APPLYING CONTEXTUAL INTEGRITY TO THE CONTEXT OF SOCIAL NETWORKING SITES TRACKING

Rath Kanha Sar and Yeslam Al-Saggaf

Abstract
Behavioural tracking is very common and many studies have shown that a user's online movements can be tracked by various parties including advertisers and data aggregators. The findings of our experiments showed that social network sites (SNS), particularly Facebook, Twitter, LinkedIn and Google Plus, have the ability to not only collect and store large volume of a user's information, but also transfer user's data to third party sites and also track that user's movement within and beyond SNS boundary, particularly among web sites embedding SNS widgets. In this paper, we analysed the privacy issue of online user's tracking by SNS from the perspective of Helen Nissenbaum's Contextual Integrity. Our aim was to answer to the question of whether or not an online user's privacy is violated by such practice.

Keywords
Contextual integrity, Social network sites, information leakages, online privacy, HTTP headers, HTTP cookies

1. Introduction

Social networking sites (SNS) like Facebook, Twitter, LinkedIn, and Google Plus, introduced a new means of communication where online users populate their online profiles with information such as name, date of birth, pictures, contact details, and other posts. Those SNS have become very popular recently. Facebook alone attracted more than one billion active users in mid 2012 (ABCNews, 2012) while Twitter was reported to have every U.S. Senate member signed up for a Twitter account early this year (Leggatt, 2013).

While providing a new means of communication and attracting millions of users, SNS became of great interest to the research community and it has been demonstrated through various studies that SNS leak and share users' information to third party sites such as advertisers and data aggregators, particularly via the HTTP headers (Krishnamurthy & Wills, 2008; Soltani, Canty, Mayo, Thomas, & Hoofnagle, 2009; Krishnamurthy & Wills, 2010a, 2010b; Mayer, 2011; Krishnamurthy & Naryshkin, 2011). Those studies were conducted by investigating large number of sites across many categories. We believe the results of those studies are important for showing the nature of information leakage from SNS and non-SNS to third party sites; however, these studies hardly reflect users' browsing habits in real life because users are not likely to visit large number of either SNS or non-SNS (e.g. 10 SNS) at a time. Instead, users tend to have a combination of online activities such as checking emails, using SNS, reading online news articles, performing online searches and doing online shopping (Purcell, 2011), within some selected sites.

Despite the fact that privacy issues in SNS have caught the attention of research community, research that examines the moral aspects of privacy on the basis of an empirical study is scarce. Answering questions regarding online privacy can be challenging. However, Helen Nissenbaum (Nissenbaum, 2011) suggests that analysing privacy online should be like those about privacy in general because Internet is not a distinctive sphere; rather, our online activities, mediated by Internet, are deeply integrated into our social life, and they are as diverse as our offline activities. She proposes a privacy benchmark known as *Contextual Integrity* (CI) in order to respond to privacy challenges posed by new emerging technologies like Internet.

This work aims to examine privacy issues, particularly those raised by SNS, by analysing the results of an experimental study in light of Nissenbaum's CI. Our experiments demonstrated that SNS, such as Facebook, Twitter, and Google Plus, have the ability not only to disseminate users' information to

third party sites, but also track users' movement across different sites. CI has been used to analyse privacy online in several contexts, for example, in the context of public records online, and the radio frequency identification (RFID) (Nissenbaum, 2004), in the context of data mining (Nissenbaum, 2004; Tavani, 2007), in the context of blogosphere (Grodzinsky & Tavani, 2010), and in the context of cloud computing (Grodzinsky & Tavani, 2011). To our knowledge, there is no study that has employed CI as a framework to analyse the privacy issues from the practice of SNS sharing users' information and tracking users' movements. Therefore, the aim of this study is to find out whether or not such practices by SNS companies violate the norms of information by using CI as a framework.

The paper is organised as follows. Section 2 summaries the findings of our experiments. We provide a brief background and description of CI framework in Section 3. In Section 4, we apply CI's informational norms to the findings and conclude in Section 5.

2. Social networking sites sharing and tracking users' information

We conducted Wireshark experiments aiming to identify the privacy implication of browsing the Internet within several browsing sessions by analysing the HTTP headers resulting from the first author of the article's browsing of various most frequently visited sites.

We first decided on the set of most common browsing activities (such as checking e-mails, doing online shopping, using SNS, performing online searches and reading online news articles (Purcell, 2011)). Then we selected the sites associated with the chosen activities by using the rankings in Alexa (www.alexa.com). In terms of online searching, we relied on the data provided by Google statistics for the frequently searched terms. We analysed the HTTP headers resulting from the browsing activities, and reported on the types of information being shared and to whom.

We found that, within just one browsing session of the selected sites, user's identifiable and non-identifiable information are being leaked or shared from first party sites being SNS and non-SNS to various third party sites, and those pieces of information were also seen to be transferred from third party to other third party sites. For example, when the first author looked up for *vanilla cup cake* recipe on *taste.com.au*, her searched term, *vanilla cup cake*, was seen to be transferred to various third party sites such as *doubleclick.net*, *Facebook*, and *Twitter*.

Our results also show that SNS, especially LinkedIn and Facebook, leaked users' identifiable information (e.g. name and SNS ID) to third party sites, and moreover, Google Plus, Twitter, and Facebook also have the ability to track users' browsing activities not only within but also beyond SNS, particularly among web sites that use SNS widgets (e.g. Google Plus's Plus one button, Facebook's Like button, and Twitter's Tweet button). SNS themselves are now holding not only large amount of personal information provided by SNS users, but also information about users' movement across different sites.

3. Contextual Integrity

While pointing out that "privacy in public ", which has been excluded or ignored in the past, is worthy to be studied and protected in the information age (Nissenbaum, 1997), and that previous philosophical and legal theories of privacy offer limited justification and mechanisms for dealing with the problem of privacy in a non-intimate real or public (Nissenbaum, 1998), Nissenbaum (2004) introduced a privacy framework called *contextual integrity*.

Contextual integrity focuses on the notion of "context "to analyse and evaluate whether or not the information gathering and dissemination is appropriate with the norms within a specific context (Nissenbaum, 2004; Barth, Datta, Mitchelle, & Nissenbaum, 2006; Nissenbaum, 2010). Nissenbaum believes that almost everything we do, act, and interact with others, take place in contexts. By contexts, she refer to "structured social settings with characteristics that have evolved over time (sometimes long periods of time) and are subject to a host of causes and contingencies of purpose, place, culture historical accident, and more "(Nissenbaum, 2010, p. 130).

She also adds that people navigate between different contexts (e.g. education, health care, employment, religion, family, and the commercial marketplace) throughout the day. Each context is characterised by different *roles*, is partly constituted by canonical *activities* that are oriented around *values*, and is governed by behaviour-guiding *norms* that prescribe and proscribe acceptable actions and practices. For example, in the *context* of education, Jane Doe's *role* is a teacher of mathematics in a public university where her job *activities* include giving lectures and marking the assignments and exam papers. The *value* of the education or her *role* as a teacher is to transmit the knowledge to the students. The *norms* at her workplace prescribe that Jane as well as other teachers design the subjects in the curriculum and prepare report cards (Nissenbaum, 2010).

Among the norms present in most contexts are ones that govern the flow of information about people involved in the contexts. Therefore, Nissenbaum (2004) posits two types of informational norms in her privacy scheme: (a) norms of appropriateness and (b) norms of distribution.

Norms of appropriateness dictates the types or nature of information about an individual that is allowable, expected or demanded to be revealed in a particular context. For example, in the context of education, it is appropriate for the admission office to know the students' past grades or GPA before granting them admission to a course of study, but it is not appropriate that they also need to know the students' family or relationship matter.

Meanwhile, *norms of distribution* govern the flow of information from one party to another - whether or not that distribution of information respects contextual norms of information flow. For example, within the health care context in Australia, it is appropriate that a patient discloses a medical condition to her general practitioner (GP) and the GP is expected to keep this piece of information confidential. It is also appropriate that the GP discloses and discusses about her medical condition with a specialist if needed; however, it is not appropriate if the GP emails her medical record to her employer.

In brief, the contextual integrity of the flow of information is maintained when both kinds of norms are respected, otherwise, a breach of privacy occurs.

4. Contextual Integrity and SNS tracking

According to the theory of contextual integrity, it is important to know the context and to identify several variables that are involved in the information flow such as the agents (who is gathering the information, who is analysing it, and who is disseminating it and to whom), the nature of the information, and the relationships among the various parties (Nissenbaum, 2004).

Facebook, Twitter, LinkedIn and Google are the main agents who have the ability to gather information, analyse it, and also disseminate it to various third party sites. The nature of the information is both identifiable (for example, LinkedIn user's name transferred to third party site like *b.scorecardresearch*) and non-identifiable information (for example, user's browsings in *taste.com.au* transferred to Facebook). Those SNS provide communication services to users, but at the same time, disseminate different pieces of information to the other agents like the advertisers and the third party application (e.g. Farmvile game application). In addition, they also have the ability to track users' movements outside the SNS. This is in addition to the fact that they already hold large amounts of personal information given by users. Does this mean SNS violate users' privacy?

In order to analyse the concerns of privacy online, Nissenbaum also proposes two recommendations derived from the contextual integrity framework (Nissenbaum, 2011). According to her, we need to locate the context online and explicate the information norms from that context. First, in an online activity, one should look for the similarity to social activities and structures. By locating the similarity, the norms or restraints on the flow of information can be drawn out. Second, without having a need to look for the comparison of online activity to the real life, she suggests that we can work out the norms by looking at the ends, purposes and values of the site or organisation. For example, when we make a bank transaction over the phone, or face-to-face with a bank teller, or online, there is a common expectation that the transaction should be done in a confidential and secure manner.

Unlike other contexts like the online bookstore, or the online pharmacy, SNS is a unique phenomena, and it is not easy to think of a comparable real life context. Looking at SNS structure, we observe that

SNS consists of multiple profiles, and each SNS profile usually comprises of different features such as profile information, wall post, album pictures, private message, and advertisement banners. Every post or activity on SNS profiles are recorded and are retrievable. The notion of a community appears to be a useful term given an SNS community has private spaces (profiles) like houses in a real community.

Let us compare an SNS profile (e.g. Facebook, Twitter, LinkedIn and Google Plus profile) to a house in a community. The process of one signing up for an SNS account and getting an SNS profile is similar to a process of one buying a house within a community. Each house comprises of different features and rooms or sections; for example, a living room area where user can hang out and interact with her friends, or her friends can leave her messages at the door when she is not present in the house, photo albums, and a mailbox where one can drop a private message to the house owner.

Making one's SNS profile private is similar to user using the key to lock up the house to limit access from public or random people. A public profile is like an unlocked house where anyone can just walk in. If the community owner provides different level of privacy setting, user has the ability to provide a different level of access to different people they befriend with.

This comparison shows that an SNS profile can be considered as private as a house where it is governed by some norms; for example, permission is needed to get into one's house. Within a community, it is appropriate that the community's mayor conducts census to get the information about the size of the population, what they do and how they live. The census data is very useful and necessary for a community leader to make informed decisions, and to support planning, administration and policy development. The census is also governed by its norms. For example, it is clearly stated in the Australian Bureau of Statistics (ABS) that any piece of identifiable information in the census data is removed and the personal information provided is kept strictly secured and confidential and is not released to other party outside ABS (ABS, 2012).

However, SNS community in this case is like a small village (place) where everyone within a community can know as much as they want about everyone else. The cookies (from both SNS and non-SNS) resided in a user's computer are seen to be like many spies who would follow each resident in every step they take - either within or outside the community (e.g. to a supermarket, or a bookstore). The information gathering by the SNS goes beyond the census norms because SNS not only collect information about user (both within and outside the community boundary), but also share user's information to third party sites.

According to the norms of appropriateness, it is appropriate that SNS community collect information about its own residents within the community, but it is not appropriate that they also collect information about the residents' movements outside the community (e.g. Facebook is able to know that users visit *taste.com.au* via Facebook's Like button). In addition, according to the norms of distribution, it is not appropriate that SNS community owner shares users' information, especially identifiable information, to third party organisation (e.g. LinkedIn user's name is shared to *b.scorecardresearch*). Applying Nissenbaum's contextual integrity to the context of SNS in this case study allows us to see that both norms are breached, thus, leads us to conclude that user's privacy is being violated by SNS.

5. Conclusion

The results of our experiments reveal that a user's visit to both SNS and non-SNS results in multiple HTTP requests to fetch different contents for the requested pages. Those HTTP requests are seen to distribute user's information within them, including identifiable and non-identifiable information, to third party sites.

In this study, the results of our experiment showed that SNS, like Facebook, LinkedIn, Twitter, and Google Plus, have the ability not only to collect and analyse users' information, but also record users' browsing activities across different sites, especially those who embedded SNS widgets (e.g. Facebook's Like button).

By using Nissenbaum's contextual integrity and by looking for the context and the variables involved in the information flow, it is clear that SNS like Facebook, LinkedIn, Twitter, and Google Plus are the main agents who have the ability to gather, analyse, disseminate users' information to third party sites.

In addition, the comparison of SNS to a real life community, specifically a profile to a house, helps us identify some norms which may govern the community and the data collection within that community context. According to the norms of appropriateness, it is appropriate that SNS collect information about their users within the site, but it is not appropriate that they also collect information about their users outside the SNS boundary. Meanwhile, according to the norms of distribution, it is appropriate that SNS use information collected to assist in maintaining the site traffic and performance, and improve users' experiences in SNS, but it is not appropriate that they also share these types of information with third party organisations. The analysis based on contextual integrity allows us to see that both norms are breached, thus, leads us to the conclusion that users' privacy is being violated by SNS.

References

ABCNews. (2012, October 5). *Facebook hits billion users amid revenue worries.* Available from http://www.abc.net.au/news/2012-10-05/facebook-hits-billion-users-amid -revenue-worries/4296792\?section=business

ABS. (2012). *About the census.* Australian Bureau of Statistics. Available from http://www.abs.gov.au/websitedbs/censushome.nsf/home/census?opendocument&navpos=10

Barth, A., Datta, A., Mitchelle, J. C., & Nissenbaum, H. (2006, may). Privacy and contextual integrity: framework and applications. In *2006 ieee symposium on security and privacy* (p. 15 -198). Berkeley/Oakland, CA: IEEE.

Grodzinsky, F. S., & Tavani, H. T. (2010). Applying the contextual integrity model of privacy to personal blogs in the blogosphere. *International Journal of Internet Research Ethics*, *3*(12).

Grodzinsky, F. S., & Tavani, H. T. (2011, October). Privacy in the cloud: applying nissenbaum's theory of contextual integrity. *SIGCAS Comput. Soc.*, *41*(1), 38–47. Available from http://doi.acm.org/ 10.1145/2095266.2095270

Krishnamurthy, B., & Naryshkin, K. (2011, May). Privacy leakage vs. protection measures: the growing disconnect. In *Proceedings of the web2.0 security and privacy workshop* (p. 1-10). Oakland, CA USA.

Krishnamurthy, B., & Wills, C. E. (2008). Characterizing privacy in online social networks. In *Proceedings of the first workshop on online social networks* (pp. 37–42). New York, NY, USA: ACM. Available from http://doi.acm.org/10.1145/1397735.1397744

Krishnamurthy, B., & Wills, C. E. (2010a, January). On the leakage of personally identifiable information via online social networks. *SIGCOMM Comput. Commun. Rev.*, *40*, 112–117. Available from http://doi.acm.org/10.1145/1672308.1672328

Krishnamurthy, B., & Wills, C. E. (2010b). Privacy leakage in mobile online social networks. In *Proceedings of the 3rd conference on online social networks* (pp. 4–4). Berkeley, CA, USA: USENIX Association. Available from http://portal.acm.org/citation.cfm?id=1863190.1863194

Leggatt, H. (2013). *U.S. congress embraces twitter - 100% of senate members signed up.* BizReport. Available from http://www.bizreport.com/2013/01/us-congress-embraces-twitter ---100-of-senate-members-signed-u.html

Mayer, J. (2011, August). *Tracking the trackers: where everybody knows your username.* Standford Law School: The centre for Internet and society. Available from http://cyberlaw.stanford.edu/ node/6740

Nissenbaum, H. (1997). Toward an approach to privacy in public: challenges of information technology. *Ethics and behavior*, *7*(3), 207 - 219.

Nissenbaum, H. (1998). Protecting privacy in an information age: the problem of privacy in public. *Law and philosophy*, *17*, 559 - 596.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review Association*, *79*(1), 119 - 158.

Nissenbaum, H. (2010). *Privacy in context: technology, policy, and the integrity of social life*. Stanford, California: Stanford University Press.

Nissenbaum, H. (2011). A contextual approach to privacy online. *The journal of the American Academy of Art and Sciences*, *140*(4), 32-48.

Purcell, K. (2011, August 9). *Search and email still top the list of most popular online activities.* Pew Internet. Available from http://pewinternet.org/ /media//Files/Reports/2012/PIP Digital differences041312.pdf

Soltani, A., Canty, S., Mayo, Q., Thomas, L., & Hoofnagle, C. J. (2009, August 10). Flash cookies and privacy. *Social Science Research Network*. Available from http://ssrn.com/abstract= 1446862orhttp://dx.doi.org/10.2139/ssrn.1446862

Tavani, H. T. (2007). Philosophical theories of privacy: implications for an adequate online privacy pol- icy. *Metaphilosophy*, *38*(1), 1–22. Available from http://dx.doi.org/10.1111/j.1467-9973 .2006.00474.x