Collateral Damage of Online Social Network Applications

Iraklis Symeonidis\textsuperscript{1}, Pagona Tsormpatzoudi\textsuperscript{2} and Bart Preneel\textsuperscript{1}
\textsuperscript{1}KU Leuven, ESAT/COSIC and iMinds, Belgium
\textsuperscript{2}ICRI/CIR, KU Leuven and iMinds
\{first.lastname\}@esat.kuleuven.be, pagona.tsormpatzoudi@law.kuleuven.be

Keywords: Online Social Networks, Applications, Facebook, Privacy by design, Privacy Risk

Abstract: Third party application providers in Online Social Networks can collect personal data of users through their friends without the user’s awareness. In some cases, one or more application providers may own several applications and thus the same provider may collect an excessive amount of personal data, which creates a serious privacy risk. Previous research has developed methods to quantify privacy risks in Online Social Networks. However, most of the existing work does not focus on the issues of personal data disclosure via the user’s friends applications and application providers. The aim of this paper is to investigate the need for solutions that can compute privacy risk related to applications and application providers. In this work we perform a legal and technical analysis of the privacy threats stemming from the collection of personal data by third parties when applications are installed by the user’s friends. Particularly, we examine the case of Facebook as it is the most popular Online Social Network nowadays.

1 INTRODUCTION

Online Social Networks (OSNs) have altered the social ecosystem in a remarkable way. OSNs supports a plethora of applications (Apps) providing games, lifestyle and entertainment opportunities developed by third parties. Such Apps may disclose a user’s personal data from their online friends to third party application providers (AppPs) (Chaabane et al., 2012; McCarthy, 2014). In some cases, AppPs may own several Apps, through which they may get access to larger amounts of personal data. As a consequence, personal data disclosure may pose significant privacy risks and prompts serious concerns among the users, the media and the research community (Consumerreports, 2012; Krishnamurthy and Wills, 2008; Wang et al., 2011).

When a user shares information with her online friends in OSNs the user is not aware whether a friend has installed Apps that may access her personal data. Moreover, she is not aware that her personal data are further exposed to the AppPs. We define as collateral damage the privacy issues that arise by: (1) the acquisition of users’ personal data by the Apps installed by a user’s friends, (2) the acquisition of personal data of a user by the AppPs, who may disclose this data outside the OSN ecosystem.

From a legal point of view, these privacy issues cause concerns with respect to data protection legislation. Data protection law provides a framework for the protection of the users’ fundamental rights and in particular the right to privacy during processing of personal data. Personal data, pursuant to Article 2(a) of the Data Protection Directive 95/46/EC (95/46/EC, 2015) “refers to any information relating to an identified or identifiable natural person (i.e., user)” which in the scope of this paper corresponds to the user’s measurable personal data processed by the App. Data protection law may be infringed in relation with two aspects: (1) data processing lacks legitimacy as the user has not given her consent to the App processing her personal data, and (2) data processing lacks transparency as the user may be totally unaware of the data processing that may take place.

There exists a considerable amount of related work regarding the user’s privacy in OSNs, including work that investigates installed Apps (Biczók and Chia, 2013; Chaabane et al., 2012; Chia et al., 2012; Huber et al., 2013). However, there is no prior work analyzing the case where a set of AppPs collects the user’s personal data via the Apps installed by the user’s friends. Moreover, there is no work that estimates the privacy risk of the users in the case of the collateral damage of the Apps in OSNs.

Based on the assumption of privacy as a good practice (Diaz and Gürses, 2012), our research is
motivated by the need to reduce the collateral damage of the Apps in OSNs. Technically, we study how an evaluation of the privacy risk can help the user to better manage her privacy. We define the privacy score of a user as an indicator of her privacy risk. The higher the privacy score, the higher the threat to the user. A privacy risk computation is a Privacy Enhancing Technology. Privacy Enhancing Technologies are able to raise awareness on personal data collection and may support the user’s decisions about personal data sharing (McDonnel et al., 2015). From a legal point of view, our solution aims to implement transparency following the principle of Data Protection by Default. Data Protection by Default was introduced in Article 23 (2) of the draft Data Protection Regulation, which requires mechanisms that, “by default ensure that the users are able to control the distribution of their personal data” (Parliament, 2015). Data protection by default intends to mitigate privacy risks stemming from users’ asymmetrical information (McDonnel et al., 2015). In the context of this paper, data protection solutions, such as privacy score, raise user’s awareness and enhance her empowerment.

The rest of this paper is organized as follows. Section 2 reviews the related work with respect to the privacy issues that Apps introduce in OSNs as well as the existing research related to the user’s privacy risk. Section 3 describes the case of Facebook OSN. Section 4 concludes and proposes the future work.

2 RELATED WORK

This section describes the related work on privacy issues that arise from the use of Apps in OSNs. Moreover, it describes the existing work on the computation of privacy risk in OSNs. Currently there exists work related to the privacy issues of Apps in OSNs. Chaabane et al. (Chaabane et al., 2012) showed that the Apps can have tracking capabilities and disseminate the collected information to “fourth party” organizations (Chaabane et al., 2014). Similarly, Huber et al. developed an automatic evaluation tool, AppInspect (Huber et al., 2013), and demonstrated security and privacy leakages of a large set of Facebook Apps. Furthermore, Biczók and Chia (Biczók and Chia, 2013) described the issue of users’ information leaked through their friends via Apps on Facebook. This work introduced a game theoretic approach to simulate an interdependent privacy scenario of two users and one App game. Extending the work of Biczók and Chia (Biczók and Chia, 2013) Pu and Grossklags (Pu and Grossklags, 2014) proposed a formula to estimate the payoffs. Finally, Frank et al. (Frank et al., 2012) showed the existence of malicious Apps that deviate from the generic permissions pattern acquiring more information from the users, while Chia et al. (Chia et al., 2012) showed that certain Apps collect more information than necessary.

To estimate the privacy risk for a user, Maximilien et al. (Maximilien et al., 2009) initially proposed a Privacy-as-a-Service formula. This formula is used to compute the privacy risk as the product of sensitivity and visibility of personal data. Liu and Terzi (Liu and Terzi, 2010) extended this work (Maximilien et al., 2009) and proposed a framework for computing the privacy risk using a probabilistic model based on the Item Response Theory (IRT). Although, IRT presents interesting results to compute the sensitivity of the user’s personal data, there is a lack of evaluation for the visibility. Moreover, Sánchez and Viejo (Sánchez and Viejo, 2015) developed a formula to assess the sensitivity of unstructured textual data, such as wall posts in OSNs. Their model aims to control the dissemination of the user’s data to different recipients of an OSN (Viejo and Sánchez, 2015). Minkus et al. (Minkus and Memon, 2014) estimated the sensitivity and visibility of the privacy settings based on a survey of 189 participants. Finally, Nepali and Wang (Nepali and Wang, 2013) proposed a privacy index to evaluate the inference attacks as described by Sweeney (Sweeney, 2000), while Ngoc et al. (Ngoc et al., 2010) introduced a metric to estimate the potential leakage of private information from public posts in OSNs.

To the best of our knowledge, there is currently no work that considers the case of the collateral damage of the Apps for computing the privacy risk. The existing related work is mainly focused on estimating a privacy score as an impact on the dissemination of the user’s information to the other members of an OSN. Our work is focused on the privacy impact that arises from the acquisition of users’ personal data via the Apps installed by the user’s friends and the user itself. These Apps expose the user’s personal data to the Apps, outside the OSN ecosystem, without users’ prior knowledge.

3 THE CASE OF FACEBOOK APPLICATIONS

This section analyses the collateral damage privacy issues of the Apps for the case of Facebook. Facebook is a popular OSN with more than 1.4 billion monthly active users (Statista, 2015). Facebook offers a plethora of easy-to-use tools
such as Apps developed by third party application providers (AppPs). Currently, there are over 25,000 Apps available on Facebook (Huber et al., 2013; SBA-Research, 2015). Users on Facebook are able to construct online profiles (Boyd and Ellison, 2008). A user’s profile consists of information that can be stored with the aim to be shared with other entities such as other users and Apps. For instance, on Facebook there is a list of more than twenty attributes in a user’s profile, such as “age”, “birthday”, “gender”, and “location” (Facebook, 2015).

A running App can retrieve information from a user’s profile; this information can subsequently be accessed and stored by the AppPs. For an App to access the user’s profile, an installation process has to be performed. Each App requests from the user a set of permissions, that allow the App to access and collect additional information. This is done by an access token provided by Facebook, that requires authorization from the user (steps 1 to 4 in Figure 1). After the user’s approval, Apps can collect the user’s personal data and store these data at the AppPs servers. Therefore, the user’s personal data are stored outside the Facebook ecosystem and out of the user’s control (steps 5 and 6 in Figure 1).

In order to control the visibility of the user’s personal data, Facebook offers a set of privacy settings to its users. The set of available privacy settings (Facebook, 2015) is broad, and it ranges from restricted to public, with settings such as “only me”, “friends”, “friends to friends”, “custom” and “public”. For the case of Apps on Facebook, the privacy setting “only me” restricts the visibility of the personal data to the user. However, privacy settings of “friends”, “friends of friends”, “custom” and “public” equally expose the user’s personal data to third party AppPs via their friends’ Apps, making them available to external servers.

Moreover, due to the server-to-server communication (steps 5 and 6 in Figure 1), the offline interaction between Facebook and AppPs makes any protection mechanisms hard to apply (Enck et al., 2014). As a result, the user’s profile information can arbitrarily be retrieved by AppPs without notification or approval of the user.

### 3.1 Users’ information is exposed by their friends

Initially, the API v.1 of Facebook provided a set of permissions to the Apps, such as friends.birthday, and friends.location. Those permissions gave the AppPs the right to access and collect users’ personal data via their friends, such as the user’s birthday and location. However, currently the Facebook API version v.1 is obsolete and the friends.xxx permissions are not present. The updated API version v.2 (Facebook, 2015) replaced the friends.xxx permissions with the user_friends. Although the newer API had to be in line with the regulations of EU and U.S. (95/46/EC, 2015; FTC, 2015), our analysis showed that it discloses up to fourteen user attributes via the user’s friends; maintaining the privacy concerns of collateral damage of the Apps as an open problem. A more detailed view on the available permissions of Apps is given in Table 3 in the Appendix.

Furthermore, Apps can request permissions through strangers (non–friends) who participated in the same conversation group with the user (i.e., personal messages). This is the case, for instance, for the permission read_mailbox. The mere exchange of text messages in a group conversation may disclose user personal data: when the user participates in a group conversation with other users (friends and non–friends) who has installed read_mailbox, the user’s personal data becomes accessible to the Apps. This personal data can be the content of the conversation as well as the time that the communication took place.

To examine the problem, we performed an analysis of the Apps on Facebook. We used the publicly available dataset provided by Hubert et al (Huber et al., 2013; SBA-Research, 2015). The dataset consists of 16,808 Facebook applications between 2012 and 2014. It contains the application name, id, number of active users (daily, weekly and monthly), the requested permissions and the Apps that an App owns. For this paper we analyzed the permissions of Apps with more than 10,000 Monthly Active Users (MAU). Among, these Apps we identified the proportion of permissions that cause the collateral damage privacy issue. Moreover, in order to calculate the number of personal data that Apps can collect we considered the number of corresponding permissions to that data such as friends.photos and user.photos to
From the list of 16,808 Apps on Facebook, we identified 2,200 Apps with more than 10,000 monthly active users (MAU). As described in Table 1, we verified that several Apps request permissions for collecting sensitive data, such as birthday 28.7%, photos 12.9%, likes 14.1%, location 9.5%, friend information 10.23%, and, more invasive, private mailbox privileges 0.45%. Among this sensitive data the most commonly requested friends related permissions for the Apps that enable the collateral damage issue, were friends,birthday 4.2%, friends,photos 4.4%, friends,likes 1.8% and friends,location 1.5%.

While the permissions affecting friends’ data seem limited, the lack of transparency and opt–out option (lack of consent) for the user is worrisome. Moreover, although read,mailbox appears to be used by only 0.45% of the Apps the severity of the risks it may entail for the user is significant.

<table>
<thead>
<tr>
<th>Permissions</th>
<th>≥ 100 Monthly</th>
<th>≥ 10000 Monthly</th>
</tr>
</thead>
<tbody>
<tr>
<td>email</td>
<td>50.5%</td>
<td>61.8%</td>
</tr>
<tr>
<td>user,birthday</td>
<td>19.4%</td>
<td>24.5%</td>
</tr>
<tr>
<td>user,likes</td>
<td>14.1%</td>
<td>12.8%</td>
</tr>
<tr>
<td>user,location</td>
<td>7.3%</td>
<td>7.9%</td>
</tr>
<tr>
<td>publish,actions</td>
<td>31.1%</td>
<td>50.3%</td>
</tr>
<tr>
<td>publish,stream</td>
<td>31.4%</td>
<td>19.3%</td>
</tr>
<tr>
<td>user,photos</td>
<td>10.6%</td>
<td>8.5%</td>
</tr>
<tr>
<td>friends,xxx</td>
<td>10.2%</td>
<td>10%</td>
</tr>
<tr>
<td>read,mailbox</td>
<td>0.9%</td>
<td>0.45%</td>
</tr>
</tbody>
</table>

Table 2: Facebook application providers and the amount of its corresponding applications for more than 100 and 10,000 Monthly Active Users.

Table 2: Facebook most requested permissions for more than 100 and 10,000 Monthly Active Users.

<table>
<thead>
<tr>
<th>Application Provider</th>
<th>≥ 100 Monthly</th>
<th>≥ 10000 Monthly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vipo Komunikacijos</td>
<td>163</td>
<td>99</td>
</tr>
<tr>
<td>Telano</td>
<td>136</td>
<td>118</td>
</tr>
<tr>
<td>Mindjolt</td>
<td>120</td>
<td>32</td>
</tr>
<tr>
<td>superplay!</td>
<td>81</td>
<td>8</td>
</tr>
</tbody>
</table>

3.3 Legal issues

From a legal perspective, one of the main challenges of data protection attached to the Apps permissions, as described above, is the fact that personal data processing may lack legitimacy. Article 7 of Data Protection Directive 1995/46/EC (95/46/EC, 2015) provides a limited number of legal grounds aiming to perform personal data processing, such as the “the user’s unambiguous consent”. As enshrined in Article 7a, the data controller, i.e., Facebook or Apps, may collect, store, use and further disclose the data, if the user has given her consent. For the consent to be valid, it has to fulfil certain criteria: it has to be given prior to the data processing, accepted in a “free” way and sufficiently specified and informed (Art. 29 WP 2011) (95/46/EC, 2015).

However, Facebook third party Apps can proceed to the users personal data processing only with the user’s friends’ consent. In other words, consent may be provided only by the user of the App and not by the user, whose data however will be processed in the end. On Facebook Apps settings, users allow by default their data to be used by their friends Apps, unless they manually uncheck the relevant boxes “Apps other use”. One could claim that consent has been theoretically given. However, the U.S. Federal Trade Commission required that Apps cannot imply consent, but this should be rather affirmatively expressed by the users. Despite this requirement, due to the default privacy setting on Facebook, users are totally unaware of the fact that they have to uncheck the boxes in order to prevent such data processing (FTC, 2015).

Further, with regards to the obligation of the data controller (Facebook or Apps) to transparency, it should be noted that in both cases, users have neither sufficient information about the nature and amount of data that will be collected nor about the purposes that the data will be used for. In other words data processing goes far beyond the user’s legitimate expectations. This interferes with the principle of fairness and transparency stemming from Article 10 of Data Protection Directive 46/1995/EC (95/46/EC, 2015). In relation with the same matter, the U.S. Federal Trade Commission stressed the need to keep users informed in case any disclosure exceeds the restrictions imposed.

3.2 Third party application providers

Third party application providers (AppPs) can be owners of several Apps. As a consequence, one App may collect through those Apps several personal data items of each user. The amount of personal data that can be retrieved is equal to the collection of all the acquired personal data under the same App. Moreover, every App retrieves the Facebook’s user ID which can identify a user and can be used to accurately correlate the collected personal data from each App. For instance, extending our analysis for the AppPs we identified that there are AppPs with up to 160 Apps with more than 10,000 MAU (see Table 2). Moreover, we repeated our analysis for the Apps and its corresponding AppPs that enable the collateral damage problem. We identified that “Astrologix” and “Social sweethearts GmbH” AppPs have 7 Apps and “Shine-zone” 5. 
by the privacy setting(s) of the Apps (FTC, 2015). This can be possibly the case of permissions such as user_friends.

4 CONCLUSION AND FUTURE WORK

In this paper, we analyse the importance of the privacy issues that arise from the collection of the user’s personal data that can be collected by third party Apps installed by the user’s friends in OSNs. Moreover, we analyse the case where Apps under a minor set of Apps can collect the users’ personal data and expose them outside of the OSN ecosystem without prior knowledge of the users. To demonstrate the importance of the problem we analyzed the case of the Facebook Apps.

Considering the privacy issues that arise from the installation of third party Apps, this paper performs a privacy risk assessment which is in line with the legal principle of privacy by default. It aims to illustrate how the user’s data disclosure takes place through the acquisition of users’ personal data via Apps installed by their friends in OSNs. A calculation of a user’s privacy risk can be useful to both users and researchers. A privacy risk assessment (Nebel et al., 2013) can help the privacy-aware users to better support their decisions when they install Apps. The increase of awareness on personal data collection is in line with the legal principle of data protection by default, as it can potentially support decisions and foster user control on personal data disclosure. From the researchers’ perspective, a numerical value describing the user’s information exposure would allow statistical inferences and comparisons for better privacy design.

A previous work proposed by Liu and Terzi (Liu and Terzi, 2010; Maximilien et al., 2009) developed a 2-dimensional matrix to compute the privacy risk, considering the sensitivity and the visibility of a user’s personal data to the users of an OSN. Our future work aims to extend this model also to Apps and Apps. Moreover, our analysis considers the Apps and Apps that are available at the time of writing. However, since API and Apps are rapidly evolving it would be interesting to update and extend the current dataset with the recent Apps available on Facebook.

5 ACKNOWLEDGMENTS

I notably want to thank Dr. Markus Hubert and SBA Research center for providing us with the necessary material for our study. A thank you to Andrea Di Maria, Dalal Azizy, Dr. Danai Symeonidou, Prof. Gergely Biczók, Dr. Mustafa A. Mustafa, Fateme Shirazi, Dr. Filipe Beato and all the anonymous reviewers who helped to better shape the idea and the quality of the text. This work was supported in part by the Research Council KU Leuven: C16/15/058.

6 APPENDIX

Table 3 illustrates the permissions available for the API v.1 and v.2 respectively.

Table 3: Facebook application permissions and the corresponding personal data. Permission availability to API v.1 and v.2

<table>
<thead>
<tr>
<th>Permissions</th>
<th>Personal data</th>
</tr>
</thead>
<tbody>
<tr>
<td>public_profile</td>
<td>full name, first name, last name, link, gender, locale, timezone, updated time, verified, bio, birthday, education, first name, last name, gender, interested in, languages, location, political, relationships, status, religion, quotes, website, work,</td>
</tr>
<tr>
<td>user_friends/v1</td>
<td>friends, friend activities, friend birthday, friend hobbies, friend education history, friend friends, friend games activity, friend groups, friend hometown, friend interests, friend likes, friend location, friend post, friend online presence, friend photo videos, photos, questions, relationship details, friendships, relationship status, relationship friends, relationship gender, relationship religion, relationship politics, friend status, friend subscriptions, friend website, friend work history, friend zips</td>
</tr>
<tr>
<td>user_friends/v2</td>
<td>friends, friend activities, friend birthday, friend hobbies, friend education history, friend friends, friend games activity, friend groups, friend hometown, friend interests, friend likes, friend location, friend post, friend online presence, friend photo videos, photos, questions, relationship details, friendships, relationship status, relationship friends, relationship gender, relationship religion, relationship politics, friend status, friend subscriptions, friend website, friend work history, friend zips, friend inbox</td>
</tr>
</tbody>
</table>

REFERENCES


Chia, P. H., Yamamoto, Y., and Asokan, N. (2012). Is this app safe? A large scale study on application permissions and risk signals. In WWW, Lyon, France. ACM.


