Investigating OSN Users' Privacy Strategies With In-Situ Observation

Andreas Poller
Fraunhofer-Institute for
Secure Information Technology
64295 Darmstadt, Germany
poller@sit.fraunhofer.de

Andreas Kramm
Goethe University
60123 Frankfurt, Germany
akramm@rz.uni-frankfurt.de

Petra Ilyes
Goethe University
60123 Frankfurt, Germany
ilyes@em.uni-frankfurt.de

Laura Kocksch
Goethe University
60123 Frankfurt, Germany
kocksch@sit.fraunhofer.de

Abstract
Studies of interactional privacy in online social networks (OSN) showed that users' privacy strategies not only rely on technical privacy controls provided by the software but also on complex workarounds like self-censorship or information obfuscation. However, it is difficult to observe users' complex behavior and practices in these digital environments. For practical reasons, researchers have to rely mainly on self-reporting techniques with all their shortcomings. We propose combining qualitative interviews and a privacy-friendly tracking software to capture users' actions in OSN. We provide first results on how collected tracking data in combination with individualized interviews allow deeper insights into user practices, can prevent problems of self-reporting, and may eventually support software design.

CSCW'14 Companion, Feb 15-19 2014, Baltimore, MD, USA
ACM 978-1-4503-2541-7/14/02.
http://dx.doi.org/10.1145/2556420.2556508

Keywords
User studies, online social networks, Facebook, socio-technical system, socio-technical design

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

Introduction
There is an increasing relevance of online social networks (OSN) like Facebook in everyday life. This phenomenon has fueled many scientific studies with the aim to provide a sound understanding of how users manage their interactional privacy in these digital environments.

Two issues stand out in existing work: first, what users do: users' privacy strategies rely only to a limited extent on privacy controls provided by the OSN software. Other means such as self-censorship, obfuscating information before sharing, or using multiple OSN accounts supplement, or replace, the use

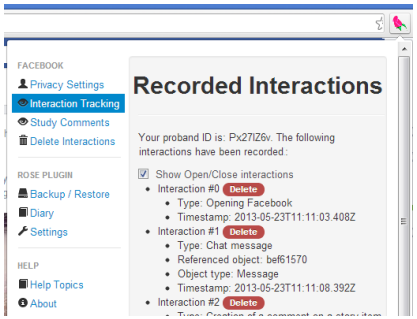

Figure 1: With the plugin control center, participants can evaluate collected data and can decide what they want to hand over to the researchers.

of OSN privacy controls. [1,3] Second, how we can study users' doings: our means to investigate privacy practices in such online environments are very limited because unlike in studies implying a physical site, it would require considerable efforts to visit users to observe them in situ. The reason is that OSN users interact with the software at varying times throughout the day, partially in private settings, and through user interfaces difficult to oversee by a researcher in an unobtrusive way.

As a result, state-of-the-art empirical studies primarily rely on self-reporting. However, self-reporting is only of limited reliability because (a) it is of retrospective nature, and users may not remember their privacy settings and actions correctly, and (b) users may hesitate to provide sensitive or even embarrassing information related to privacy concerns. Moreover, as researchers cannot compare what participants tell them with their own observations, they may miss important information, or may even find inexplicable contradictions in their data. For example, researchers often report users' privacy settings as not necessarily corresponding to privacy attitudes reported in the literature that does not provide additional data about lateral actions user may take to manage their privacy.

Researchers have made a number of suggestions to mitigate this methodological problem, notably diary studies [4], walkthroughs, or collecting data on participants' OSN usage through third-party-application interfaces [5]. As these approaches may be intrusive with respect to study participants, are not easily scalable, challenge participants' skills, or do not provide comprehensive data, we propose an alternative approach: we show how an in-situ data collection by means of a privacy-preserving web browser plugin installed on participants' web browsers combined with focused interviews allows us (a) to gather rich data on participants' practices, (b) to focus interviews on aspects that matter to interviewees, and (c) to contrast interview statements and collected data to better understand privacy strategies.

A Privacy-Preserving Web Browser Plugin to Capture Participants' OSN Usage Behavior

Before we interview participants in our studies – all Facebook users - we collect data on their Facebook use practices with a web browser plugin we developed for research purposes. This plugin consists of three major components: (1) a tracking engine running in the background while participants use Facebook, (2) a comment function participants can use to comment in situ on what kept their attention, and (3) a control center enabling participants to check the collection process and to look at the gathered data in human-readable form (see Figure 1).

The plugin's design followed three principles to preserve participants' privacy: first, participants can check the collected data at any time, and can exclude data records from being handed over to the researchers, should they wish so. Second, researchers have no way to access the plugin and collected data at participants' computers. Instead, participants have to email collected data to the researchers after checking that the data records do not contain private content they do not wish to submit. Third, the plugin does not collect any shared content, chat messages, names of contacts or similar information. Instead the plugin stores pseudo-unique identifiers referring to items affected by user actions and to item owners.

Type and meaning of data tracked by the web browser plugin during the study

Timestamps and action types: At what time do participants use particular functions? Which functions do participants use in combination, e.g., in quick succession?

Target item types: What are targets of users' actions, e.g., what do they comment or "like"?

Target item IDs: Does a participant interact with the same item multiple times, e.g., "likes" or comments on the same post? In peer groups, do different group members interact with the same item?

Target item owners: Do participants interact repeatedly with items created by specific other users?

Privacy Settings: Which privacy settings do users apply?

Target person IDs: Who is target of a participant's action, e.g., who is a message recipient or was accepted as a new "friend"?

In-situ comments: What was the purpose of or reason behind an action?

	Participants															Color coding			
Action	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	negligible	low	medium	high
Like item	6	6	1	4	9	15	32	48	11	11	0	33	0	1	24	0	1-12	13-20	>20
Comment status	5	2	2	3	6	4	10	6	5	11	0	8	1	0	25	0	1-5	5-15	>15
Update own status	2	0	0	0	0	0	1	0	0	0	0	1	0	0	3	0	1	>1	n/a
Share an item	0	0	0	0	0	0	5	0	3	0	1	1	0	0	1	0	1-3	>3	n/a
Send chat message	43	0	3	0	0	3	68	242	245	0	44	186	1	183	46	0-5	6-100	101-200	>200
Accept/sent friend request	2	0	1	0	1	0	1	0	3	4	0	1	0	0	1	0	1-2	>3	n/a
Open Facebook	95	46	28	76	54	178	40	239	127	125	17	32	9	46	40	0-50	51-100	101-200	>200
Settings for friends' posts and tags on one's own Timeline																restrictive			permissive
Who sees others' posts/tags						cust	cust						cust		cust	Me-Me	Frnd-Me	Frnd-Frnd	FoF-FoF
Allow posts for.../Review posts																Me-On	Frnd-On	Frnd-Off	n/a
Audience for tags/Review tags							cust								cust	Me-On	Frnd-On	Frnd-Off	n/a

Table 1: Frequency of user actions and applied general privacy settings per participant as tracked with the plugin; the table shows diverse use of Facebook functions and Facebooks' privacy settings (Abbreviations: Frnds-Friends, FoF-Friends of Friends, cust – customized)

The plugin derives these identifiers from shortened, salted hash values of user content and user names. For the researchers it is practical infeasible to reconstruct the original plaintext content or names. But they can still use the identifiers to analyze how actions relate to each other, without actually knowing the shared information. For example, researchers can investigate if participants regularly chat with a specific group of persons, or whether they comment on same Timeline items multiple times. The sidebar on this page gives an overview of the data collected by the plugin and the questions they can answer.

Study: Investigating Diverse Privacy Strategies of Facebook Users

For a first study, we recruited 10 active users of Facebook by bulletin board notices on the university campus, and a peer group of 5 members through personal recruitment by one of our researchers. Our average participant was 24.8 years old (range from 20 to 28, σ=2.3, 11 students, 4 with other occupations, 7 male, 8 female).

All participants consent to install our self-developed web browser plugin on their personal computer for two weeks after being instructed about the way it works, and about their options. We analyzed plugin-collected data regarding (1) how permissive were the privacy settings and (2) which interaction functions of Facebook were used and how was this done. After the data analysis, we conducted focused one-on-one interviews. We asked participants to (1) describe their use practices, (2) explain their privacy settings, and (3) to tell us about their experiences with Facebook. The data collected with the plugin provided cues to talk about specific use practices of the participants.

First Results

Focusing Interviews with Plugin-Collected Data

Selecting significant actions from the action flow (see Figure 2) and presenting them to interviewees proved helpful to initiate discussions on specific experiences that matter to the participants. For example, P7 reported on his use of chat-like comment threads on Timeline items only after the interviewer asked to explain a sequence of comments visible in the participant's activity data.

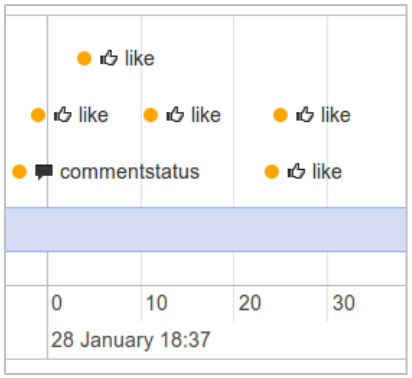

Figure 2: Sequence of one status comment and five "likes" within half a minute found in P8's data; the participant explained it with a spontaneous reaction on a birthday posting. Snapshot taken from analysis software used to evaluate plugin-collected data.

Relating Self-Reports and Actual Behavior

Plugin-collected data helped to better relate interview statements to users actual practices. For instance, while P9 stated that he did not chat a lot, he overlooked that he frequently used the chat function to stay in contact with his brother and his girlfriend (cp. Table 1). Another case in point is P10 who conveyed himself as carefully selecting his "friends". But tracked data revealed that he accepted quite a lot of friend requests during the tracking period (cp. Table 1). When discussing this fact with him during the interview he described in detail how he decides on friend requests.

Supporting Software Design

Contrasting plugin-collected data with the information on the usage context gathered from participants' interview statements provides two types of findings that may support software design: (a) software functions experienced by users as interfering with their practices can be identified and improved; (b) we may obtain information on use contexts that could be applied by context-sensitive software functions to mitigate users' problems. For example, we observed situations where commenting evolved into chat-like conversations of a small group of users. OSN software may detect such situations and could suggest group chats instead to protect users' privacy.

Conclusion and Outlook

Our first results seem promising: with our methodology we were able to reveal what users actually do in OSN by contrasting collected data with participants' statements on their usage context gathered in tailored interviews. We consider both information to suggest ideas for improving OSN software. However, several issues remain to be tackled: if participants decide not to hand over specific data to the researchers, we are not provided with supportive data to understand and discuss use practices with them. So far, this did not occur in our studies. Second, we cannot recognize some privacy strategies in our collected data such as self-censorship within postings. This is a limitation due to the privacy-respecting design of our software.

In the future, we plan to use our methodology to investigate both how peer groups interact in OSNs and in which situations users prefer one OSN to another. Future research might include tracking actions in other OSNs, could include quantitative surveys, or could focus on a subset of Facebook functions with refined action tracking. Our software is available for download on our project website http://dipo.sit.fraunhofer.de.

References

[1] boyd, D., Marwick, A.E. Social Privacy in Networked Publics: Teens' Attitudes, Practices, and Strategies. In *A Decade in in Internet Time: Symposium on the Dynamics of the Internet and Society.* 2011

[2] Rotman, D., Preece J., He, Y., Druin, A. Extreme Ethnography: Challenges for Research in Large Scale Online Environments. In *Proc. 2012 iConference*, ACM (2012), 207–214

[3] Wisnewski. P.J., Richter Lipford, H., Wilson, D.C. Fighting for My Space: Coping Mechanisms for SNS Boundary Regulation. In: *Proc. CHI'12*, ACM (2012), 609-618

[4] Sleeper, M., Balebako, R., Das, S., McConahy, A.L., Wiese, J., Cranor, L.F. The post that wasn't: exploring self-censorship on facebook. In: *Proc. CSCW'13*, ACM (2013), 793-802

[5] Madejski, M., Johnson, M., Bellovin, S.M. A study of privacy setting errors in an online social network. In: *Proc. SESOC'12*, IEEE (2012)