

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

Iraklis Symeonidis<sup>1</sup>, Faruk Gologlu<sup>2</sup>, and Bart Preneel<sup>1</sup>

<sup>1</sup>ESAT COSIC, KU Leuven and iMinds, Belgium.  
firstname.lastname@esat.kuleuven.be

<sup>2</sup>Claude Shannon Institute, University College Dublin.

January 18, 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.,  $\mathcal{F} = \{u_1, \dots, u_k\}$ . 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.,  $f \in F^u$ . Clearly,  $F^u \subseteq \mathcal{F}$ . 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.,  $\mathcal{T} = \{a_1, \dots, a_n\}$ . 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.,  $f^* \in F^{u*}$ . Clearly,  $F^{u*} = \{u\} \cup F^u$  and  $F^u \cap \{u\} = \emptyset$ , as  $u$  is not a friend of  $u$ . For instance,  $F^{u*} = \{u, f_1, \dots, f_{k'}\}$  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.,  $\mathcal{L} = \{A_1, \dots, A_s\}$ . Let  $A_j$ , for  $1 \leq j \leq s$ , be the set of attributes that each  $A_j$  can collect, i.e.,  $A_j \subseteq \mathcal{T}$ . Each  $A_j$  is owned and managed by an appP denoted as  $P_j$ . The set of  $A_j$ s that belong to  $P_j$  it is denoted as  $\mathcal{P}_j$ , i.e.,  $\mathcal{P}_j \subseteq \mathcal{L}$ . The set of all  $P_j$ s is denoted as the  $\mathcal{AP}$  and  $m$  the size of the set, i.e.,  $\mathcal{AP} = \{P_1, \dots, P_m\}$ . 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.,  $\mathcal{P}_j = \{A_1 \dots A_{s'}\}$  with  $1 \leq s' \leq 160$  [1].

## 1.1 Profiling

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

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

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

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

Remark:

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

For instance, an  $A_j^{u, F^{u*}} = \{a_1, a_i, a_n\}$  is represented as  $\vec{A}_j = \vec{a}_1 \vee \vec{a}_i \vee \vec{a}_n = [1 \ 0 \dots 0 \overset{i}{1} 0 \dots 0 \ 1]$ . 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  $AP^{u, F^{u*}}$  the set of appPs whose  $A_j^{u, F^{u*}}$  installed by users in  $F^{u*}$  which can collect attributes of

$u$ 's profile Hence,

$$\text{AP}^{u, \mathcal{F}^{u*}} = \bigcup_{f^* \in \mathcal{F}^{u*}} \text{AP}^{u, f^*} \quad (2)$$

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

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

$$\mathcal{P}_j^{u, \mathcal{F}^{u*}} = \bigcup_{\substack{A \in \mathcal{P}_j^{u, f^*} \\ f^* \in \mathcal{F}^{u*}}} A_j^{u, f^*} = \bigcup_{A \in \mathcal{P}_j^{u, \mathcal{F}^{u*}}} A_j^{u, f^*} \quad (3)$$

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

- Note that:  $\mathcal{P}^{u, u} = \mathcal{P}^u$

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

$$\vec{\mathcal{P}}_j^{u, \mathcal{F}^{u*}} = \bigvee_{\substack{A \in \mathcal{P}_j^{u, f^*} \\ f^* \in \mathcal{F}^{u*}}} \vec{A}^{u, f^*} = \bigvee_{A \in \mathcal{P}_j^{u, \mathcal{F}^{u*}}} \vec{A}^{u, \mathcal{F}^{u*}} \quad (4)$$

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

- Note that:  $\vec{\mathcal{P}}^{u, u} = \vec{\mathcal{P}}^u$

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

**Example.** Let  $\mathcal{F}^{u*} = \{u, f_1, f_2\}$  the user  $u$  and  $f$  the  $u$ 's friends. The set of  $A_j$ s that all  $\mathcal{F}^{u*}$  have activated is  $\mathcal{L}^{u, \mathcal{F}^{u*}} = \{A_1^{u, \mathcal{F}^{u*}} \dots A_7^{u, \mathcal{F}^{u*}}\}$ . The set of  $\mathcal{P}_j$ s for all  $A_j$  that all  $\mathcal{F}^{u*}$  has activated is described as  $\text{AP}^{u, \mathcal{F}^{u*}} = (\text{AP}^u \cup \text{AP}^{u, f_1} \cup \text{AP}^{u, f_2}) = (\{\mathcal{P}_1^u, \mathcal{P}_2^u, \mathcal{P}_3^u\} \cup \{\mathcal{P}_1^{u, f_1}, \mathcal{P}_2^{u, f_1}\} \cup \{\mathcal{P}_4^{u, f_2}\}) = \{(\mathcal{P}_1^u \cup \mathcal{P}_1^{u, f_1}), (\mathcal{P}_2^u \cup \mathcal{P}_2^{u, f_1}), \mathcal{P}_3^u, \mathcal{P}_4^{u, f_2}\}$ . Each  $\mathcal{P}_1^{u, \mathcal{F}^{u*}} = \mathcal{P}_1^u \cup \mathcal{P}_1^{u, f_1} = (A_1^u \cup A_2^u) \cup (A_1^{u, f_1} \cup A_2^{u, f_1} \cup A_3^{u, f_1})$ ,  $\mathcal{P}_2^{u, \mathcal{F}^{u*}} = \mathcal{P}_2^u \cup \mathcal{P}_2^{u, f_1} = A_4^u \cup A_5^{u, f_1}$ ,  $\mathcal{P}_3^{u, \mathcal{F}^{u*}} = A_6^u$  and  $\mathcal{P}_4^{u, \mathcal{F}^{u*}} = A_7^{u, f_2}$ . Each  $A_j$  activated by  $u$  or  $u$ 's friends can collect a set of attributes  $a_i$  from  $u$ 's profile such that,  $A_1 = \{a_1, a_2, a_3, a_4\}$ ,  $A_2 = \{a_1, a_4, a_5\}$ ,  $A_3 = \{a_4, a_6, a_7\}$ ,  $A_4 = \{a_8\}$ ,  $A_5 = \{a_9\}$ ,  $A_6 = \{a_{10}, a_{11}\}$ ,  $A_7 = \{a_{12}\}$ . The total collection of  $a_i$ s for  $\mathcal{P}_1^{u, \mathcal{F}^{u*}} = (\{a_1, a_2, a_3\} \cup \{a_1, a_4\}) \cup (\{a_1, a_4\} \cup \{a_5\} \cup \{a_4, a_6, a_7\}) = \{a_1 \dots a_7\}$ ,  $\mathcal{P}_2^{u, \mathcal{F}^{u*}} = \{a_8\} \cup \{a_9\} = \{a_8, a_9\}$ ,  $\mathcal{P}_3^{u, \mathcal{F}^{u*}} = \{a_{10}, a_{11}\}$ ,  $\mathcal{P}_4^{u, \mathcal{F}^{u*}} = \{a_{12}\}$ .

## 1.2 Degree of *collateral* information collection

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

$$\vec{\Pi}_{A_j^u, A_j^{u, F^u}}^u = \vec{A}_j^{u'} \wedge \vec{A}_j^{u, F^u} \quad (5)$$

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

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

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

$$\vec{\Pi}_{P_j^u, P_j^{u, F^u}}^u = \vec{P}_j^{u'} \wedge \vec{P}_j^{u, F^u}. \quad (6)$$

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.
$F^{u^*} = \{u, f_1, \dots, f_{k'}\}$	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., $A_j$ s).
$\mathcal{AP} = (\{P_1, \dots, P_m\})$	Set of $m$ appPs (i.e., $P_j$ s).
$\mathcal{L}^{u, f^*} / \mathcal{L}^{u, F^{u^*}}$	Set of $A_j$ s installed by a user $f^*$ / all users in $F^{u^*}$ , having access to $u$ 's profile.
$\mathcal{AP}^{u, f^*} / \mathcal{AP}^{u, F^{u^*}}$	Set of $P_j$ s whose $A_j$ s installed by a user $f^*$ / all users in $F^{u^*}$ , having access to $u$ 's profile.
$A_j^{u, f^*} / A_j^{u, F^{u^*}}$	Set of $a_i$ s each $A_j$ , installed by a user $f^*$ / all users in $F^{u^*}$ , having access to $u$ 's profile.
$P_j^{u, f^*} / P_j^{u, F^{u^*}}$	Set of $a_i$ s all $A_j$ s 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.,  $A_j$ s) and appPs (i.e.,  $P_j$ s) 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  $A_j$ s 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  $A_j$ s and  $P_j$ s that have more than 10.000 MAU; there are 207 and 88 respectively.<sup>1</sup>

**Profiling,  $F^{u*}$ .** Performing the analysis over the dataset, we found that 72.4% of  $A_j$ s and 62.5% of  $P_j$ s can collect one attribute from  $F^{u*}$ . For all  $A_j$ s and all  $P_j$ s, 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  $A_j$ s and 23.2% from  $P_j$ s.

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

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  $A_j$ s or  $P_j$ s.

## 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  $A_j$ s 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. 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/>.

---

<sup>1</sup>[http://iraklissymeonidis.info/Fb\\\_apps\\\_statistics/](http://iraklissymeonidis.info/Fb\_apps\_statistics/)

- [2] M. Huber, M. Mulazzani, S. Schrittwieser, and E. R. Weippl. Apinspect: 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.