Analysis of Information Disclosure on a Social Networking Site

Katherine Peterson[1] and Katie A. Siek[2]

[1] University of Colorado at Boulder, Department of Applied Mathematics, 526 UCB, Boulder, Colorado 80309-0526 USA
Katherine.Peterson@colorado.edu
[2] University of Colorado at Boulder, Department of Computer Science, 430 UCB, Boulder, Colorado, 80309-0430 USA
ksiek@cs.colorado.edu

Abstract. We present a small study about information disclosure and awareness of disclosure implications on Couchsurfing.com. Couchsurfing is an online social networking site where users connect with others interested in traveling and staying at each other's homes. Since users are looking for someone to stay or travel with, they must develop a rapport and trust before traveling. This leads users to share more information on their Couchsurfing profile than they ordinarily would share on mainstream social networking sites such as Facebook or MySpace. After a survey with twenty Couchsurfing users and semi-structured interviews with nine participants, we found participants were generally not concerned with the information they disclosed online and were not aware of how this information could be used against them by malicious third parties. We conclude the paper with a brief discussion of how designers and developers could utilize personas to better inform participants of the implications of their disclosure decisions.

Keywords: Social Networking, Information Disclosure, Privacy.

1 Introduction

In this paper, we look at information disclosure on Couchsurfing[1] to further understand how social networking participants determine what information they feel comfortable sharing. Couchsurfing is a social networking site where users meet other members to stay on their couches while traveling the world. Since the site is focused on travel and hosting travelers, many users disclose information about themselves such as dates they will be away on travel and detailed descriptions of their home locations that would not normally be found on mainstream social networking sites (e.g., MySpace or Facebook). Couchsurfing users are also asked for more detailed personal information than MySpace or Facebook users. For example, the profile template contains fields to describe a user's personal philosophy, knowledge to share, the most

[1] http://www.couchsurfing.com

A.A. Ozok and P. Zaphiris (Eds.): Online Communities, LNCS 5621, pp. 256–264, 2009.

amazing thing done/seen, and basic interests included in a MySpace or Facebook profile. In addition to creating a network of friends users can vouch for other users who they believe are trustworthy. To further promote a sense of security, users can verify their name and address by associating their name with a bank account or credit card and by sending a verification code in the regular mail.

We chose to analyze this social networking site because Couchsurfing users must develop some kind of rapport and trust before allowing an online friend to stay at their home. Thus, we initially assumed, these relationships developed online and are the perfect place to study what makes users comfortable with information disclosure because users must share some personal information to establish relationships.

In addition to studying information disclosure, we studied how aware social networking site users are of the malicious acts that can be committed with their personal data. While the risk of burglary based on travel dates and house location may be apparent to a user, there are many other malicious acts that can be committed based on basic information that users may not be aware of. For example, it has been found that an identity theft scheme can be created just by using a full name or email address [8].

The main contribution of this paper is a general awareness of self reported information on social networking site profiles that requires disclosure for user safety. Although users generally felt the information they provided in profiles would not lead to privacy concerns, we show that the information available could be problematic for most of the study participants. We, as a community of designers, must integrate an awareness of information disclosure into social networking sites so people who are unaware can learn and vary their participation and privacy appropriately.

2 Related Work

Social networking sites have been shown to play an important role in maintaining personal relationships. Researchers have shown that it is easier for individuals to keep relationships with people that they would not ordinarily keep in physical contact with because it is easier to distribute information electronically [3, 6]. The prominence of social networking sites makes it a necessity that users are aware of the implications of the information they share. It has been shown that people self disclose information on online communities for many reasons – reciprocation [13], online peer pressure, or naïveté of information disclosure [1]. We want to find out how users decide which pieces of information to share.

Online friendships take on a different meaning than in person friendships. Researchers have found that social network site users tend to add anyone as a friend that they know and do not have a strong negative feeling towards [2]. This means that a user might not know their online friends in person or trust them. All their online friends can view their profile unless specific privacy settings are modified. However, research has shown that users tend not to change default privacy settings [10]. On Couchsurfing the default privacy settings let anyone see a user's profile, even if the onlooker is not a member of the site.

Malicious acts can be executed using basic profile information that users likely do not consider jeopardizing. For example, an identity theft scheme can be created using only a full name or email address [8]. Basic profile information can also be used to create a digital dossier of the user. This is a cached record of all the digital data

available about a person over a given period of time which could be used by adversaries to track a person's life based on previously saved profile data [7]. In fact, a study of the recent census data showed that 63.3% of the population in the United States reported characteristics that likely made them unique based only on gender, 5-digit ZIP code and full date of birth (day, month, and year) [12]. This puts some users at a special risk for re-identification, which is the process of linking datasets without explicit identifiers such as name or address to datasets without explicit identifiers. Sweeney was able to link someone with a unique gender, ZIP code and birth date to their sensitive medical information that was received from a group responsible for purchasing health insurance for state employees and thought to be anonymized [14]. Different pieces of personal trivia in profiles could provide several different opportunities for re-identification, for example if a user lists a favorite book on their profile page and writes a review of the book on Amazon - they could be re-identified with their Amazon user account.

3 Study Design

The study consisted of two parts: a survey and an hour-long semi-structured interview. Here we discuss the design considerations for the study and information about the Couchsurfing members who participated in the survey and interviews.

3.1 Methods

The Survey: The Westin/Harris Privacy Segmentation Model was mentioned in several papers looking at privacy in social networking sites. Based on the answers to three questions, respondents are categorized into three groups: Privacy Fundamentalists, Privacy Unconcerned, and Privacy Pragmatists [9]. Although this survey would be a good indicator of users' privacy beliefs and practices, it has been concluded that it is not a good predictor of how people act in social settings since it was originally intended to analyze privacy beliefs in a business environment [5].

Our survey asked users about basic demographic information, behavior on social networking sites, what information they disclose, and knowledge of publicly available data. State government websites provided us with information about what personal records can be obtained by the public. Since this information varies a bit by state and country, we analyzed respondent's answers with respect to laws in their own area. For example, different states have different requirements for what authorization an individual must have to obtain a birth certificate or wedding license.

We received approval from the university's Human Research Committee before collecting data. Before launching the survey we had a test group of users complete it to verify the clarity of the questions and the system we used to collect data. An invitation to participate in our research project was posted on Couchsurfing forums. As an extra incentive to take the survey we put email addresses of respondents in a raffle for Amazon gift cards. Respondents were also asked if they wanted to participate in a semi-structured interview. Initially our invitation to participate was posted in a few general message boards on Couchsurfing. Users were very responsive, with 9 users responding within 24 hours of our post. Survey data was collected with a Google Form and analyzed by the team with basic statistical analysis.

The Interview: The first author conducted semi-structured interviews over the phone because of the broad geographic span between participants and the study team. Interview topics included understanding of malicious acts with user data, privacy, social networking site membership, and Internet information sharing philosophies. Our analysis was informed by the constant comparative method where we iteratively analyzed each transcript individually for thematic content.

3.2 Participants

We were able to recruit 20 Couchsurfing members to complete the survey and coordinated interviews with 9 participants. The survey participants were between the ages of 19 and 59 years old (average age = 35; s.d. = 11.8). Similarly, the interview participants were on average 34.1 years old (s.d. = 8.85). Sixteen participants lived in the United States, two lived in Canada, and one participant each lived in France and Belgium. Participants were distributed all over the United States – they primarily lived in the mid-west (7 participants) followed by the western part (4 participants) and then some on the East Coast and Southern parts of the country. Six participants had an advanced graduate degree, five participants had completed a four-year college degree, four participants had some college, and three participants had completed high school.

On average, participants had been part of the Couchsurfing community for 14.6 months (s.d = 12.3). Participants were part of 2 other social networking sites, on average, outside of Couchsurfing with Facebook (13 participants) and MySpace (7 participants) listed as the most popular alternative social networking sites. On average participants had 34.95 Couchsurfing friends (s.d. = 55.5; min = 0 and max = 231). Surprisingly, one participant who had zero friends had been on Couchsurfing for eight months, whereas the other participant had been on the site for less than a month. Fourteen of the eighteen participants who had Couchsurfing friends reported knowing all of their friends, three participants had never met one of their friends, and one participant had not met three of their friends in real life. Eleven participants had not changed the privacy setting on their Couchsurfing profile. Thirteen participants reported being very eager to meet new people on Couchsurfing and seven were somewhat interested in meeting new people.

4 Findings

Overall, we found that:

- Most (19 out of 20) participants could be identified through census data based on the information shared on their Couchsurfing profiles.
- Participants were mildly concerned about the information disclosed, but mostly thought that a third party would not take the time to target them.
- Most of the participants would disclose information via Couchsurfing with little information (e.g., a request to stay with them) about the other party.

4.1 Profile Information Disclosure

All of the participants listed their gender and zip code in their profile, as shown in Table 1. Based on work by Golle [12], we know 10 participants in this study are at

Table 1. Self-reported information disclosure - what participants disclose on their Couchsurfing profiles. Participants could select multiple pieces of information in the survey, thus percentages may add up to more than 100%.

Information in Profile	# Participants	Percentage
Phone Number	2	10
Dates of travel for upcoming trips	2	10
Detailed description of house location	2	10
Street Address	3	15
Description of daily hang-outs/habits	3	15
Pictures of home exterior	4	20
Email	7	35
Full Birthday	10	50
Full Name	15	75
Occupation	16	80
Personal Pictures	19	95
Age	19	95
Gender	20	100
Zip Code	20	100

risk of being identified through census data since they list gender, zip code, and full birth date on their profiles. We asked about participants' age because given a person's name and zip code, it is fairly easy to get a person's birth date and gender using a people search web site like Intellius.com. Nine participants could further be identified with Golle's method, with indirect people-searches to obtain gender and full birth date. Thus, all but one participant could be identified through census data by a motivated third party. Based on this self-reported data, six participants are at risk of the identity theft schemes reported in [8] based on putting their full name and email address in their profiles.

A person's house could be identified, depending on the area a participant lives in, with zip code and pictures of the home's exterior thanks to Google's Street View functionality [11]. Indeed, this would take significant time to traverse an entire zip code on Google Maps, however if a person had detailed descriptions of their local hang-outs (e.g., how far they are from their favorite Starbucks), the location of the person could be identified quicker. Likewise, phone numbers could be reverse looked-up to find out the location of the phone number. Fortunately, only two participants can be categorized into this risk.

4.2 Concerns about Disclosure

During the survey portion of the study, we found that none of the participants were "very concerned" with the amount of information they disclosed on Couchsurfing, whereas 13 participants were somewhat concerned and seven participants were not concerned with what they disclosed. Those who were somewhat concerned remarked in interviews that a decrease in privacy was expected when you use the Internet and the information they disclosed could be found in other resources as shown in the following quotes:

> *As much as I share seemingly personal information on my profile, you can find a lot of the same information by googling my name. Anyone who was interested enough could easily find newspaper articles or my friend's blogs saying much the same things. – P7*

> *Everything available about myself would also be found in the phone book. - P13 (listed - Full name, Personal pictures, Age, Gender, ZIP code, and Occupation in profile)*

One participant honestly disclosed that, "*I don't know what I'm doing exactly but I've nothing to hide, so no problem- P5*." This participant had disclosed his full name, age, gender, zip code, full birthday, occupation, and pictures of house exterior. As we discussed in Section 4.1, the information disclosed by P13 and P5 open themselves up to possible privacy schemes.

Those who were not concerned with the amount of information they disclosed either did not think anyone would take the time to do something with their information or acknowledge the risk but did not care as shown below:

> *I suppose that people could perhaps track me down at work and harass me, or else harass my friends. I don't know why anyone would want to do this, though. –P6*

> *Sure... identity theft and all that, but I don't worry about it. - P11*

Although participants who were not concerned with the amount of information disclosed, phishing and other malicious schemes creators do not necessarily care who is targeted - just that information can be used and exploited. Thus not worrying or not thinking anyone cares is not a realistic assumption if a person wants to protect her personal information.

4.3 Decision Process for Information Disclosure

Since Couchsurfing is a social networking site to help members find people to stay with during their travels, information disclosure is important so that both parties understand what type of person is staying with them. Indeed, four participants believed in disclosing as much information as possible – full disclosure – so others could decide if they wanted to stay with them:

> *The information on my profile is there to allow other surfers to get an idea of who I am, what it would be like to host/surf with/travel with me. References are all quite repetitive. I actually added information about the time I "scared off" a Couchsurfer, since he never left me a reference. I think people should know what they're getting into. – P7*

This idea even branched out into the global community for one participant:

> *I think that people should put a lot of information on the Internet. Being as open as possible about our lives to each other can only make the world a better place. – P6*

Table 2. Information participants thought was publicly available. Participants could select multiple pieces of information in the survey, thus percentages add up to more than 100%.

Publicly Available Information	# Participants	Percentage
Student records	4	22%
Stock purchases	5	28%
Voting registration information	7	39%
Birth certificate	7	39%
Military records	7	39%
Change of address form	8	44%
Property records	13	72%
Marriage license	14	78%
Arrest/court records	14	78%
Divorce record	15	83%

Two participants discussed how they gradually add more information to social networking sites as they become more active in the community. The final three participants described how they either limited views of their profiles to only friends – friends they knew in real life – or did not put much information on their profile.

Participants primarily took two approaches when considering the appropriate time to give the Couchsurfing requestor more information. The first approach, that four participants used, dealt looking at the requestor's profile, verifications, and references they had. Sometimes this verification went outside of the Couchsurfing site as P13 told us about verifying one Couchsurfer's travels:

> *For instance somebody said that they had been to a monastery in Asia and I checked their list of places they had been to see that they had really been there. - P13*

The second approach, used by another four participants, was simply to give the requestor information when a visit was confirmed. As we stated earlier, all of the participants said they knew most of their friends in real life. However, one participant discussed how when he first started Couchsurfing, he did not know anyone. Through the social networking site, similar to the relationships reported by boyd [3], he was slowly able to create relationships and a network of referrals:

> *Now I ask for their contact information only so I have a backup way to contact them if my train is late or I am delayed for some other reason. In the beginning my hosts and surfers were total strangers, but now they are very often friends-of-friends since I know so many people through the site. - P7*

Other approaches to deciding when to disclose more information included meeting the requestor face to face prior to the Couchsurfing visit, asking the requestor to disclose just as much information as was being disclosed by the host, and verifying that they knew someone in common.

4.4 Awareness of Information Disclosure Implications

For the final part of the survey, we asked participants to identify what pieces of information are publicly available in their country or state. In most cases in the United States, the information shown in Table 2 is publicly available to people if enough information is provided. For example, most universities provide the piece of information thought least publicly available to a third party. The information includes the student's full name, degrees awarded, schools, majors, and distinctions without alerting the student. Students do have the right to cease this information from becoming publicly available. This type of information can give a third party confirmation on where the person previously lived and possibly an income range given degree, major, and graduation year.

Birth certificates and wedding licenses sometimes require the requesting party to prove their relationship to the person(s) in question. The other records, although public, require more information than would most likely be available on a public social networking site. For example, voting registration requires registration date and military records require the years the person was in and addresses of time of entry, time of release, and post-separation. However what is interesting here is that participants do not know what information is public and how the data they currently disclose in their profiles (Table 1) can assist a malicious third party gain access to this public information.

4.5 Implications of Information Disclosure

All of the participants were aware of some adverse events that happened because of information disclosure on social networking sites. Six participants had experienced or heard of people being hacked because of information they posted on social networking sites. Two participants specifically mentioned the teen that committed suicide because a parent had pretended to be a potential love interest and then shunned her [4]. Two participants had adverse events specific to Couchsurfing where they received false information or the person who stayed with them had not disclosed enough information. For example, P11 shared:

> *When I was going to [place], the person gave me a wrong number and a false address. I ended up on a bad part of town at night.*

Interestingly enough, despite hearing about these adverse events or, in two cases, experiencing them, participants were still not that concerned about the information they disclosed. They generally believed, "this cannot happen to me."

5 What Next?

In this brief paper, we have given an overview of information disclosure on Couchsurfing and participants understanding about the implications of their information

disclosure. Although the amount of information needed is formidable to obtain some of the publicly available information for harmful schemes, participants lack of understanding or care about what is accessible and what they provide people shows that we need more assistance in teaching online social networking members how information they post can have implications elsewhere. Since we have studies that specify the information needed for specific schemes (e.g., name, zip code, and date of birth), developers could create filters that scan profiles and alert users when information they disclose puts them at risk of a malicious scheme. These messages, however, must be accurate, timely, and appropriate for the user group. For example, assuming a sixteen digit number is a credit card number, when in reality it is someone's geocaching data would frustrate the user and lead to mistrust of the filtering system. Instead, we suggest designers consider utilizing personas that resonate with the target population to provide information. The teen suicide, although tragic, resonated with some of the participants – but the message there was it is not okay to forge a profile. Thus, if we create personas that specifically discuss a person who experienced a malicious act because of the data the user has disclosed, they may be more interested in learning about the malicious act and deciding how to proceed or modify their practices.

References

1. Alessandro, A.: Privacy in electronic commerce and the economics of immediate gratification, pp. 21–29. ACM, New York (2004)
2. Boyd, D.: Reflections on Friendster, Trust and Intimacy, Seattle, WA (2003)
3. Boyd, D.M.: Friendster and publicly articulated social networking, pp. 1279–1282. ACM Press, New York (2004)
4. Collins, L.: Friend Game: Behind the online hoax that led to a girl's suicide, The New Yorker (2008)
5. Consolvo, S., et al.: Location disclosure to social relations: why, when, & what people want to share, pp. 81–90. ACM, New York (2005)
6. Donath, J., Boyd, D.: Public Displays of Connection. BT Technology Journal 22, 71–82 (2004)
7. Gross, R., Acquisti, A., Heinz, J.H.: Information revelation and privacy in online social networks, pp. 71–80. ACM Press, New York (2005)
8. Kay, H.C., et al.: The Internet Hunt Revisited: Personal Information Accessible via the Web (2004)
9. Kumaraguru, P., Cranor, L.F.: Privacy Indexes: A Survey of Westin's Studies, Carnegie Mellon University of Computer Science (2005)
10. Mackay, W.E.: Triggers and barriers to customizing software, pp. 153–160. ACM, New York (1991)
11. Musil, S.: Google wins Street View privacy suit. CNET News, Digital Media. CNET (2009)
12. Philippe, G.: Revisiting the uniqueness of simple demographics in the US population, pp. 77–80. ACM, New York (2006)
13. Preece, J.: Sociability and usability in online communities: determining and measuring success. In: Behaviour and Information Technology, pp. 347–356. Taylor and Francis Ltd., Abington (2001)
14. Sweeney, L.: Uniqueness of Simple Demographics in the U.S. Population, Carnegie Mellon University, Laboratory for International Data Privacy (2004)