= 12% of 778 hide violations, network member = 8%, and friend of friend = 8%). In general, the participants are not motivated to correct show violations (85% of 413 show violations), though show violations that involve the friend group are slightly more likely to motivate action (8% of 413 show violations). While some violations motivated changes, the most frequent response was 'no action' (76% of 1191 confirmed violations).

In Figure 4, we present the confirmed violations by information category. The high number of violations for academic (14% of 1191) may have been an artifact of our sample of students. Similarly, the high number of hide violations for alcohol (9% of 1191) may have been due to the fact that many participants were under the legal drinking age. Hide violations for alcohol, sexual, explicit, and religious were most likely to motivate action, and show violations for family, personal, and religious were most likely to motivate action.

V. Discussion

We measured the accuracy of users' Facebook privacy settings by comparing their sharing intentions with their actual privacy settings. We found that every person in our sample had at least one confirmed inconsistency between their sharing intentions and privacy settings. Even though the participant's reported that they would not correct many of the violations, the existence of these violations presents a clear message: not only are Facebook's existing privacy settings flawed but improvements must be made to minimize risk to users.

We suspect that the basic access control mechanism used by Facebook is irreparably flawed. Previous studies on the usability of access control mechanisms (e.g., [9], [11]) have shown that this style — a list of items, and a set of permissions for various users which must be set manually by the owner — is difficult to use. A drawback of past studies is that they use contrived scenarios, and synthetic data which users may not feel motivated to protect. A benefit of studying Facebook is that the data is personal, and users are, presumably, motivated to protect it. Our results, however, show that even with personal data our participants were not able to protect it successfully. An unfortunate result given that our survey data indicate they are concerned with privacy, take steps like untagging or deleting content to protect their privacy, and believe their privacy settings are correct. Furthermore, the results of a related study suggest that users do not understand the limitations of the current Facebook mechanism [11]. We believe it is reasonable to conclude the problem is inherent in the basic design and further research is needed.

Even though every participant had at least one confirmed violation, about the same number of participants supplement the existing privacy settings by untagging and deleting content. Such privacy preserving behaviors have been observed in other research as well, such as a survey by Pew Internet [12]. For example, the data from the Pew survey show that in the 18-29 age group many OSN users had deleted unwanted comments (47%).

One of the largest culprits for privacy flaws is Facebook's reliance on data types (e.g., photos, events, and status updates) for defining privacy settings. These data types are misrepresentative of the real world that Facebook attempts to model. Outside of a social network, an individual does not determine visibility of personal data by its format but instead by the context of its information. A key improvement would be to automatically categorize information with a predicted context, and define privacy settings per context that reflect the user's intent. Our data suggest that users are strongly opinionated about showing or hiding specific categories of information. Prior work has explored the possibility of using content-based access control for blog posts, further investigation is necessary to determine if a similar approach can be used for OSN posts [13].

The recommended privacy settings contradict how users interact with other Facebook users. The responses to our question about how users interact (Figure 1) and the overall sentiment expressed in the sharing intentions (Figure 2) suggest that users have little to no use for 'strangers' on OSN websites. Thus, the recommended settings should be updated to reflect users' needs.

Our study investigated users' sharing intentions and actual privacy settings in search of violations. The fact that every participant confirmed at least one sharing violation indicates that additional research on the usability of privacy settings is necessary. Determining the root cause of violations is one possible follow-up study; this is better suited to an in-person interview (as opposed to the remote study reported

here), which would allow study coordinators to adjust the questions to identify the source of the violation. Participants who have violations may not understand the privacy settings well enough to identify the reason behind a violation in any format other than interview.

A. Limitations of Study Design

Our sample contains only students and could be larger. Typically, a sample of students is a weakness but for our study it may be an advantage. Most of our participants are tech-savvy and experienced Facebook users. If any subset of users would be adept at managing privacy settings it might be students. Also, students will almost certainly be on the job market in the near future, which means the correct use of privacy settings is critical. The size of our sample is defensible given the extreme nature of our results, i.e. that every participant had at least one violation.

The statistics we present on confirmed sharing violations are a lower bound. We hypothesize that, in practice, each participant has more violations than were counted, which is an artifact of our identification algorithm and the study design. Across the 65 participants, the study instrument identified a total of 70,402 potential violations (M = 1083, $S.D.$ = 1056). Rather than present each violation to the participant individually, the application grouped potential violations by data category then by profile group and asked the participant if at least one of the data items in the group was a true violation. Furthermore, the algorithm only classified the textual posts, a future study might identify additional violations if photo content and videos were also considered. Thus, our final count is most likely an underestimate of the number of sharing violations.

We are unable to analyze the nature of the potential and confirmed violations beyond the analysis presented in the results section because our application did not retain, or even download, the posts in question. The application temporarily stored the Facebook identifier of each post that was potential violation. We implemented the application in this way to protect the privacy of our participants.

VI. Conclusion

We conducted a survey to evaluate the correctness of OSN users' privacy settings. Our results indicate that OSN users have trouble correctly specifying their privacy settings with the current mechanism of the most popular OSN. Every one of our 65 participants confirmed that our application correctly identified at least one sharing violation. In other words, every participant was sharing something they wished to hide or was hiding something they wished to share with a group of people on Facebook. Both cases represent a mismatch between the user's ideal policy and their actual policy, which suggests a shortcoming of the existing privacy settings. We recommend improvements to their mechanism based on our findings, and suggest directions for future work.

References

[1] Facebook, "http://www.facebook.com/press/info.php?statistics."

[2] A. Acquisti and R. Gross, "Imagined communities: Awareness, information sharing, and privacy on the Facebook," in *Proc. of Privacy Enhancing Technologies (PETS)*. Springer-Verlag, 2006.

[3] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in *Proc. of Workshop on Privacy in the Electronic Society (WPES)*. ACM, 2005.

[4] B. Krishnamurthy and C. E. Wills, "Characterizing privacy in online social networks," in *Proc. of First Workshop on Online Social Networks (WOSN)*. ACM, 2008.

[5] Y. Liu, K. Gummadi, B. Krishnamurthy, and A. Mislove, "Analyzing Facebook privacy settings: User expectations vs. reality," in *Proc. of Internet Measurement Conference (IMC)*. ACM, 2011.

[6] A. N. Joinson, "Looking at, looking up or keeping up with people?: motives and use of Facebook," in *Proc. of Conference on Human Factors in Computing Systems (CHI)*. ACM, 2008.

[7] C. Lampe, N. Ellison, and C. Steinfield, "A Face(book) in the crowd: social searching vs. social browsing," in *Proc. of the 20th conference on Computer Supported Cooperative Work (CSCW)*. ACM, 2006.

[8] T. Das, R. Bhagwan, and P. Naldurg, "Baaz: a system for detecting access control misconfigurations," in *Proc. of the 19th USENIX Conference on Security*. USENIX Association, 2010.

[9] R. W. Reeder and R. A. Maxion, "User interface dependability through goal-error prevention," in *Proc. of International Conference on Dependable Systems and Networks*. IEEE, 2005.

[10] M. Madejski, M. Johnson, and S. M. Bellovin, "The failure of online social network privacy settings," Dept. of Computer Science, Columbia University, Tech. Rep. CUCS-010-11, February 2011. [Online]. Available: http://mice.cs.columbia.edu/getTechreport.php?techreportID=1459

[11] S. Egelman, A. Oates, and S. Krishnamurthi, "Oops, I did it again: Mitigating repeated access control errors on Facebook," in *Proc. of Conference on Human Factors in Computing Systems (CHI)*. ACM, 2011.

[12] M. Madden and A. Smith, "Reputation management and social media," http://pewinternet.org/Reports/2010/Reputation-Management.aspx, May 2010.

[13] M. Hart, R. Johnson, and A. Stent, "More content-less control: Access control in the web 2.0," in *Proc. of First Workshop on Online Social Networks (WOSN)*. ACM, 2008.