Privacy Threats Related to User Profiling in Online Social Networks

Fredrik Erlandsson, Martin Boldt, Henric Johnson
School of Computing
Blekinge Institute of Technology, Sweden
{fredrik.erlandsson, martin.boldt, henric.johnson}@bth.se

Abstract—The popularity of Online Social Networks (OSNs) has increased the visibility of users profiles and interactions performed between users. In this paper we structure different privacy threats related to OSNs and describe six different types of privacy threats. One of these threats, named public information harvesting, is previously not documented so we therefore present it in further detail by also presenting the results from a proof-of-concept implementation of that threat. The basis of the attack is gathering of user interactions from various open groups on Facebook which then is transformed into a social interaction graph. Since the data gathered from the OSN originates from open groups it could be executed by any third-party connected to the Internet independently of the users’ privacy settings. In addition to presenting the different privacy threats we also we propose a range of different protection techniques.

Index Terms—Online social network, privacy, user profiling, social interactions

I. INTRODUCTION

In the beginning of 2012 Facebook had about 800 million users and the company was valued to over 100 billion dollars which to large extent originate from advertisement and user profiling possibilities based on user interaction. Besides Facebook there are a number of different OSNs that has reached a considerable user-base, e.g. Google+, Twitter and LinkedIn.

It is therefore important to address the privacy implications of how the published information within OSNs is handled. Information that is published by users within a limited group, or perhaps shared with a single user is often of a nature that can cause significant inconvenience, or even harm to concerned users. As OSNs grow in size the methods and knowledge among its users about how to configure privacy settings is crucial. In this paper we list different privacy threats within OSNs together with potential protection mechanisms. In addition to this we also add a new privacy threat that originates from scraping publicly available information which is published in open groups within the OSN.

OSNs like Facebook, Google+ and Twitter all provide open interfaces (i.e. APIs) for third-party applications to interact with the OSN by accessing and publishing data. This is very convenient for the user as it open up possibilities for value-increasing applications to interact directly with the social network. For instance, there are more than 500'000 third-party applications, such as online games, that interact and coexists with Facebook. What people does not reflect upon is the fact that most of these applications have the abilities to interact with the OSN on behalf of the user, which also includes the possibility to gather information that the user posted as private correspondence. Add to this that users on OSNs share information that could be harmful for the user itself, or even the user’s friends. As an effect Trojan applications that use deceptive and covert behavior can gather such sensitive information from users. However, Trojan applications can also retrieve information among a user’s friends including their posts, which threaten the privacy of the OSN users.

II. PRIVACY THREATS

In this section we will present six different types of privacy threats illustrated in Fig. 1. All of these threats result in user information leakage from the OSN to third parties. These privacy threats exist because social information about OSN users have a value, and can be refined into revenues within the context of targeted advertisements etc.

A. OSN Information Leakage

The first type of privacy threat, illustrated in case (a) in Fig. 1, is based on that the owner of a OSN, e.g. Facebook or Google, continuously gather detailed information regarding users activities within the OSN. This is probably the most obvious privacy threat and as such it is well known within research community and it is also the threat that OSN users first come to reflect upon [1], [2]. We therefore expect OSN users to understand that information they share within the OSN, e.g. user profile content, messages, and photos, can be mined, refined and sold by the owner of the OSN. Exactly how the OSN owner is allowed to use and benefit from this information is regulated within policy documents, e.g. the statement of rights and responsibilities and the data use or privacy policy for Facebook. A problem is that a large extent of OSN users don’t reflect upon how their interaction within OSNs affect their privacy, which could be a threat to their privacy [3]. As a natural consequence these users do not bother to investigate the content of the OSN policy documents.

There is also a risk that the OSN infrastructure get compromised, giving third parties unauthorized access to sensitive information [4].

B. Friend-in-the-Middle Threat

Case (b) in Fig. 1 shows a type of privacy threat where user information is leaked through a trusted friend within the OSN. Because of this threat the OSN infrastructure often provide users the possibility limit their posts and information spread to smaller group, which (if used correctly) could be used as one method for avoiding public scrutinization. Unfortunately a chain is not stronger than its weakest link, which goes for friendships within OSNs as well. A large portion of OSN users act irresponsible by more or less allowing anybody to establish a friendship, which not only affect the user but potentially also that particular user’s friends.

One must also consider the current state of social gaming, where users require a certain number of friends in order to achieve certain tasks (level up)\(^4\). This tend to cloud users’ judgements regarding whom they are accepting as friends, as they instead focus on the primary task ahead, i.e. leveling up.

C. Trojan Application

The third type of privacy threat is associated with Trojan applications leaking information about its OSN users to third parties, see Fig. 1 case (c) [5]. The user is deceived to install a Trojan application which claims to provide some desired functionality, but also hides unwanted and shady behavior, and as a result leak valuable information.

D. Public Information Harvesting

Case (d) in Fig. 1 illustrates a new type of threat that we present in this paper, and as such it is previously unknown within both academia and among OSN users. The basis of the threat is that third parties collect user information published in open groups within OSNs like Facebook. Such open groups exist in the boundary between the OSN and the publicly available Internet. Since the information is gathered from open OSN groups there is no need for using covert or deceiving methods when collecting the information. It is simply a matter of scraping the information available on these web pages, which can be done by anyone connected to the Internet. Using the harvested information it is possible for third parties such as profit-driven companies or national security agencies to create social interaction graphs, which details how users interact among a certain topic, e.g. the Occupy Wall street movement. This privacy threat is described further in Section III.

E. Socialbot

Recently automated software programs, called socialbots, have been seen influencing OSN users [6]. These socialbots are designed to control OSN accounts, by autonomously performing basic tasks such as posting messages and sending friend requests. Socialbots are not applications within the OSN itself, but rather software programs that impersonate the human beings behind user accounts by imitating human behavior towards the OSN, and as such the socialbots fool both the OSN infrastructure itself and the users populating it. Socialbots with these features have been seen infiltrating private and trusted areas shared by Friend relationships in Facebook, and as a consequence harvesting sensitive data from the concerned user accounts.

The threat from socialbots increase since many users are irresponsible when accepting new friend requests from unknown users. In a practical demonstration a socialbot were accepted as friend by OSN users at a rate of 19.3% out of 4493 requested users during the initialization phase and by 59.1% during the socialbot’s propagation phase [7]. Given this high acceptance rate regarding unknown users’ friend requests it is questionable what the effect of privacy settings that limit information access to friends, or friends-of-friends within a OSN really have in practice. If a user’s friend is routinely accepting friend-requests from unknown sources, this friend is a privacy threat, even though this might be unintentional, to both himself and his friends. With respect to our privacy we have therefore come to a situation where we no longer can fully trust the integrity of our friends within OSNs.

F. Friend-in-the-Middle Trojan Application

This type of threat is indirectly affecting a user when one of the user’s friends add a deceptive Trojan application. The effects on the user and the user’s friends privacy is similar to Trojan application threat described previously. As such, a user’s privacy is dependent not only on his/her own ability and judgement, but also on his/her friends competences, or even weakest friend in this regard.

III. PROOF-OF-CONCEPT

The threat we describe as public information harvesting is based on that users within Facebook can interact in open groups that are publicly available from the Internet. User

---

\(^4\)(2011, Nov.) How to increase Facebook Farmville levels and farm dollars fast. http://www.ehow.com/how_5466674_increase-levels-farm-dollars-fast.htm
interaction within these groups is in the form of “Likes”\(^5\) on the group itself, comments within the group, or “Likes” on other users comments. By systematically gathering this public information it is possible to create interaction profiles identifying and profiling users based on the interactions made, i.e. through social interaction graphs as shown in Fig. 2.

**A. Gathering of Information from Open Groups**

Facebook provide different methods for third-party application interaction, for instance using the Graph API\(^6\). The use of this API is straightforward, in a few hours we built an application acting as a data extraction tool that gathered information as an authenticated user on Facebook. Then we created a dummy-user without any interactions or affiliates to begin with. Next our newly created dummy-user accepted our application with just basic permissions. It was then through this dummy-user's application we gathered data from various open groups on Facebook. However, it is important to stress that the content of open groups are freely available on the Internet so there is no requirement of using a dummy-user to extract this information, we only used it due to convenience reasons.

The information gathered have traditionally been seen by research community as simple post and user information. We have however seen that the information gathered follows such a structured form that different users’ interactions can be combined and form a social interaction graph. Any third-party can gather this user information independent on the user’s privacy settings without their knowledge.

**B. Creating a Social Interaction Graph**

From the information gathered in the previous step we created a social interaction graph shown (Fig. 2). The figure shows the interactions between different networks before (Fig. 2 (a) ) and after (Fig. 2 (b) ) the Pepper-spray incident that happened in Davis, CA. This pepper-spray incident resulted in not only more intense interactions, but also that users involved in their representative community started to interact with other “Occupy” groups.

When looking at the created social interaction graph we can conclude that even if a user have strict privacy settings the user’s actions are hard to hide. We were able to gather not just the name of the users, but also the profile ID making it possible to find out more information about the human behind the user account. The users in today’s OSNs must understand that no matter how strict they are trying to protect their user profile with policies, they are still at risk of being profiled based on their behaviors in various groups.

**C. Privacy Implications**

Since public information harvesting can be carried out by basically anyone it definitely pose a threat to user privacy.

---

5“Like” is a term found in Facebook where an user can show that they agree or in other way would like to show that they share the same thought as the message, this is called +1 in Google+.

One such example is countries where the regime is interested in targeting and monitoring citizens engaged in various issues that are uncomfortable for the regime. For users living in countries that respect human rights the threat might come from corporations and advertisers to larger extent.

IV. PROTECTION MECHANISMS

In this section we suggest different protection mechanisms against the threats described in section II. This list of protection mechanisms is with no means complete. Using encryption for instance it would be possible to address several of these threats if the OSNs could act and facilitate PKI, but due to the space-limitation we exclude that in this paper.

A. Information leakage from the OSN

Since it is impossible to reach absolute security in any system it is important to inform the OSN users in an adequate manner regarding how their information is handled. Here public discussions to raise user awareness is an important component. It should also be possible to benefit from existing techniques to increase transparency of the OSN policies towards the users, e.g. “Privacy Simplified”\(^7\) that help summarize privacy policies using standardized icons. In addition to improving user awareness it is of paramount importance that the OSN infrastructure is properly secured, and that a continuous security process is established.

B. Trojan applications

To improve the protection against Trojan applications, case (c) and (f), we suggest the use of an application certification and reputation program, which just recently has been announced by Facebook under the name “App Center”\(^8\). We suggest that a more privacy-driven application certification program is added to this initiative, where not only the overall application quality is evaluated, but also the privacy-implications of the data gathered. Combined with a privacy policy of what data the application will retrieve and how the application will handle this information would make a valuable addition. It is also preferable that interested users should have the possibility to see a audit trail of the interactions the application has carried out on behalf of the user [8].

C. Questionable friends

The most important issue to focus on is the lack of user awareness about the problem shown in case (b) and (e), which could be addressed through end-user education. Instructing the users about socialbots, Trojan applications, and the implications of the “Friends-of-Friends” privacy setting. Users should also be instructed to keep their friend-list up to date as far as possible.

D. Social interaction profiling

To protect against social interaction profiling, case (d) in Fig.1, we suggest the use of pseudonyms or virtual profiles [9]. However, by hiding the real identity of the end-user, for instance using anonymity techniques, will also remove one fundamental value of the OSN, i.e., that the OSN transcends the real world since each user accounts (more or less) corresponds a human being. By using pseudonyms it is possible for a user to interact under separate pseudonyms in different open groups, which renders it impossible to make connections between different groups at least.

V. CONCLUSION

In this paper we present different privacy threats in OSNs. One of these threats is previously undocumented and we therefore describe this threat in more detail together with the results from our own proof-of-concept implementation, which includes the resulting social interaction graph that could be used for user profiling. The proof-of-concept shows that with limited resources it is possible to profile users within an OSN through open groups and then build a social interaction graph of their interactions. Any user within the OSN is vulnerable to this threat, independent on their privacy settings. Finally we suggest a number of different protection mechanisms against the threats identified.

REFERENCES
