Collateral Damage in Online Social Networks: computing the significance of information collection. *

Iraklis Symeonidis[1], Faruk Gologlu[2], and Bart Preneel[1]

[1]ESAT COSIC, KU Leuven and iMinds, Belgium.
firstname.lastname@esat.kuleuven.be
[2]Claude Shannon Institute, University College Dublin.

January 14, 2016

Abstract

Third-party apps enable a personalized experience on social networking platforms; however, they give rise to privacy interdependence issues. Apps installed by a user's friends can collect and potentially misuse her own personal data inflicting *collateral damage* on the user herself while leaving her without proper means of control. In this paper, we present a study on the *collateral information collection* of apps in social networks. Based on real data, we compute the proportion of exposed user attributes including the case of profiling, when several apps are offered by the same provider.

1 Significance

In this section, we develop a mathematical model and compute the volume of the user's attributes that can be collected by apps and appPs when installed by the user's friends. Our calculations are based on several snapshots of the most popular apps on Facebook using the Appinspect dataset [1].

OSN, users and users' friends. Let an Online Social Network (OSN) with k users and the corresponding set denoted as the set $\mathcal{F}$, i.e., $\mathcal{F} = \{u_1, \ldots, u_k\}$. The user under consideration is denoted by u, such that u is an element of $\mathcal{F}$, i.e., $u \in \mathcal{F}$. Let f be a friend of u and F^u the set of u's friends, i.e., $f \in \mathsf{F}^u$. Clearly, $\mathsf{F}^u \subseteq \mathcal{F}$. For instance, currently Facebook has $k = 1.3 \times 10^9$ users (i.e., MAU) [3].

*This work consists a part of our study that has been submitted to IFIP SEC 2016 conf.

A u's profile consists of attributes a_i such as name, email, birthday and hometown. We denote the set of attributes of a u's profile as $\mathcal{T}$ and n as the size of $\mathcal{T}$, i.e., $\mathcal{T} = \{a_1, \ldots, a_n\}$. For instance, Facebook currently operates with a set of $n = 25$ profile attributes. Let F^{u*} be the union of u's friends and the u itself and f^* an element of F^{u*}, i.e., $f^* \in \mathsf{F}^{u*}$. Clearly, $\mathsf{F}^{u*} = \{u\} \cup \mathsf{F}^u$ and $\mathsf{F}^u \cap \{u\} = \emptyset$, as u is not a friend of u. For instance, $\mathsf{F}^{u*} = \{u, f_1, \ldots, f_{k'}\}$ describes a user u and its k' friends, where $1 \leq k' \leq k$.

Applications and Application providers. Let $\mathcal{L}$ be the set of apps an app provider (appP) can offer to every u_i in an OSN and s the size of this set, i.e., $\mathcal{L} = \{A_1, \ldots, A_s\}$. Let A_j, for $1 \leq j \leq s$, be the set of attributes that each A_j can collect, i.e., $\mathsf{A}_j \subseteq \mathcal{T}$. Each A_j is owned and managed by an appP denoted as P_j. The set of A_js that belong to P_j it is denoted as P_j, i.e., $\mathsf{P}_j \subseteq \mathcal{L}$. The set of all P_js is denoted as the $\mathcal{AP}$ and m the size of the set, i.e., $\mathcal{AP} = \{P_1, \ldots, P_m\}$. From our analysis we identified $s = 16.808$ apps and $m = 2055$ appPs on Facebook indicating that a P_j can have more than one A_j, i.e., $\mathsf{P}_j = \{A_1 \ldots A_{s\prime}\}$ with $1 \leq s\prime \leq 160$ [1].

1.1 Profiling

Application j**.** When A_j is activated by f^* (i.e., $f^* \in \mathsf{F}^{u*}$), a set of attributes a_i can be collected from u's profile. It is considered as $A_j^{u,\mathsf{F}^{u*}}$ an A_j that users in F^{u*} installed and as $\mathsf{A}_j^{u,\mathsf{F}^{u*}}$ the set of attributes a_i that $A_j^{u,\mathsf{F}^{u*}}$ can collect from u's profile. Clearly, $\mathsf{A}_j^{u,\mathsf{F}^{u*}} \subseteq \mathsf{A}_j \subseteq \mathcal{T}$. The set of all $A_j^{u,\mathsf{F}^{u*}}$s installed by the users in F^{u*} is denoted as $\mathsf{L}^{u,\mathsf{F}^{u*}}$. Clearly, $\mathsf{L}^{u,\mathsf{F}^{u*}} \subseteq \mathcal{L}$.

We denote as $\vec{a}_i$ a vector of length n which corresponds to a_i, i.e., $\vec{a}_i = [\overset{1}{0} \ldots 0\overset{i}{1}0 \ldots \overset{n}{0}]$. Moreover, we consider $\vec{A}_j^{u,\mathsf{F}^{u*}}$ as a vector of length n, which corresponds to $A_j^{u,\mathsf{F}^{u*}}$, i.e.,

$$\vec{A}_j^{u,\mathsf{F}^{u*}} = \bigvee_{a \in A_j^{u,\mathsf{F}^{u*}}} \vec{a} \Leftrightarrow \vec{A}_j^{u,\mathsf{F}^{u*}}[i] = \begin{cases} 1 & \text{if } a_i \in A_j^{u,\mathsf{F}^{u*}}, \\ 0 & \text{if } a_i \notin A_j^{u,\mathsf{F}^{u*}}, \end{cases} \quad (1)$$

for $1 \leq i \leq n$ and $1 \leq j \leq s$.

Remark:

- $x \vee y = \begin{cases} 0 & if\ x = y = 0, \\ 1 & otherwise, \end{cases}$ and $\vec{x} \vee \vec{y} = \vec{z}$ where $\vec{z}[i] = \vec{x}[i] \vee \vec{y}[i]$

For instance, an $\mathsf{A}_j^{u,\mathsf{F}^{u*}} = \{a_1, a_i, a_n\}$ is represented as $\vec{A}_j = \vec{a}_1 \vee \vec{a}_i \vee \vec{a}_n = [\overset{1}{1}0 \ldots 0\overset{i}{1}0 \ldots 0\overset{n}{1}]$. It represents the attributes that can be collected by A_j when is installed by f (i.e., the user's friend).

Application provider j**.** It is denoted as $\mathsf{AP}^{u,\mathsf{F}^{u*}}$ the set of appPs whose $A_j^{u,\mathsf{F}^{u*}}$ installed by users in F^{u*} which can collect attributes of

u's profile Hence,

$$\mathsf{AP}^{u,\mathsf{F}^{u*}} = \bigcup_{f^* \in \mathsf{F}^{u*}} \mathsf{AP}^{u,f^*} \tag{2}$$

- Note that: $\mathsf{AP}^{u,u} = \mathsf{AP}^{u}$.

Each appP consists of a set of $A_j^{u,\mathsf{F}^{u*}}$s denoted as $\mathsf{P}_j^{u,\mathsf{F}^{u*}}$ which the users in F^{u*} installed and have access to the u's profile. To identify which a_is can be collected by P_j we consider $\mathsf{P}_j^{u,\mathsf{F}^{u*}}$ as a set of size n, where $n \in \mathcal{T}$, i.e., Hence,

$$\mathsf{P}_j^{u,\mathsf{F}^{u*}} = \bigcup_{\substack{A \in \mathsf{P}_j^{u,f^*} \\ f^* \in \mathsf{F}^{u*}}} \mathsf{A}_j^{u,f^*} = \bigcup_{A \in \mathsf{P}_j^{u,\mathsf{F}^{u*}}} \mathsf{A}_j^{u,f^*} \tag{3}$$

- Remark: $\mathsf{P}_j^{u,\mathsf{F}^{u*}} = \bigcup_{f^* \in \mathsf{F}^{u*}} P_j^{u,f^*} = (\mathsf{P}_1^{u} \cup \mathsf{P}_1^{u,f_1} \cup \dots \cup \mathsf{P}_1^{u,f_i})$, where $\mathsf{F}^{u*} = \{u, f_1, \dots, f_i\}$
- Note that: $\mathsf{P}^{u,u} = \mathsf{P}^{u}$

Similarly, we consider $\vec{P}_j^{u,\mathsf{F}^{u*}}$ as a vector of length n (i.e., $n \in \mathcal{T}$), which corresponds to $P_j^{u,\mathsf{F}^{u*}}$, i.e.,

$$\vec{P}_j^{u,\mathsf{F}^{u*}} = \bigvee_{\substack{A \in \mathsf{P}_j^{u,f^*} \\ f^* \in \mathsf{F}^{u*}}} \vec{A}^{u,f^*} = \bigvee_{A \in \mathsf{P}_j^{u,\mathsf{F}^{u*}}} \vec{A}^{u,\mathsf{F}^{u*}} \tag{4}$$

- Remark: $\vec{P}_j^{u,\mathsf{F}^{u*}} = \bigvee_{f^* \in \mathsf{F}^{u*}} \vec{P}_j^{u,f^*} = (\vec{P}_j^{u} \vee \vec{P}_j^{u,f_1} \vee \dots \vee \vec{P}_j^{u,f_i})$, where $\mathsf{F}^{u*} = \{u, f_1, \dots, f_i\}$
- Note that: $\vec{P}^{u,u} = \vec{P}^{u}$

The complexity of this operation for all f^* in F^{u*} is $\mathcal{O}(n \times |\mathsf{P}_j^{u,\mathsf{F}^{u*}}|)$.

Example. Let $\mathsf{F}^{u*} = \{u, f_1, f_2\}$ the user u and f the u's friends. The set of A_js that all F^{u*} have activated is $\mathsf{L}^{u,\mathsf{F}^{u*}} = \{A_1^{u,\mathsf{F}^{u*}} \dots A_7^{u,\mathsf{F}^{u*}}\}$. The set of P_js for all A_j that all F^{u*} has activated is described as $\mathsf{AP}^{u,\mathsf{F}^{u*}} = (\mathsf{AP}^{u} \cup \mathsf{AP}^{u,f_1} \cup \mathsf{AP}^{u,f_2}) = \big(\{\mathsf{P}_1^{u}, \mathsf{P}_2^{u}, \mathsf{P}_3^{u}\} \cup \{\mathsf{P}_1^{u,f_1}, \mathsf{P}_2^{u,f_1}\} \cup \{\mathsf{P}_4^{u,f_2}\}\big) = \{(\mathsf{P}_1^{u} \cup \mathsf{P}_1^{u,f_1}), (\mathsf{P}_2^{u} \cup \mathsf{P}_2^{u,f_1}), \mathsf{P}_3^{u}, \mathsf{P}_4^{u,f_2}\}$. Each $\mathsf{P}_1^{u,\mathsf{F}^{u*}} = \mathsf{P}_1^{u} \cup \mathsf{P}_1^{u,f_1} = (\mathsf{A}_1^{u} \cup \mathsf{A}_2^{u}) \cup (\mathsf{A}_1^{u,f_1} \cup \mathsf{A}_2^{u,f_1} \cup \mathsf{A}_3^{u,f_1})$, $\mathsf{P}_2^{u,\mathsf{F}^{u*}} = \mathsf{P}_2^{u} \cup \mathsf{P}_2^{u,f_1} = \mathsf{A}_4^{u} \cup \mathsf{A}_5^{u,f_1}$,$\mathsf{P}_3^{u,\mathsf{F}^{u*}} = \mathsf{A}_6^{u}$ and $\mathsf{P}_4^{u,\mathsf{F}^{u*}} = \mathsf{A}_7^{u,f_2}$. Each A_j activated by u or u's friends can collect a set of attributes a_i from u's profile such that, $\mathsf{A}_1 = \{a_1, a_2, a_3, a_4\}$, $\mathsf{A}_2 = \{a_1, a_4, a_5\}$, $\mathsf{A}_3 = \{a_4, a_6, a_7\}$, $\mathsf{A}_4 = \{a_8\}$, $\mathsf{A}_5 = \{a_9\}$, $\mathsf{A}_6 = \{a_{10}, a_{11}\}$, $\mathsf{A}_7 = \{a_{12}\}$. The total collection of a_is for $\mathsf{P}_1^{u,\mathsf{F}^{u*}} = (\{a_1, a_2, a_3\} \cup \{a_1, a_4\}) \cup (\{a_1, a_4\} \cup \{a_5\} \cup \{a_4, a_6, a_7\}) = \{a_1 \dots a_7\}$, $\mathsf{P}_2^{u,\mathsf{F}^{u*}} = \{a_8\} \cup \{a_9\} = \{a_8, a_9\}$, $\mathsf{P}_3^{u,\mathsf{F}^{u*}} = \{a_{10}, a_{11}\}$, $\mathsf{P}_4^{u,\mathsf{F}^{u*}} = \{a_{12}\}$.

1.2 Degree of *collateral* information collection

Friends f of u ($f \in F^u$) allow access to u's profile by installing A_js. We denote with $\Pi^u_{A_j^u, A_j^{u,\mathsf{F}^u}}$ the amount of attributes that can be collected by A_j exclusively from u's friends (and not trough the user herself, i.e., $u \notin \mathsf{F}^u$) . Let $\vec{\Pi}^u_{A_j^u, A_j^{u,\mathsf{F}^u}}$ be a vector of length n which $\Pi^u_{A_j^u, A_j^{u,\mathsf{F}^u}}$ provides, where $n = |\mathcal{T}|$, where

$$\vec{\Pi}^u_{A_j^u, A_j^{u,\mathsf{F}^u}} = \vec{A}'^u_j \bigwedge \vec{A}^{u,\mathsf{F}^u}_j \tag{5}$$

- Remark: $\vec{x}' \wedge \vec{x} = [\overset{1}{0} \dots \overset{n}{0}]$ and $\vec{x}' \vee \vec{x} = [\overset{1}{1} \dots \overset{n}{1}]$.

The complexity of this operation for all f^* in F^{u*} is $\mathcal{O}(n^4 \times |\mathsf{A}^u_j| \times |\mathsf{A}^{u,\mathsf{F}^u}_j|)$.

Similarly, to compute the amount of attributes can be collected by P_j exclusively from u's friends in F^u we denote as $\vec{\Pi}^u_{P_j^u, P_j^{u,\mathsf{F}^u}}$, i.e.,

$$\vec{\Pi}^u_{P_j^u, P_j^{u,\mathsf{F}^u}} = \vec{P}'^u_j \bigwedge \vec{P}^{u,\mathsf{F}^u}_j . \tag{6}$$

An overall notation description is given in Table 1.

Table 1: Notations

Notation	**Description**
$\mathcal{F} = \{u_1, \dots, u_k\}$	Set of k users in an OSN and u the user under consideration.
$\mathsf{F}^{u*} = \{u, f_1, \dots, f_{k'}\}$	Set of u's friends (i.e., f) and u herself.
$\mathcal{T} = \{a_1, \dots, a_n\}$	Set of n attributes of u's profile.
$\mathcal{L} = \{A_1, \dots, A_s\}$	Set of s apps (i.e., A_js).
$\mathcal{AP} = (\{P_1, \dots, P_m\}$	Set of m appPs (i.e., P_js).
L^{u,f^*} / $\mathsf{L}^{u,\mathsf{F}^{u*}}$	Set of A_js installed by a user f^* / all users in F^{u*}, having access to u's profile.
AP^{u,f^*} /$\mathsf{AP}^{u,\mathsf{F}^{u*}}$	Set of P_js whose A_js installed by a user f^* / all users in F^{u*}, having access to u's profile.
A^{u,f^*}_j / $\mathsf{A}^{u,\mathsf{F}^{u*}}_j$	Set of a_is each A_j, installed by a user f^* / all users in F^{u*}, having access to u's profile.
P^{u,f^*}_j / $\mathsf{P}^{u,\mathsf{F}^{u*}}_j$	Set of a_is all A_js installed by a user f^* / all users in F^{u*} and belong to P_j, having access to u's profile.

1.3 The case of Facebook Applications

To examine the problem, we extended our analysis for the apps (i.e., A_js) and appPs (i.e., P_js) on Facebook, using the Appinspect dataset [2, 1]. For each A_j, apart from the application name and id, the dataset provide us with the requested permissions and the A_js each P_j owns. We computed the proportion of attributes an A_j and P_j can collect through: 1) the user's friends and the user herself (i.e., *profiling*, F^{u*}) and 2) only the user's friends (i.e., *degree of collateral information collection*, F^u). From 16.808 apps, 1202 enables *collateral information* *collection*. Our analysis focuses on A_js and P_js that have more than 10.000 MAU; there are 207 and 88 respectively. [1]

Profiling, F^{u*}. Performing the analysis over the dataset, we found that 72.4% of A_js and 62.5% of P_js can collect one attribute from F^{u*}. For all A_js and all P_js, 48.6% and 28.7% of attributes which are considered sensitive by the participants of our survey (such as *photos*, *videos*, *location* and *family-relationships*) can be collected. Considering location related attributes such as *current location*, *hometown*, *work_history* and *education_history*, the proportion of attributes that can be collected are 23.5% from A_js and 23.2% from P_js.

Degree of collateral information collection, F^{u}. For A_js installed only by F^{u}, 28.9% of them show a degree of *collateral information collection* equal to 1; similarly, 36.3% of all P_js. Moreover for F^u, we identified that the proportion of sensitive attributes that can be collected from A_js and P_js is 46.8% and 37%, respectively; while the proportion of collectable location related attributes is 22.5% for A_js and 36.9% for P_js.

We conclude that the size of the two sets of sensitive attributes, collected via profiling versus exclusively through friends, are both significant and, surprisingly, comparable to each other. We also found that a considerable amount of attributes concerning the user's location can be collected by either A_js or P_js.

2 Conclusion

In this paper we have presented a study concerning the *collateral damage* caused by friends' apps in social networking sites. Based on real data, we have quantified the significance of *collateral information collection* by computing the proportion of attributes collected by apps installed by the users' friends. We have found that a significant proportion of sensitive attributes, such as photos, videos, relationships and location, can be collected from A_js either by the user's friends and the user herself (i.e., 48.6%) or exclusively from the user's friends (i.e., 46.8%); surprisingly, these values are comparably high. Furthermore, a considerable amount of location-related attributes are collected by both friends' apps and profiling appPs. Furthermore, to the best of our knowledge, we have been first to report the potential user *profiling* threat that could be achieved by application providers: they can gain access to complementary subsets of user profile attributes by offering multiple apps.

References

[1] AppInspect. A framework for automated security and privacy analysis of osn application ecosystems. http://ai.sba-research.org/.

[1] `http://iraklissymeonidis.info/Fb\_apps\_statistics/`

[2] M. Huber, M. Mulazzani, S. Schrittwieser, and E. R. Weippl. Appinspect: large-scale evaluation of social networking apps. In *Conference on Online Social Networks, COSN'13, Boston, MA, USA, October 7-8, 2013*, pages 143–154, 2013.

[3] Statista. Leading social networks worldwide as of august 2015.