Collateral Damage by Facebook Applications: a Comprehensive Study (under review)

Iraklis Symeonidis[1](), Gergely Biczók[2], Fatemeh Shirazi[1], Cristina Pérez-Solà[3], Jessica Schroers[3], Bart Preneel[1]

[1] COSIC, KU Leuven, Belgium `first.last@esat.kuleuven.be`
[2] Budapest Univ. of Technology and Economics, Dept. of Networked Systems and Services, CrySyS Lab `biczok@crysys.hu`
[3] Universitat Autònoma de Barcelona, dEIC `cperez@deic.uab.cat`
[4] KU Leuven, Centre for IT & IP Law `jessica.schroers@kuleuven.be`

Abstract. Third-party applications on Facebook can collect personal data of the users who install them, but also of their friends. This raises serious privacy issues as these friends are not notified by the applications nor by Facebook and they have not given consent. This paper presents a detailed multi-faceted study on the *collateral information collection* of the applications on Facebook. To investigate the views of the users, we designed a questionnaire and collected the responses of 114 participants. The results show that participants are concerned about the collateral information collection and in particular about the lack of notification and of mechanisms to control the data collection. Based on real data, we compute the likelihood of collateral information collection affecting users: we show that the probability is *significant* and greater than 80% for popular applications such as TripAdvisor. We also demonstrate that a substantial amount of profile data can be collected by applications, which enables application providers to *profile* users. To investigate whether collateral information collection is an issue to users' privacy we analysed the legal framework in light of the new General Data Protection Regulation. We provide a detailed analysis of the entities involved and investigate which entity is accountable for the collateral information collection. To provide countermeasures, we propose a privacy dashboard extension that implements privacy scoring computations to enhance transparency towards collateral information collection. Furthermore, we discuss alternative solutions highlighting other countermeasures such as notification and access control mechanisms, cryptographic solutions and application auditing. To the best of our knowledge this is the first work that provides a detailed multi-faceted study of this problem and that analyses the threat of user *profiling* by application providers.

1 Introduction

Online Social Networks (OSNs) have become a dominant platform for people to express themselves, share information and interact with each other. By design and popularity, Facebook has morphed into an immense information repository [1], storing users' personal data and logging their interactions with friends, groups, events and pages. The sheer amount and potentially sensitive nature of such data have raised a plethora of privacy issues for the Facebook users including the lack of awareness from users [2], the cumbersome privacy controls [3], the accidental information disclosure [4] and the reconstruction of the identity of users [5]. This paper focuses on the interdependent information collection [6, 7]: apps installed by users collect data about users but also about their friends, which clearly brings major privacy implications.

Due to Facebook's popularity, third-party apps on Facebook use the Developers Platform [8] to launch their applications (apps), benefiting from the existing massive user base [1]. Currently, Facebook offers a set of more than 25,000 apps [9], such as Criminal Case [10], Candy Crush Saga [11], Angry Birds [12] and TripAdvisor [13]. When a user installs an application (app) from

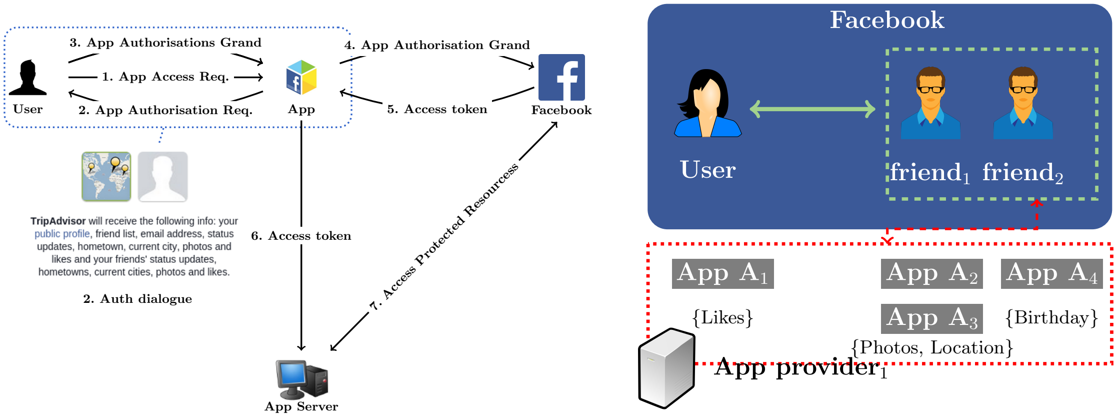

3. App Authorisations Grand
1. App Access Req.
2. App Authorisation Req.
User
App
4. App Authorisation Grand
5. Access token
Facebook
TripAdvisor will receive the following info: your public profile, friend list, email address, status updates, hometown, current city, photos and likes and your friends' status updates, hometowns, current cities, photos and likes.
2. Auth dialogue
6. Access token
7. Access Protected Resourcess
App Server
Facebook
User
friend₁ friend₂
App A₁
{Likes}
App A₂
App A₃
{Photos, Location}
App A₄
{Birthday}
App provider₁

Fig. 1: (a) Facebook Developers app architecture, (b) *Collateral information collection* scheme on Facebook

the Facebook app store, the app may collect their information on Facebook. For instance, Candy Crush Saga collects the name, profile picture, country, friend list and email address of a user. Other apps may collect other types of information.

When a friend of a user installs an app on Facebook, this app not only can collect the friend's information but, it may also collect information about the user herself. For instance, if a friend of a user installs travelling apps such as *TripAdvisor*, these may collect the user's current location, hometown, work location and likes. This allows for sharing travel preferences [14] aiming to notify both the user and their friends whether they have visited the same point of interest. If a friend of a user installs a horoscope app such as *Horoscopes*, the app may collect the friend's and the user's birthday.

App installation and information flow From a technical standpoint, Facebook relies on permission-based platform security and applies the least privilege principle to third-party apps [8]. For installation and operation, each app requests from the user a set of *permissions*, granting the app the right to access and collect information such as the profile name (steps 1 to 3 in Fig. 1 (a)). After the user's and Facebook approval, apps can collect the profile information of the user (i.e., personal data) and store it on servers outside Facebook's ecosystem and thus out of the user's control (steps 7 in Fig. 1 (a)) [15, 16].

Facebook API and friend permissions Initially, the API v1.x of Facebook (April 2010 – April 2015) provided a set of permissions to the apps, such as $friends\_birthday$ and $friends\_location$. These permissions gave the apps the right to access and collect personal data of users via their friends, such as their birthdays and locations. Currently, the updated API version v2.x [8] (April 2014 – present) replaced the $friends\_xxx$ permissions with the single $user\_friends$ permission. In our previous work, we analysed the extent of friend permission requests among popular apps [7] using the publicly available AppInspect dataset [9] provided by Huber et al. [17]. The proportion of such permission requests were more than 10%. Although the v2.x API has had to be in line with the regulations of the EU [18] and FTC [19], it still discloses up to fourteen user profile attributes that can be collected via the apps installed by friends of a user.

Privacy interdependence and collateral information collection The sub-optimal privacy controls and the server-to-server (and potential offline) communication between Facebook and app providers make any privacy protection mechanism hard to apply [20]. As a result, the user's profile attributes can be arbitrarily retrieved by an app provider without automatic notification or on-demand approval by the user through their friends.

Table 1: The mutually amplifying effect of collateral information collectionȧnd multi-app data fusion

	App provider		
	App$_1$	App$_2$	
u	$\{a_1, a_2\}$	$\{a_3, a_4\}$	Direct information collection
f^u	$\{a_5\}$	$\{a_6\}$	Collateral information collection

Theorem 1. *We define collateral information collection as the acquisition of users' personal data through any mean or tool initiated by anyone but the users themselves.*

Such type of information collection may inflict a privacy loss due to the lack of transparency and consent given by the users. Fundamentally speaking, collateral information collection is the manifestation of *interdependent privacy*; the scenario when the privacy of an individual user is affected by the decisions of other users [6]. From an economic point of view, sharing a user's information without their direct consent can lead to the emergence of externalities, i.e., unwanted side-effects. While sharing someone else's information may yield benefits for them (positive externalities, such as personalised experience in third-party apps), it is also almost certain to cause a decrease in their utility (negative externality, e.g., exposed profile items). Existing research is limited to pointing out the existence of and risks stemming from such negative externalities on the Facebook app ecosystem [6], and its potential impact on app adoption [21, 22].

App providers, data fusion and profiling Third party app providers can be owners of several apps in Fig. 1 (b) app provider$_1$ offers the apps A_1, A_2, A_3 and A_4. To the best of our knowledge, this has not yet been studied in the literature. For instance, the app providers Vipo Komunikacijos and Telaxo offer 163 and 130 apps; among those, 99 and 118 have more than 10 000 monthly active users, respectively (extracted from the AppInspect dataset [9]). As a consequence, an app provider may cluster several apps and thus may collect more personal data from the user's profile. Moreover, every app retrieves the Facebook user's ID, that uniquely identifies a user over apps. Hence, the app provider could utilise a type of data fusion over all apps offered [23], and construct a more complete representation of the profiles of the users. Such data fusion partly constitutes and greatly enables *profiling* as defined in Article 4 GDPR [18]. Note that this definition is analogous to its common meaning in marketing in the context of consumer behaviour [24].

Table 1 illustrates the mutually amplifying interaction between collateral information collection and multi-app data fusion. It is clear that collateral information collection allows an increasing amount of data collection through vertical coupling for example, by adding $\{a_5\}$ to the directly collected $\{a_1,\ a_2\}$. On the other hand, data fusion over multiple apps alone allows for horizontal coupling such that combining $\{a_1, a_2\}$ and $\{a_3,\ a_4\}$ into a consistent $\{a_1,\ a_2,\ a_3,\ a_4\}$. With both mechanisms in full effect, the app provider is able to compile an extended attribute set of $\{a_1,\ a_2,\ a_3,\ a_4,\ a_5,\ a_6\}$. Therefore, with the help of multiple apps installed by the friends of a user, an app provider could profile a user partly or entirely without her consent, which constitutes a privacy breach, which has legal consequences [18].

Contribution In this work, we investigate the collateral information collection through Facebook applications (apps) installed by one or more friends of a user, taking into account the fact that an app provider may own multiple apps. Specifically, we identified five research questions to advance our understanding of indirect and collateral information collection in the case of Facebook apps.

- *Are users concerned that apps installed by their friends are able to collect their profile data on Facebook?* To identify whether users were concerned, we designed a questionnaire and distributed it among 114 participants. We aimed at identifying their concerns on *collateral information collection, lack of transparency (notification) and not being asked for their approval*

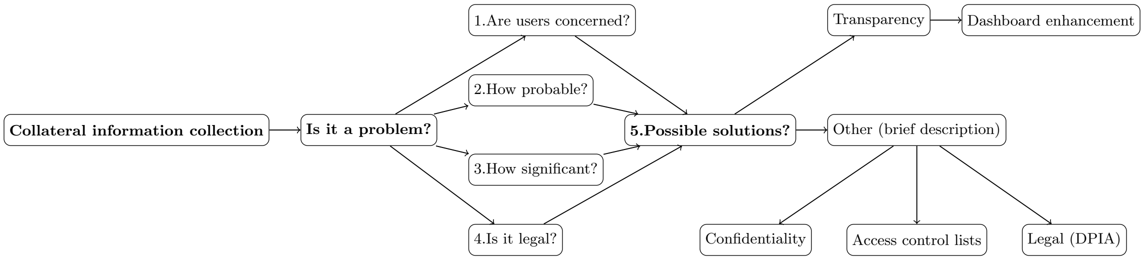


Fig. 2: Paper contribution to collateral information collection on Facebook.

(consent). Our user opinion study serves as evidence that the majority of participants are indeed concerned. On top of that, their concern is bidirectional: they would like to both notify their friends and be notified by their friends when installing apps enabling collateral information collection. They would also like to be able to restrict which attributes are shared that way.

- *What is the likelihood that an installed app enables collateral information collection?* To answer this question, we estimated the probability of that event to occur in the Facebook ecosystem. We show that the likelihood of collateral information collection for a user depends on the number of friends a user has and the popularity of apps (number of active users). We run simulations with a network size of 1 million users, and investigated various network topologies and app adoption models tailored to the Facebook ecosystem. Based on the results obtained, we demonstrate that for an average user (≈ 200 friends) this likelihood is greater than 80% for popular apps such as TripAdvisor.
- *How significant is the collateral information collection?* To answer how much information is collected, we quantified the amount of attributes collected by apps which enable collateral information collection. The quantification depends on popular apps available on the Facebook ecosystem, and we estimated the number of attributes that each app is collecting. We also considered that several apps belong to the same app provider. To investigate that, we developed a mathematical model that also takes into account the access control mechanisms that Facebook provides to its users. For our calculations, we used the Appinspect dataset [9] which is a real world snapshot of the apps on Facebook.
- *Under the data protection legislation, is collateral information collection considered a risk for the protection of the personal data of Facebook users?* We investigated this research question under the prism of the new General Data Protection Regulation (GDPR) [18]. First, our analysis clarifies who the data controllers and data processors are. Second, it scrutinises whether collateral information collection is an adequate practice from a data protection point of view. Finally, it identifies who is merely accountable.
- *How can we mitigate collateral information collection?* For this end, we analysed various countermeasures as follows. First, we propose the use of transparency tools aiming at raising user awareness, and helping the users make informed decisions. For this purpose, we outline a privacy dashboard extension that implements privacy scoring computations for enhancing transparency towards collateral information collection. Furthermore, we discuss alternative solutions highlighting other countermeasures such as notification and access control solutions, cryptographic tools and a legal application assessment framework.

The rest of the paper is organised as follows (see Fig. 2). Section 2 characterises collateral information collection and presents our user opinion study. Section 3 constructs a mathematical model of collateral information collection and estimates the likelihood of a user being affected. Section 4 extends the model and quantifies the significance of collateral information collection illustrated by a case study of popular apps. Section 5 conducts a legal analysis focusing on the

GDPR. Section 6 outlines potential solutions to mitigate collateral information collection including the the high-level design for a customised privacy dashboard. Section 7 summarises the related work. Finally, Section 8 describes future work and concludes the paper.

2 Relevance of collateral information collection on Facebook: evidence of user concerns

In this section we investigate the research question: *are users concerned that apps installed by their friends are able to collect their profile data on Facebook?* To answer this question, we designed a questionnaire and distribute it to 114 participants. We first characterise collateral information collection with five shortcomings, which we refer to as the five pillars of collateral information collection. These five pillars establish the relevance of collateral information collection as a key challenge on Facebook privacy. Thereafter, we present the users' concerns which serve as evidence of the relevance of collateral information collection. Our questionnaire covers concerns with regard to the *collateral information collection, lack of transparency (notification) and not being asked for their approval (consent).*

2.1 The five pillars of collateral information collection on Facebook

Lack of transparency The primary problem of collateral information collection lies in the fact that users are not aware of their data being collected [7, 25]. Awareness has been reported as a key issue for Facebook users [6, 7, 22, 26, 27]. If a friend of a user installs an application, it is unclear whether the friend is aware of the data collection by this application. This lack of transparency makes it impossible for the user and the friends to give informed consent or to influence the decision of use of the personal data of users in any way.

Lack of control by Facebook users The lack of control is one of the fundamental privacy problems in OSNs [2, 4, 7, 27, 28]. However, in the case of collateral information collection the lack of transparency and the inability to prevent such information collection without un-friending the Facebook friend, who has installed the app, limits the user even further in having control. For some users, the price for protecting their privacy might become too high.

Informed consent and its (in)transitivity Lack of informed consent is often an integral part of privacy loss on the web and in particular in OSNs [29]. Due to lack of transparency, informed consent cannot be given by a user when one of their friends installs an app that can collect information about the user themselves. One can argue that the friends that installed the app might be informed about such data collection, hence they can make a decision for the user. In this case, even assuming that the friend is aware and privacy savvy, the question arises whether consent is transitive? Transitivity of consent and preferences has been the subject of debate in the literature [30]. One can assume that the user is giving indirect consent to their friends to share their data by making their personal profile attributes accessible to them. This stance can be criticised on (at least) two grounds. First, the user does not know whether their friend installed an app collecting such data. Hence, their assumed indirect consent cannot be informed. Second, the default settings on Facebook are clearly pro-sharing when it comes to the collateral information collection; also, Facebook users often leave the default settings unchanged [3, 31].

Facebook friends are not (always) real friends Even if we assume that privacy decision-making can be transitive in some cases among friends, there have been reports in the literature that the relationship between Facebook friends cannot be equated to real-life friendships [32]. Therefore, it is debatable whether the user trusts their Facebook friends to make the right (privacy) decision on their behalf.

False sense of privacy and over-sharing Due to the lack of transparency, users may assume that their data are only visible to their friends. In such cases, users might be less restrictive with information sharing [33]. Hence, in an environment where apps installed by the user's friends can collect information from the user themselves, the average user may experience a false sense of privacy. Such a gap between a Facebook user's general privacy expectation and reality has been well-investigated by Liu et al. [3] and Madejsk et al. [34]. The added potential collateral information collection stemming from interdependent privacy can further widen this gap.

2.2 User opinion study

In order to investigate the research question: "are users concerned about collateral information collection?", we conducted an online questionnaire. We explored the users' concerns about the disclosure of personal data by Facebook apps installed by the friends of a user, and we investigated the users' concerns about un-consented information collection on Facebook. We observed that the participants' opinion about collateral information collection, can be characterised as remarkably concerned in general and in particular when the information collection is un-consented. Furthermore, the majority of users prefer to take action to prevent the collateral information collection.

Methodology After an introduction, our questionnaire consisted of four main parts (see A). First, we asked users about their concerns on *collateral information collection*. We assessed users' standpoints and concerns about default privacy settings and the lack of notification for indirect and un-consented information collection. This assessment is necessary to be able to differentiate users who are concerned independent of their intentions to take actions against such practices. The second part explores which attributes users were more concerned about. We investigated the type of personal data on Facebook users find most sensitive. The third part is twofold: 1) whether users want to be notified when their friends' apps can collect their personal data or when their installed apps can collect personal data from their friends; 2) which actions users prefer to take in such cases. Users replied the questions by marking their responses on a scale of 1 to 5 where 1 stands for "not concerned at" all and 5 stands for "extremely concerned" [35, 36]; we also provided a text field where necessary. The fourth part of the questionnaire collects demographics and information regarding the participants' and the use of Facebook apps.

Participants Our response pool consisted of 114 participants. Participants were recruited from the authors' direct and extended friend circles (including mostly, but not only, Facebook friends). A large proportion of participants are aged between 20 and 35 and are well educated.

Questionnaire Design Limitations Our response pool is a convenience sample of the Facebook users. As such a sample is usually limited in size and diversity, we do not extrapolate our findings to the general population of Facebook and do not lean on quantitative results from the user opinion study. While a full-scale quantitative survey might constitute important future work (see Sect. 8), we would like to point out that notable previous research works on Facebook users have been limited to student populations. Acquisti and Grossklags [2] used a sample of 294 were more than 89% were undergraduate and graduate college students; whereas the 89 respondents of Boyd and Hargittai [31] were almost exclusively students aged between 18 and 19.

2.3 Results

For the first part, we observe that for all four statements users show concern (see Fig. 3). For instance, 66% (i.e., $37 + 38/114 \approx 0.66$) of users are at least very concerned about the default privacy setting of Facebook that allows the collateral information collection. Similarly, 77% (i.e., $38 + 49/114 \approx 0.77$) of users are at least very concerned about not being notified when their

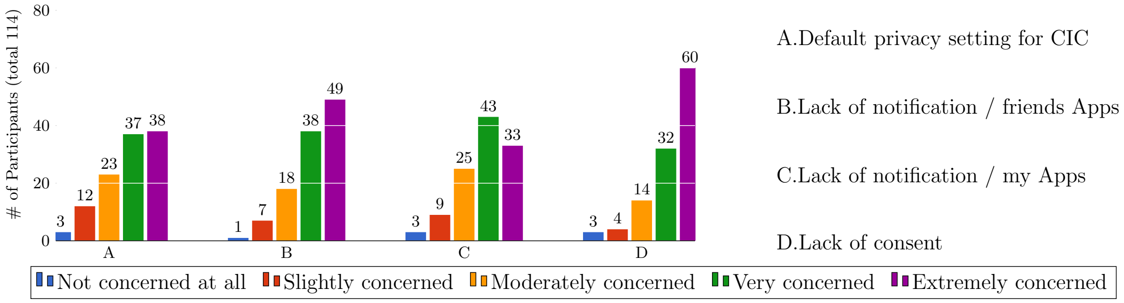


Fig. 3: Results for the first part of the questionnaire where we asked participants about their opinions on four statements regarding default settings, lack of notification (for friends and for the user themselves), and lack of consent for the collateral information collection (CIC).

friends enable collateral information collection and 67% (i.e., $43 + 33/114 \approx 0.67$) for not being notified when one of the user's own apps can collect their friends' information. Finally, 81% (i.e., $32+60/114 \approx 0.81$) of users are at least very concerned about the collateral information collection and the lack of their approval. Note that Golbeck and Mauriello [37] have investigated how informed users are regarding the privacy risks of using Facebook apps. Their findings show that users do not always comprehend what type of data is collected by apps even when they have installed the app themselves. Therefore, it is reasonable to assume an incomplete understanding of apps installed by their friends, which is in line with our results.

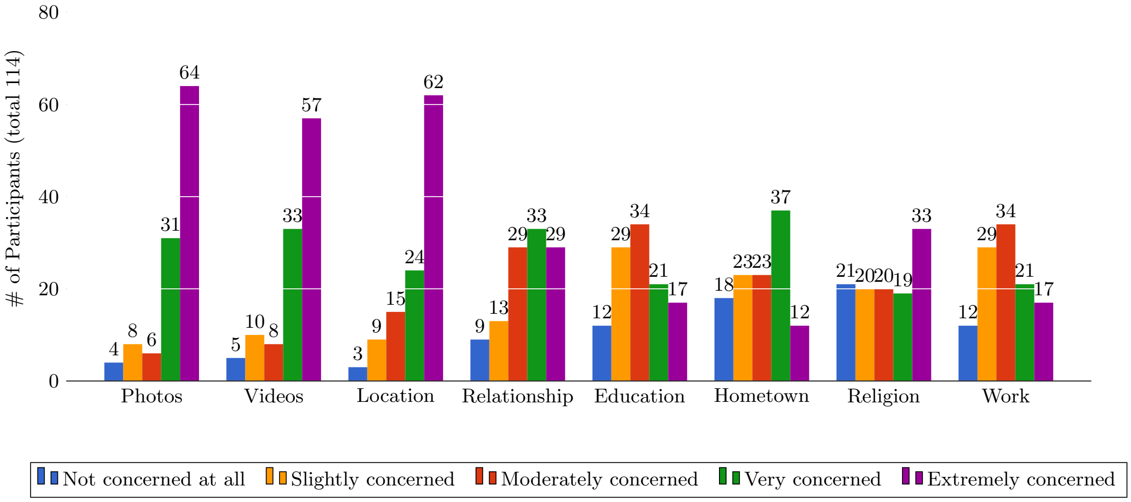


Fig. 4: Total number of apps and appPs requesting collateral information collection of sensitive attributes (per attribute)

For the second part of our questionnaire (see Fig. 4), we found that although users are concerned about a number of attributes, their concern is relatively subjective and differs between users. However, it is noteworthy that certain attributes clearly stand out and have been marked as more concerned about than others by a large proportion of the participants. For example, most of the users identify photos (84% are at least very concerned), videos (79%), their current location (76%), and family and relationships (54%) as profile attributes that participants are concerned the most about. The profile attributes that participants are least concerned about are proved to be birthday and sexual orientation. Note that the level of concern about the attributes is likely to depend on the context. For example, although a birthday attribute might seem harmless on its own, participants might feel different if a weather app would be collecting this information, or when a flashlight app requests the location of users [38].

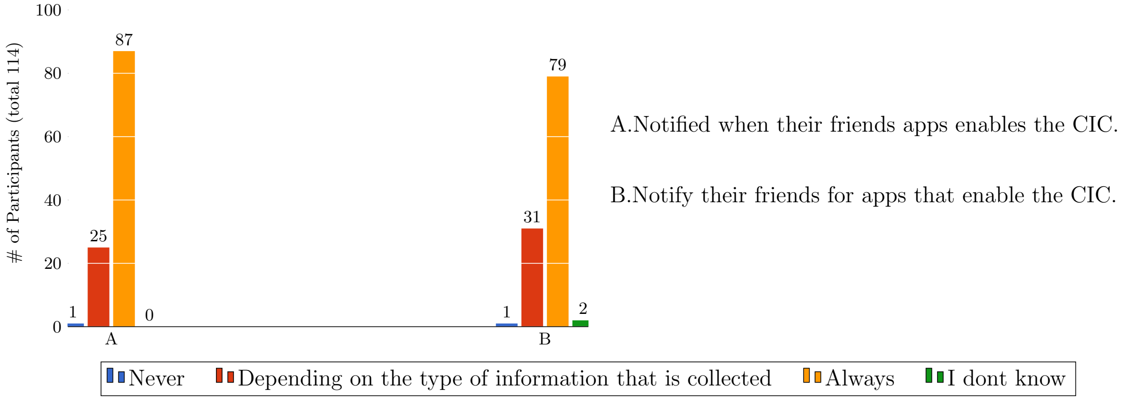


Fig. 5: Number of participants who want to a) be notified by their friends b) notify their friends when installing apps enabling collateral information collection (CIC).

For the third part of our questionnaire, we asked participants whether they want to be notified and also take an action for the collateral information collection. Concerning notification (see Fig. 5), we identify that 77% of users always want to be notified when friends' apps can collect their personal data, 22% only want to be notified in particular cases, while only about 1% do not want to be notified at all. Moreover, 69% of users always want to be notified when their apps are collecting information from their friends, 27% in particular cases and only about 1% not at all. We observe that users are also seriously concerned about harming their friends' privacy. This corroborates the finding on attested shreds of evidence concerning other-regarding preferences [39, 40]. Concerning notification, there exist tools that can be very useful to enhance privacy awareness for un-consented data collection. Note that Golbeck et al. [37] have shown that the privacy awareness of users can be changed significantly through educational methods.

When participants were asked which actions (see Fig. 6) they would take if they are notified that their friends' apps are about to collect their information (multiple answers allowed), 99 out of 114 participants answered that they would restrict access to their personal data while 8 participants answered that they would un-friend their Facebook friend. Only 5 participants answered that they would take no action. We emphasise that the reaction of a user may strongly depend on the relationship between the user and their friends. When participants were asked what action they would take if they are notified that one of their apps is about to collect their friends' information (multiple answers allowed), 64 out of 114 replied that they would restrict access to their friends' personal information for this app. Only 5 out of 114 answered that they would take no action. The answers to the questions in the third part help to confirm that the answers of our participants in the first part were not due to salience bias; participants who were concerned in the first part about not being notified for the *collateral information collection* replied that they also want to take an action in the third part.

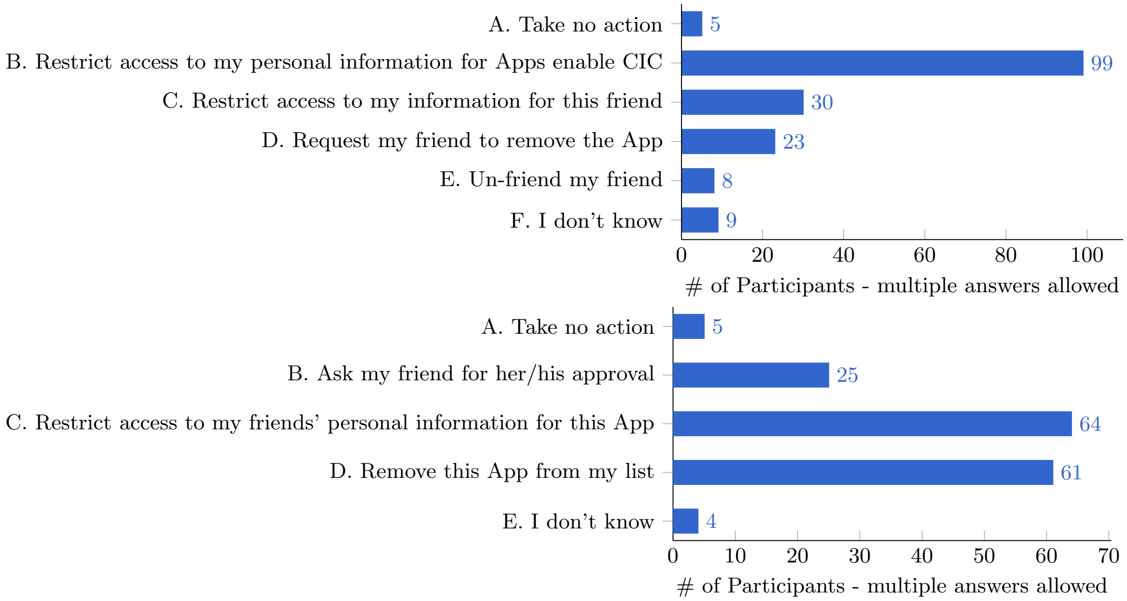


Fig. 6: Number of participants preferring to take a given action when a) their friends install an app, b) they install an app enabling collateral information collection (CIC).

The last part of our questionnaire collected demographics and statistics about Facebook and app usage. Participants were between 16 and 53 years old with an average age of 29 years. They have had their Facebook accounts for between 6 months and 10 years. Moreover, 69% of our participants have installed an app at least once, and among those 87% have installed 1 or 2 apps in the last six months. 54% of the participants were female, 42% male while 4% preferred not to disclose their gender. Participants varied greatly in their number of friends, from 10 to 1000. 51% changed their privacy settings on Facebook; 79% restricted who could see their profile information, 41% who could see them in searches, and 35% who can collect their information through friends apps (multiple answers were allowed). Interestingly, among users who already took an action by restricting their permissions to their friends' apps, 90% choose to be notified too. One explanation could be that privacy settings on Facebook are constantly changing and tracking these changes might be cumbersome [5]. Furthermore, 82% of our participants is pursuing or has obtained a higher education degree, where 55% had an IT background based on personal interest and 44% through higher education.

We conclude from our questionnaire that there is indeed evidence that users are concerned about collateral information collection. Our participants' concern is bidirectional, meaning that the large majority of them prefer to be notified and even take actions, whether this occurs from their friends or their own side, to prevent collateral information collection. While we do not want to generalise our findings to all Facebook users, we argue that these findings justify that collateral information collection is a privacy issue on Facebook, and thus, it merits a deeper investigation.[5]

3 Likelihood of Collateral Information Collection

In this section, we investigate the research question: *what is the likelihood that an installed app enables collateral information collection?* To answer this question, we estimate the likelihood of at

[5] http://iraklissymeonidis.info/fbapps/Survey/survey.html

least one Facebook friend of a user installing an app which enables collateral information collection. In order to estimate this likelihood, we build a model incorporating Facebook friends, the application space and the probabilities of installing apps. Then, we conduct two case studies to estimate this likelihood. Our first case study assumes a uniform app adoption model, which takes into account the popularity of an app and the Facebook users. Our second case considers a more realistic, non-uniform app adoption model, alongside different network topologies tailored to the Facebook ecosystem. We have instantiated our numerical estimations with the AppInspect dataset [9].

3.1 Model

Let an Online Social Network (OSN) with k users and the corresponding set be denoted by $\mathcal{U}$, i.e., $\mathcal{U} = \{u_1, \ldots, u_k\}$. The user is denoted by u, with $u \in \mathcal{U}$. Let $f \in \mathsf{F}^u$ be a friend of u and $\mathsf{F}^u \subseteq \mathcal{U}$ the set of u's friends. Moreover, let A_j be an app and $\mathcal{L}$ the set of all A_js that are offered by the OSN to every u_i, and s the size of the set, i.e., $\mathcal{L} = \{A_1, \ldots, A_s\}$. Moreover, let AU_j be the number of users who have installed A_j. For our likelihood estimation, we consider the number of Monthly Active Users (MAU) to represent the number of active users. For instance, when we did our research Facebook had $k = 1.3 \times 10^9$ users (i.e., MAU) [1] and more than $s = 25{,}000$ apps [9] (January 2015).

In order to estimate the likelihood that u's personal data can be collected by A_j, installed by f, we compute the probability of at least one arbitrary f installing any available A_j. Let Q^f be the probability of f installing A_j which enables collateral information collection. For all the friends of u (i.e., F^u) the probability of not installing any A_j is the product of probabilities for each f (assuming that these probabilities are independent; a reasonable approximation). Let Ω be the probability of at least one of u's friends installing A_j (regardless if u has installed A_j), i.e.,

$$\Omega = 1 - \prod_{f \in \mathsf{F}^u} (1 - Q^f) \ . \tag{1}$$

In order to estimate the likelihood Ω, we compute the probability of a friend of a user installing an app using two different app adoption models (uniform and a non-uniform) as follows.

3.2 Case study 1 – uniform distribution.

Each f decides whether to install A_j without considering any app adoption signals from other users. The probability of at least one friend of u installing A_j is uniformly distributed among u's friends and is equal to $1 - Q$. Note that $Q = Q^{f_1} = \cdots = Q^{f_{k'}}$ for uniform distribution with $1 \leq k' \leq k$. The probability Q, is then computed as the number of all users who installed the app, i.e., AU_j, divided by the number of users of the OSN, i.e., the cardinality of $\mathcal{U}$ (with regard to active users):

$$Q = \frac{AU_j}{|\mathcal{U}|} \ . \tag{2}$$

We used the publicly available dataset provided by Huber et al. [9, 17] to extract the range of MAU for apps which enable collateral information collection. The dataset consists of 16,808 Facebook apps from the period between 2012 and 2014. It contains the application name, ID, number of active users (daily, weekly and monthly) and the requested permissions. To illustrate the influence of different values of MAUs on Ω, we only consider the upper tier of apps, i.e., over 500,000 MAU; while the most popular app that collects friends' data has 10,000,000 MAU, therefore $5 \cdot 10^5 \leq AU_j \leq 1 \cdot 10^7$. To cover most users, we assume the number of friends for a given u (i.e., $|\mathsf{F^u}|$) to be between 0 and 1,000. Finally, we estimate the population of Facebook to be $1.1 \cdot 10^9$ MAU for the period of 2012 to 2014 [1].

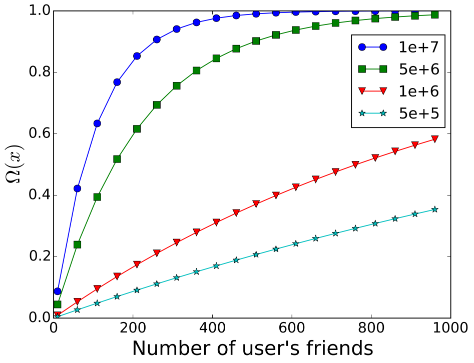
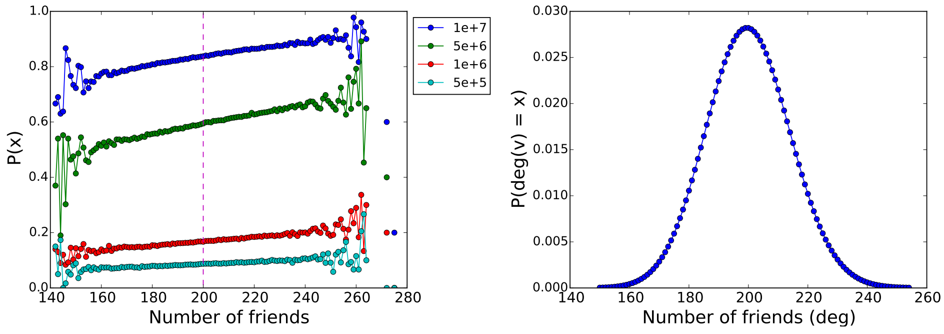


Fig. 7: Likelihood of collateral information collection based on real data [9] (per MAU).

For A_js with $AU_j \geq 5 \cdot 10^6$, the probability Ω grows steeply with the average number of friends (see Fig. 7). For a median of 200 friends the probability Ω is larger than 0.6. For a user with 300 friends and more, the probability Ω exceeds 0.8. Note that most of Facebook users have more than 200 friends [41]. From Eqns. (1) and (2) it is clear that Ω depends strongly on AU_j. For instance, our most popular app TripAdvisor [13] has approximately $1 \cdot 10^7$ MAU (i.e., $AU_j \approx 1 \cdot 10^7$). Assuming that on average a user has 200 friends [41] (i.e., $|\mathsf{F}^u| \approx 200$), and considering $\mathcal{U} = 1.1 \cdot 10^9$ (i.e., the population of Facebook), we estimate that the probability of at least one of u's friends installing TripAdvisor is larger than 78% ($\Omega \geq 0.78$).



Probability of being affected

Degree distribution

Fig. 9: Simulation results on an `ER` graph with k = 1,000,000 and d = 200.

To further elaborate our results, we empirically computed the probability Ω on a synthetically generated social network graph. We used the Erdős-Rényi [42] model (`ER`), to create a graph with 10^6 nodes and a mean node degree of $d = 200$. To emulate the theoretical results depicted by Eqn. (2), the simulation assigned app installations randomly following a uniform distribution. Therefore, each node in the graph has the same probability to install the app, i.e., $Q = \frac{AU_j}{|\mathcal{U}|}$.

Figure 8a shows the probabilities Ω observed in the simulation. Note that, the results are consistent with Fig. 7, with Ω increasing with the degree of the nodes. The plot is noisy for degrees less than 170 and higher than 230, since there are just a few nodes with these degrees in the simulated graph, as expected for an ER graph (see Fig. 8b). Indeed, ER graphs are created by including each edge uniformly at random with a certain probability p. Therefore, the resulting degree distribution is binomial with an expected value of 200. Fig. 8b shows the theoretical degree distribution of an ER graph with $k = 10^6$ and $d = 200$. The standard deviation is very low (14.14), thus most of the nodes have a degree close to 200.

The probability of a user having at least one friend with the app installed is computed assuming uniformity. Both Eqn. (2) and the simulation (see Fig. 8a) are based on the assumption that the probability of a friend installing A_j is equal to the mean app adoption rate of the network, i.e., $Q^f = Q$ for all friends of u. Case study 2 deals with the analysis of Ω when the uniformity assumption is relaxed.

3.3 Case study 2 – non-uniform distribution

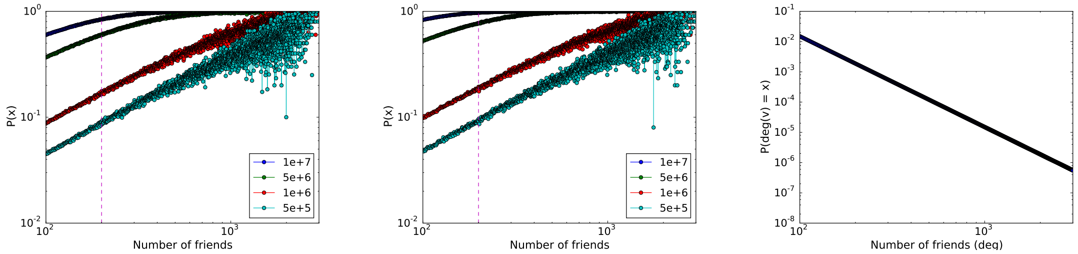
Realistic social networks do not usually conform to the uniformity assumption, thus assuming $Q = \frac{AU_j}{|\mathcal{U}|} = Q^{f_1} = \cdots = Q^{f_{k'}}$ where $1 \leq k' \leq k$ may not be realistic. App adoption has been proclaimed to be affected by different signals [21], which in turn may be affected by the underlying network structure. Research in social networks has reported node degree distributions following a power law and clustering coefficients much higher than in a random network [43–45]. We have resorted to simulations in order to introduce all these factors into the estimations of the probability Ω.

Each simulation uses two models: one to generate the synthetic network in which the probability under study is computed and the other to decide which users of the network install the application. Regarding the network topology, we have considered two different models: Barabási-Albert [46] (BA); that generates networks with a power-law degree distribution; and Watts-Strogatz [47] (WS), that is able to generate small-world networks with high clustering and small diameter. Regarding app adoption, two different models have been implemented: uniform (unif), where all users install an app with the same probability (and each installation is independent from other installations and the underlying network topology); and preferential (prop), where the probability of a user installing an app is proportional to the number of friends that have already installed the app owing to local network effects [48].

Therefore, the *configuration* of a simulation is defined by the network model, the app adoption model, and a set of parameters which configure both models: number of nodes denoted by k, expected mean node degree d, and a fraction of users installing the app denoted by e. The number of nodes k determines the network size for every model. The expected mean node degree d (and its relation to k) is used to adjust the additional specific parameters of the underlying network models. Finally, the fraction of users in the network installing the app e is used by the app adoption model to compute the actual user base.

We performed simulations with *configurations* using all possible pairs of network and app adoption models, for network sizes $k \in [10^4, 10^5, 10^6]$, mean degree $d \in [200]$ and $e = [5 \cdot 10^5/1.1 \cdot 10^9, 10^6/1.1 \cdot 10^9, 5 \cdot 10^6/1.1 \cdot 10^9, \cdot 10^7/1.1 \cdot 10^9]$. Figures 11 and 13 present the results for $k = 10^6$ and all tested e values. Owing to computational limitations, the largest simulated network size is $k = 10^6$; however, the trends are consistent across all network sizes. We have omitted the results for smaller k for the sake of brevity.

Figure 11 shows the results of the simulation using the BA model to generate the graph, together with the theoretical degree distribution of those graphs. The BA model generates graphs with a few very high degree nodes (known as hubs) and lots of low degree nodes, as shown in the theoretical degree distribution of a BA graph (see Fig. 10c). Regarding the simulation results for the two app adoption models, there are small differences between the observed Ω values. When combining networks generated with the BA model with a uniform app adoption model (see Fig. 10a), the probability for a hub to install an app is the same as for any other node. To the contrary, when combining BA with the prop app adoption model (see Fig. 10b), hubs have a higher probability of installing the app than non-hubs, since having a higher degree makes them more likely to have (more) friends with the app installed. As a consequence, each installation affects more users on average, and thus, Ω increases.



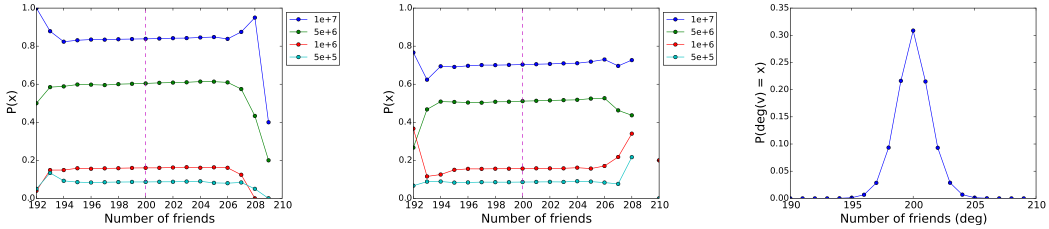
Uniform app adoption

Preferential app adoption

Degree distribution

Fig. 11: Simulation results on a BA graph with k = 1,000,000 and d = 200.

Fig. 13 shows the results of the simulation using the WS model, together with the theoretical degree distribution of those graphs. The WS model generates highly clustered networks[6] with very low variability in node degrees. Indeed and as it is illustrated in Fig. 12c, the vast majority of nodes will have a degree between 195 and 205. As a consequence, Fig. 12a and Fig. 12b are noisy for degrees outside this interval. The simulation results show, on the one hand, that Ω is about the same for all nodes in the graph, which is consistent with the fact that the degrees of the nodes are similar. On the other hand, the results also show a significant difference between using the unif (see Fig. 12a) and the prop (see Fig. 12b) app adoption models. This is caused by the strong clustering coefficient exhibited by the network. With unif app adoption, each node installs the app with the same probability. On the contrary, with prop app adoption, when an app is installed by a member of a community, it gets adopted by all other members easily. However, each new installation inside the same community implies only a small increase in the overall Ω. That is because most of the users affected by the installation were already affected by installations from other members of the community.



Uniform app adoption

Preferential app adoption

Degree distribution

Fig. 13: Simulation results on a WS graph with k = 1,000,000 and d = 200.

[6] The expected clustering coefficient can be adjusted with the rewiring probability parameter.

In summary, irrespective of app adoption models, the likelihood that a friend of a given user installs an app enabling collateral information collection is pretty significant. If the app is really popular, i.e., has millions of users, it is *very likely* that a user is affected by collateral information collection.

4 Significance of Collateral Information Collection

In this section, we answer the research question: *how significant is the collateral information collection?* In order to estimate how much information is collected, we build a model including friends, profile attributes, and Facebook access control, i.e., privacy settings and app permissions. This model can be used to compute the volume of the user's attributes that can be collected by apps (and thus app providers) when installed by the friends of a user. We investigate different ways of acquiring data: direct collection from the user themselves and indirect collection through the friends of a user. To instantiate our model, we use several snapshots of the most popular apps on Facebook using the AppInspect dataset [9].

4.1 Basic model

Users and users' friends Each user u_i in an OSN (i.e., $u_i \in \mathcal{U}$) has a personal profile where they can store, update, delete and administer their personal data [49]. A given u_i's profile consists of attributes a_i such as name, email, birthday and hometown. We denote the set of attributes of a u_i'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. Note that we use the term attribute and profile item interchangeably.

For any user u_i in $\mathcal{U}$, we consider u to be a user under consideration and $f_i \in \mathsf{F}^u$ one of their friends. Let F^{u*} be the union of u's friends and the u itself, or $\mathsf{F}^{u*} = \{u\} \cup \mathsf{F}^u$, and let f^* be an element of F^{u*}, i.e., $f^* \in \mathsf{F}^{u*}$. Clearly $\mathsf{F}^u \cap \{u\} = \emptyset$, as u is not a friend of u. For instance, $\mathsf{F}^{u*} = \{u, f_1, \dots, f_{k'}\}$ describes a user u and her k' friends, where $1 \leq k^{'} \leq k$.

Applications and application providers Let $\mathcal{L}$ be the set of apps an app provider (app provider) 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., $\mathsf{A}_j \subseteq \mathcal{T}$. Each A_j is owned and managed by an app provider denoted by P_j. The set of A_js that belong to P_j it is denoted by P_j, i.e., $\mathsf{P}_j \subseteq \mathcal{L}$. The set of all P_js is denoted by $\mathcal{AP}$ 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 $m = 2055$ app providers on Facebook indicating that a P_j can have more than one A_j, i.e., $\mathsf{P}_j = \{A_1 \dots A_{s\prime}\}$ with $1 \leq s\prime \leq 160$ [9].

Collateral information collection by an application A_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. We denote by A_j^{u,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,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,F^u}s installed by the users in F^u is denoted by $\mathsf{L}^{u,\mathsf{F}^u}$. Clearly, $\mathsf{L}^{u,\mathsf{F}^u} \subseteq \mathcal{L}$.

We denote by $\vec{a}_i$ a vector of size $n \times 1$ which corresponds to a_i, i.e.,

$$\vec{a}_i = [\overset{1}{0} \dots 0\overset{i}{1}0 \dots \overset{n}{0}] \ .$$

Moreover, we consider $\vec{A}_j^{u,\mathsf{F}^u}$ a vector of length n, which corresponds to $\mathsf{A}_j^{u,\mathsf{F}^u}$, i.e.,

Table 2: Notations

Notation	Description
$\mathcal{U} = \{u_1, \ldots, u_k\}$	Set of k users u_i in an OSN.
$\mathsf{F}^{u*} = \{u, f_1, \ldots, f_{k'}\}$	Set of k' friends (i.e., $f_i \in \mathsf{F}^u$) and u themselves (i.e., the user under consideration), where $k' \leq k$, $\mathsf{F}^{u*} = \{u\} \cup \mathsf{F}^u$ and $\{u\} \cap \mathsf{F}^u = \emptyset$.
$\mathcal{T} = \{a_1, \ldots, a_n\}$	Set of n attributes a_i of u's profile.
A_j / $A_j^{u,f}$ / A_j^{u,F^u}	An app j / an app j installed by: a user f / all users in F^u which can collect u's profile attributes.
$\mathsf{A}_j^{u,f}$ / $\mathsf{A}_j^{u,\mathsf{F}^u}$	Set of a_is for each A_j installed by: a user f / all users in F^u which can collect attributes of u's profile.
$\mathcal{L} = \{A_1, \ldots, A_s\}$	Set of s apps A_j hosted by an OSN.
$\mathsf{L}^{u,f}$ / $\mathsf{L}^{u,\mathsf{F}^u}$	Set of A_js installed by: a user f / all users in F^u, that can collect attributes of u's profile.
AU_j	The number of Monthly Active Users (MAU) of an A_j.
P_j / $P_j^{u,f}$ / P_j^{u,F^u}	An app provider P_j offering a set of A_js / a set of A_js installed by user f / a set of A_js installed by all users in F^u offered by P_j that can collect u's profile attributes.
$\mathsf{P}_j^{u,f}$ / $\mathsf{P}_j^{u,\mathsf{F}^u}$	Set of a_is all A_js installed by, a user f / all users in F^u and belong to the P_j that can collect u's profile attributes.
$\mathcal{AP} = \{P_1 \ldots P_m\}$	Set of m app providers P_js hosted by an OSN.
$\mathsf{AP}^{u,f}$ /$\mathsf{AP}^{u,\mathsf{F}^u}$	Set of P_js whose A_js: installed by a user f / all users in F^u which can collect u's profile attributes.
$\mathsf{C}_{PM,A_j}^{u,f}$ / $\mathsf{C}_{PM,A_j}^{u,\mathsf{F}^{u*}}$	Set of permissions r_i an A_j can request and, a user f (i.e., $f \in \mathsf{F}^{u*}$) / all users in F^{u*} (i.e., $\mathsf{F}^{u*} = \{u\} \cup \mathsf{F}^u$), accept(s) to collect u's profile attributes.
$\mathsf{C}_{PV}^{u,f}$ / $\mathsf{C}_{PV}^{u,\mathsf{F}^u}$	Set of privacy settings that an OSN offers and allows: a user f / all users in F^u, to access u's profile attributes.
$\mathsf{C}_{A_j}^{u,f}$ / $\mathsf{C}_{A_j}^{u,\mathsf{F}^u}$	Set of access settings (i.e., a combination of permissions and privacy settings) granted and accepted by u and: a user f / / all users in F^u to access and collect u's profile attributes by an A_j.
$\mathsf{\Pi}_{A_j}^{u,\mathsf{F}^u}$ / $\mathsf{\Pi}_{P_j}^{u,\mathsf{F}^u}$	Set of a_is for each, A_j / P_j, installed by all users in F^u that can collect u's profile attributes.
$\mathsf{\Delta}^u_{A_j^u, A_j^{u,\mathsf{F}^u}}$ / $\mathsf{\Delta}^u_{P_j^u, P_j^{u,\mathsf{F}^u}}$	Set of a_is for each, A_j / P_j, *exclusively* installed by all users in F^u (and not through the user themselves, i.e., $\mathsf{A}_j'^{u} \cap \mathsf{A}_j^{u,\mathsf{F}^u}$) that can collect u's profile attributes.
d / e	The mean node degree / the fraction of users that installed an app, i.e., $\frac{AU}{\lvert\mathcal{U}\rvert}$

$$\vec{A}_j^{u,\mathsf{F}^u} = \bigvee_{a_i \in A_j^{u,\mathsf{F}^u}} \vec{a_i} \ := \vec{A}_j^{u,\mathsf{F}^u}[i] \ = \begin{cases} 1 & \text{if } a_i \ \in \mathsf{A}_j^{u,\mathsf{F}^u} \ , \\ 0 & \text{if } a_i \ \notin \mathsf{A}_j^{u,\mathsf{F}^u} \ , \end{cases} \tag{3}$$

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

Note that:

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

For instance, an $\mathsf{A}_j^{u,\mathsf{F}^u} = \{a_1, a_i, a_n\}$ is represented as $\vec{A}_j = \vec{a}_1 \vee \vec{a}_i \vee \vec{a}_n = [\overset{1}{1}0\ldots0\overset{i}{1}0\ldots0\overset{n}{1}]$. It represents the attributes that can be collected by A_j when is installed by f (i.e., the user's friend).

Collateral information collection by application provider P_j We denote by $\mathsf{AP}^{u,\mathsf{F}^u}$ the set of app providers whose apps A_j^{u,F^u}s are installed by users in F^u and who can collect attributes of user u's profile. Hence,

$$\mathsf{AP}^{u,\mathsf{F}^u} = \bigcup_{f \in \mathsf{F}^u} \mathsf{AP}^{u,f} \quad . \tag{4}$$

Each P_j^{u,F^u} consists of a set of A_j^{u,F^u}s denoted by $\mathsf{P}_j^{u,\mathsf{F}^u}$ which users in F^u installed. Each $\mathsf{P}_j^{u,\mathsf{F}^u}$ can collect attributes of u's profile. To identify which a_is can be collected by P_j we consider $\vec{P}_j^{u,\mathsf{F}^u}$ a vector of length n (i.e., $n \in \mathcal{T}$), which corresponds to $\mathsf{P}_j^{u,\mathsf{F}^u}$, i.e.,

$$\vec{P}_j^{u,\mathsf{F}^u} = \bigvee_{\substack{A \in \mathsf{P}_j^{u,f} \\ f \in \mathsf{F}^u}} \vec{A}^{u,f} = \bigvee_{A \in \mathsf{P}_j^{u,\mathsf{F}^u}} \vec{A}^{u,\mathsf{F}^u} \quad . \tag{5}$$

Note that: $\vec{P}_j^{u,\mathsf{F}^u} = \bigvee_{f \in \mathsf{F}^u} \vec{P}_j^{u,f} = (\vec{P}_j^{u,f_1} \vee \cdots \vee \vec{P}_j^{u,f_i})$.

The complexity of evaluating Eqn. (5) for all f in F^u is $\mathcal{O}(n \times |\mathsf{P}_j^{u,\mathsf{F}^u}|)$.

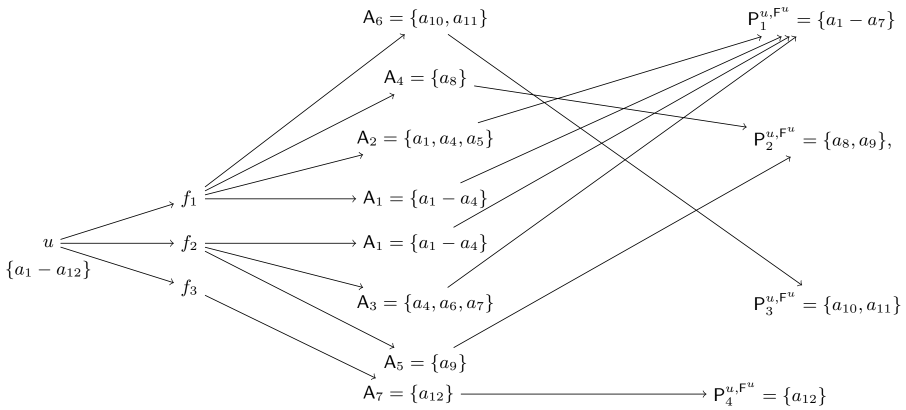

$\mathsf{A}_6 = \{a_{10}, a_{11}\}$
$\mathsf{P}_1^{u,\mathsf{F}^u} = \{a_1 - a_7\}$
$\mathsf{A}_4 = \{a_8\}$
$\mathsf{A}_2 = \{a_1, a_4, a_5\}$
$\mathsf{P}_2^{u,\mathsf{F}^u} = \{a_8, a_9\}$,
f_1
$\mathsf{A}_1 = \{a_1 - a_4\}$
u
$\{a_1 - a_{12}\}$
f_2
$\mathsf{A}_1 = \{a_1 - a_4\}$
f_3
$\mathsf{A}_3 = \{a_4, a_6, a_7\}$
$\mathsf{P}_3^{u,\mathsf{F}^u} = \{a_{10}, a_{11}\}$
$\mathsf{A}_5 = \{a_9\}$
$\mathsf{A}_7 = \{a_{12}\}$
$\mathsf{P}_4^{u,\mathsf{F}^u} = \{a_{12}\}$

Fig. 14: Example for collateral information collection while enabling profiling to P_j.

Example (see Fig. 14. Let $\mathsf{F}^u = \{f_1, f_2, f_3\}$ friends of u. The set of A_js that all $f \in \mathsf{F}^u$ installed 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_js installed is described as $\mathsf{AP}^{u,\mathsf{F}^u} = \mathsf{AP}^{u,f_1} \cup \mathsf{AP}^{u,f_2} \cup \mathsf{AP}^{u,f_3} = \{\mathsf{P}_1^{u,f_1}, \mathsf{P}_2^{u,f_1}, \mathsf{P}_3^{u,f_1}\} \cup \{\mathsf{P}_1^{u,f_2}, \mathsf{P}_2^{u,f_2}\} \cup \{\mathsf{P}_4^{u,f_3}\} = \{(\mathsf{P}_1^{u,f_1} \cup \mathsf{P}_1^{u,f_2}), (\mathsf{P}_2^{u,f_1} \cup \mathsf{P}_2^{u,f_2}), \mathsf{P}_3^{u,f_1}, \mathsf{P}_4^{u,f_3}\}$. Each $\mathsf{P}_1^{u,\mathsf{F}^u} = \mathsf{P}_1^{u,f_1} \cup \mathsf{P}_1^{u,f_2} = \{(\mathsf{A}_1^{u,f_1} \cup \mathsf{A}_2^{u,f_1}) \cup (\mathsf{A}_1^{u,f_2} \cup \mathsf{A}_3^{u,f_2})\}$, $\mathsf{P}_2^{u,\mathsf{F}^u} = \mathsf{P}_2^{u,f_1} \cup \mathsf{P}_2^{u,f_2} = \{\mathsf{A}_4^{u,f_1} \cup \mathsf{A}_5^{u,f_2}\}$, $\mathsf{P}_3^{u,\mathsf{F}^u} = \{\mathsf{A}_6^{u,f_1}\}$ and $\mathsf{P}_4^{u,\mathsf{F}^u} = \{\mathsf{A}_7^{u,f_3}\}$. Each A_j installed by u's friends can collect a set of a_i attributes 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 each P_j is $\mathsf{P}_1^{u,\mathsf{F}^u} = (\mathsf{A}_1^{u,f_1} \cup \mathsf{A}_2^{u,f_1}) \cup (\mathsf{A}_1^{u,f_2} \cup \mathsf{A}_3^{u,f_2}) = (\{a_1, a_2, a_3, a_4\} \cup \{a_1, a_4, a_5\}) \cup (\{a_1, a_2, a_3, a_4\} \cup \{a_4, a_6, a_7\}\}) = \{a_1 - a_7\}$, $\mathsf{P}_2^{u,\mathsf{F}^u} = \mathsf{A}_4^{u,f_1} \cup \mathsf{A}_5^{u,f_2} = \{a_8\} \cup \{a_9\} = \{a_8, a_9\}$, $\mathsf{P}_3^{u,\mathsf{F}^u} = \mathsf{A}_6^{u,f_1} = \{a_{10}, a_{11}\}$ and $\mathsf{P}_4^{u,\mathsf{F}^u} = \mathsf{A}_7^{u,f_1} = \{a_{12}\}$.

Exclusive collateral information collection by application A_j and application provider P_j We denote by $\Delta^u_{A_j^u, A_j^{u,\mathsf{F}^u}}$ the set of a_i attributes that can be collected by A_j exclusively from u's friends (and not through the user themselves), i.e., $\mathsf{A}_j'^u \cap \mathsf{A}_j^{u,\mathsf{F}^u}$.

Let $\vec{\varDelta}^u_{A_j^u, A_j^{u,\mathsf{F}^u}}$ be a vector of length n which $\Delta^u_{A_j^u, A_j^{u,\mathsf{F}^u}}$ provides, where $n = |\mathcal{T}|$, i.e.,

$$\vec{\varDelta}^u_{A_j^u, A_j^{u,\mathsf{F}^u}} = \vec{A}_j'^u \bigwedge \vec{A}_j^{u,\mathsf{F}^u} \quad . \tag{6}$$

Note that $\vec{x}' \wedge \vec{x} = [\overset{1}{0} \dots \overset{n}{0}]$ and $\vec{x}' \vee \vec{x} = [\overset{1}{1} \dots \overset{n}{1}]$. The complexity of evaluating Eqn. (6) for all $f \in \mathsf{F}^u$ is $\mathcal{O}(n^4)$.

Similarly, we denote by $\Delta^u_{P_j^u, P_j^{u,\mathsf{F}^u}}$ the set of a_is that can be collected by A_js exclusively from u's friends and belong to P_j^{u,F^u}, and $\vec{\varDelta}^u_{P_j^u, P_j^{u,\mathsf{F}^u}}$ be a vector of length n which $\Delta^u_{P_j^u, P_j^{u,\mathsf{F}^u}}$ provides, i.e.,

$$\vec{\varDelta}^u_{P_j^u, P_j^{u,\mathsf{F}^u}} = \vec{P}_j'^u \bigwedge \vec{P}_j^{u,\mathsf{F}^u} \quad . \tag{7}$$

The complexity of evaluating Eqn. (7) for all $f \in \mathsf{F}^u$ is $\mathcal{O}(n^4 \times |\mathsf{P}_j^u| \times |\mathsf{P}_j^{u,\mathsf{F}^u}|)$.

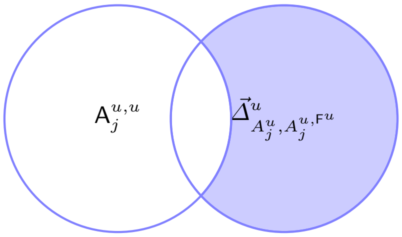


Fig. 15: Exclusive collateral information collection by application A_j

4.2 Application permissions and user's privacy settings

In order to control the collateral information collection of A_j from a u's profile, *access control settings* are provided by the OSN. On Facebook the *access control settings* consist of permissions and privacy settings [8]. Permissions depend on the friend f, where $f \in \{u\} \cup \mathsf{F}^u$ and on the application A_j, as each f should accept the permissions that each A_j is requesting. Privacy settings also depend on the user u, as each u chooses with whom their profile information is shared.

Permissions and application A_j Each A_j can request a set of permissions r_i to be accepted from a user f such as $u$ser_$e$mails, $u$ser_$f$riends and $f$riends_$b$irthday. We denote by $\mathcal{AC}$ the set of permissions an application A_j can request from a friend f (with cardinality n), i.e., $\mathcal{AC} = \{r_1, \ldots, r_n\}$, where $f \in \mathsf{F}^{u*}$. Moreover, we denote by $\mathsf{C}^{u,\mathsf{F}^{u*}}_{PM,A_j}$ the set of permissions an A_j can request, and F^{u*} accepts; thus A_j can collect attributes of u's profile, where $1 \leq j \leq s$. Clearly $\mathsf{C}^{u,\mathsf{F}^{u*}}_{PM,A_j} \subseteq \mathcal{AC}$.

We consider $\vec{r_i}$ a vector of length n which corresponds to r_i, for $1 \leq i \leq n$, i.e.,

$$\vec{r_i} = [\overset{1}{0} \ldots 0\overset{i}{1}0 \ldots \overset{n}{0}] \ .$$

Moreover, we consider $\vec{C}^{u,\mathsf{F}^{u*}}_{PM,A_j}$ a vector of length n, which corresponds to $\mathsf{C}^{u,\mathsf{F}^{u*}}_{PM,A_j}$, i.e.,

$$\vec{C}^{u,\mathsf{F}^{u*}}_{PM,A_j} = \bigvee_{r \in \mathsf{C}^{u,\mathsf{F}^{u*}}_{PM,A_j}} \vec{r} \ := \vec{C}^{u,\mathsf{F}^{u*}}_{PM,A_j}[i] \ = \begin{cases} 1 & \text{if access is provided by } \mathsf{F}^{u*} \\ & \text{to } u\text{'s profile for a given } A_j, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

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

Privacy settings and user u Each u can allow a user f (i.e., $f \in \mathsf{F}^u$) such as a friend, friend of a friend or any user to access the attributes of u's profile. We denote by $\mathsf{C}^{u,\mathsf{F}^u}_{PV}$ the set of attributes of u's profile that u allows to access for F^u using the privacy settings of an OSN.

We consider $\vec{C}^{u,\mathsf{F}^u}_{PV}$ a vector of length n, which corresponds to $\mathsf{C}^{u,\mathsf{F}^u}_{PV}$, i.e.,

$$\vec{C}^{u,\mathsf{F}^u}_{PV} = \begin{cases} [\overset{1}{1} \ldots \overset{i}{1} \ldots \overset{n}{1}] & \text{if an access to } u\text{'s profile is provided by } u \text{ to } \mathsf{F}^u, \\ [\overset{1}{0} \ldots \overset{i}{0} \ldots \overset{n}{0}] & \text{otherwise} \ , \end{cases} \tag{9}$$

for $1 \leq i \leq n$.

Permissions vs. privacy settings We denote as $\mathsf{C}^{u,\mathsf{F}^u}_{A_j}$ the set of access settings provided by u and F^u to u's profile for an A_j, as a combination of permissions (i.e., $\mathsf{C}^{u,\mathsf{F}^{u*}}_{PM,A_j}$) and privacy settings (i.e., $\mathsf{C}^{u,\mathsf{F}^u}_{PV}$).

We consider $\vec{C}^{u,\mathsf{F}^u}_{A_j}$ a vector of length n which correspond to $\mathsf{C}^{u,\mathsf{F}^u}_{A_j}$, i.e.,

$$\vec{C}^{u,\mathsf{F}^u}_{A_j} = \vec{C}^{u,\mathsf{F}^u}_{PV} \wedge \vec{C}^{u,\mathsf{F}^{u*}}_{PM,A_j} \ , \tag{10}$$

for $1 \leq j \leq s$.

- Remark: $a \wedge b = \begin{cases} 1 & if \ a = b = 1, \\ 0 & \text{otherwise,} \end{cases}$
- Extension: $\vec{a} \wedge \vec{b} = \vec{c}$ where $\vec{a}[i] \wedge \vec{b}[i] = \vec{c}[i]$.

The complexity of evaluating Eqn. (10) for all f in F^u is $\mathsf{O}(n^2)$.

Collateral information collection with permissions We denote by $\mathsf{\Pi}^{u,\mathsf{F}^u}_{A_j}$ and $\mathsf{\Pi}^{u,\mathsf{F}^u}_{P_j}$ the set of attributes that can be collected by A_j and P_j respectively, for the accepted access settings to u's profile by u and F^u.

Let $\vec{\mathit{\Pi}}^{u,\mathsf{F}^u}_{A_j}$ be a vector of length n which $\mathsf{\Pi}^{u,\mathsf{F}^u}_{A_j}$ provide, i.e.,

$$\vec{\mathit{\Pi}}^{u,\mathsf{F}^u}_{A_j} = \vec{A}^{u,\mathsf{F}^u}_j \wedge \vec{C}^{u,\mathsf{F}^u}_{A_j} \ . \tag{11}$$

for $1 \leq j \leq s$.

The complexity of evaluating Eqn. (11) for all f in F^u is $\mathcal{O}(n^3)$.

Let $\vec{\Pi}^{u,\mathsf{F}^u}_{P_j}$ be a vector of length n which $\Pi^{u,\mathsf{F}^u}_{P_j}$ provides, i.e.,

$$\vec{\Pi}^{u,\mathsf{F}^u}_{P_j} = \bigvee_{A_j \in \mathsf{P}^{u,\mathsf{F}^u}_j} (\vec{A}^{u,\mathsf{F}^u}_j \wedge \vec{C}^{u,\mathsf{F}^u}_{A_j}) \ . \tag{12}$$

for $1 \leq j \leq s$.

The complexity of evaluating Eqn. (12) for all f in F^u is $\mathcal{O}(n^3 \times |\mathsf{P}^{u,\mathsf{F}^u}_j|)$.

Exclusive collateral information collection with permissions We denote by $\Delta^u_{A^u_j, A^{u,\mathsf{F}^u}_j}$ and $\Delta^u_{P^u_j, P^{u,\mathsf{F}^u}_j}$ the set of attributes that can be collected by A_j and P_j (respectively) exclusively from u's friends (and not through the user themselves, i.e., $\mathsf{A}'^u_j \cap \mathsf{A}^{u,\mathsf{F}^u}_j$), for the accepted access settings to u's profile by u and F^u.

Let $\vec{\Delta}^u_{A^u_j, A^{u,\mathsf{F}^u}_j}$ be a vector of length n which $\Delta^u_{A^u_j, A^{u,\mathsf{F}^u}_j}$ provides, i.e.,

$$\vec{\Delta}^u_{A^u_j, A^{u,\mathsf{F}^u}_j} = (\vec{A}^u_j \wedge \vec{C}^u_{A_j})' \wedge (\vec{A}^{u,\mathsf{F}^u}_j \wedge \vec{C}_{A^{u,\mathsf{F}^u}_j}) \ , \tag{13}$$

for $1 \leq j \leq s$.

The complexity of evaluating Eqn. (13) for all f in F^u is $\mathcal{O}(n^{12})$ and $\mathcal{O}(n^{12} \times |\mathsf{P}^u_j| \times |\mathsf{P}^{u,\mathsf{F}^u}_j|)$

Let $\vec{\Delta}^u_{P^u_j, P^{u,\mathsf{F}^u}_j}$ be a vector of length n which $\Delta^u_{P^u_j, P^{u,\mathsf{F}^u}_j}$ provides, i.e.,

$$\vec{\Delta}^u_{P^u_j, P^{u,\mathsf{F}^u}_j} = \vec{\Pi}'^u_j \bigwedge \vec{\Pi}^{u,\mathsf{F}^u}_j \ . \tag{14}$$

for $1 \leq j \leq s$.

The complexity of evaluating Eqn. (14) for all f in F^u is $\mathcal{O}(n^{12} \times |\mathsf{P}^u_j| \times |\mathsf{P}^{u,\mathsf{F}^u}_j|)$.

4.3 Numerical study of collateral information collection: the case of Facebook apps

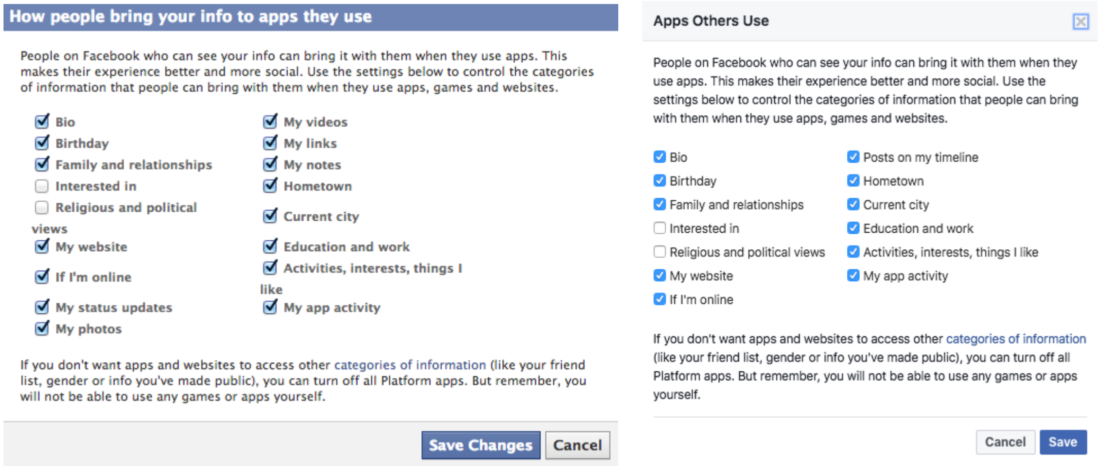
In order to illustrate the significance of collateral information collection, we extended our analysis to Facebook apps (i.e., A_js) and app providers (i.e., P_js) using the AppInspect dataset [9, 17]. For each A_j, apart from the application name and ID, the dataset provides us with the requested permissions and the A_js each P_j owns. We compute and compare the proportion of attributes A_js and their respective P_js can collect: 1) through the user themselves (i.e., direct collection by apps and potential data fusion by app providers), 2) through the user and the user's friends combined (i.e., collateral information collection) and 3) exclusively through the user's friends (i.e., collateral information collection). Out of the 16,808 apps in the dataset, 1,202 enable *collateral information collection*˙ corresponding to 7.15%. Out of these 1,202, our analysis focuses on A_js and P_js that have more than $AU \geq 10{,}000$ MAU; there are 207 and 88, respectively.[7]

Before delving into the details, we point our two limitations of our numerical study. First, the dataset does not contain information about the individual users of the listed apps, but only their aggregate numbers. Therefore, it is not possible to perform a per-user evaluation of the privacy loss due to collateral information collection. Instead, we had to resort to per app and per app provider analysis as it would impact an average user and an average friend of the user.

Second, the dataset does not contain information about the privacy settings of individual users. Therefore we are not able to factor in those when numerically evaluating the extent of the collateral information collection; we assume that all permission requests are granted. This is not a far-fetched assumption, as default privacy settings for apps have been and still are very much pro-sharing (see

[7] `http://iraklissymeonidis.info/fbapps/fb_apps_statistics/index.html`.

Fig. 17). In a nutshell, most attributes are accessible by default for apps installed by the friends of a user. Furthermore, the more recent "Apps Others use" dialogue (see Fig. 17 (b)), does not list photos, potentially indicating that these might not be accessible for apps through Facebook friends.



How people bring your info to apps they use

People on Facebook who can see your info can bring it with them when they use apps. This makes their experience better and more social. Use the settings below to control the categories of information that people can bring with them when they use apps, games and websites.

- ☑ Bio
- ☑ Birthday
- ☑ Family and relationships
- □ Interested in
- □ Religious and political views
- ☑ My website
- ☑ If I'm online
- ☑ My status updates
- ☑ My photos
- ☑ My videos
- ☑ My links
- ☑ My notes
- ☑ Hometown
- ☑ Current city
- ☑ Education and work
- ☑ Activities, interests, things I like
- ☑ My app activity

If you don't want apps and websites to access other categories of information (like your friend list, gender or info you've made public), you can turn off all Platform apps. But remember, you will not be able to use any games or apps yourself.

Apps Others Use

People on Facebook who can see your info can bring it with them when they use apps. This makes their experience better and more social. Use the settings below to control the categories of information that people can bring with them when they use apps, games and websites.

- ☑ Bio
- ☑ Birthday
- ☑ Family and relationships
- □ Interested in
- □ Religious and political views
- ☑ My website
- ☑ If I'm online
- ☑ Posts on my timeline
- ☑ Hometown
- ☑ Current city
- ☑ Education and work
- ☑ Activities, interests, things I like
- ☑ My app activity

If you don't want apps and websites to access other categories of information (like your friend list, gender or info you've made public), you can turn off all Platform apps. But remember, you will not be able to use any games or apps yourself.

Fig. 17: Default permissions settings for apps on (a) 2012 (left) and (b) 2017 (right) on Facebook.

Direct information collection First, we aimed to investigate how many attributes apps can collect: from users giving explicit consent when installing an app. Figure 18 (a) shows that more than half of the 207 apps collect a single attribute from the users. Other characteristic number of attributes are 2 (14%) and 7 (12%), respectively. Furthermore, there was a single app collecting 9 attributes. Note that 10% of the apps did not ask for any extra attribute (i.e., $u$ser_$x$xx) outside of the basic information granted to all apps.

Collateral information collection Second, and more importantly, we were interested in the extent of collateral information collection, when apps and app providers become entitled to collect attributes of a user without direct consent, through their friend installing the app. Performing the analysis over the dataset with regard to apps, using Eqn. (3) we found that 72.4% of apps can collect exactly one attribute from F^u (see Fig. 18 (b) for details). There are some apps which collect between 2 and 7 attributes in a collateral fashion, with a peak at 6 attributes. Furthermore, there were several apps which collected 11 attributes from the friends of the user.

Exclusive collateral information collection With regard to attributes which can be collected *exclusively* from the friends of a user (i.e., collected from friends but not from the user themselves, see Fig. 15), using Eqn. (6), we found that more than 30% of the apps collect at least one attribute. Specifically, 28.9% of the apps under consideration can collect exactly one attribute, 1.45% as much as 10, and one app 11 attributes (see Fig. 18 (c) for details).

Potential data fusion by app providers offering multiple apps Considering direct information collection and turning towards the 88 app providers, Fig. 18 (d) shows that providers predominantly collect 1 or 2 attributes per user. The slight increase in collected attributes indicates that app providers offering multiple apps are able to gather an enriched set of personal information from a single user. Extremes also do exist: one app provider collects as much as 17

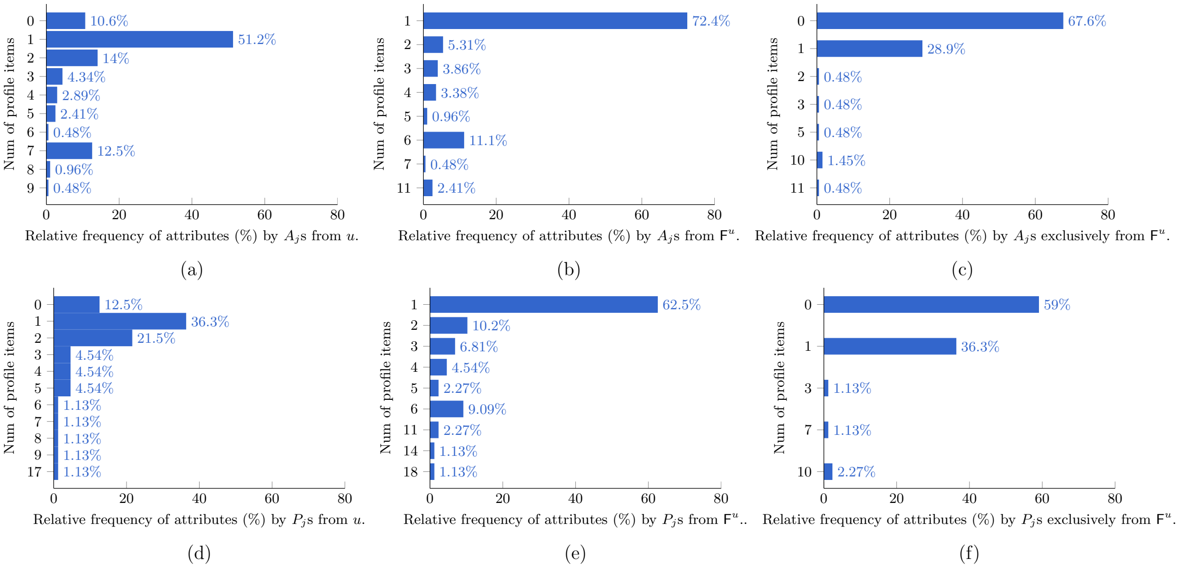


Fig. 18: Number of attributes gathered via direct, collateral and exclusive *Collateral Information Collection* wrt apps (A_j, top) and app providers (P_j, bottom)

attributes from a single user via a combined offering of multiple apps. It is interesting to see that are 4 app providers collecting at least 7 attributes, while there are almost 30 apps collecting the same amount; this indicates that data-ravenous apps are concentrated in a few app providers.

Data fusion also strengthens the extent of collateral information collection, as seen in Fig. 18 (e) compared to Fig. 18 (b): there is a slight but visible increase towards collecting more attributes compared to the single app scenario throughout the whole range of the histogram. Furthermore, some app providers collect as much as 14 and 18 attributes, exhibiting strong potential for data fusion and thus *profiling*. A similar effect of data fusion can be seen with regard to *exclusive* collateral information collection in Fig. 18 (f). A surprising nuance is the disappearance of the 11-attribute app: the *exclusive* collateralness of a given attribute vanishes, if another app from the same app provider collects the same one directly.

The most important observation here is that collateral information collection and *data fusion* are orthogonal mechanisms, which, surprisingly, amplify each other's effects, resulting in more pronounced and obscure information gathering. As a countermeasure, bringing transparency to such a scenario should provide straightforward benefits to Facebook users.

Potentially sensitive profile items The collection of profile items deemed potentially sensitive by users deserve a more detailed look. Both related literature and our user study (see Sect. 2) found that attributes with connection to location, relationships, religious/ political views and photos/videos are particularly guarded more closely by users (i.e., where participants were at least very concerned). Analysing the AppInspect dataset with an eye on permissions granting access to such information, we derived the following results concerning collateral and exclusive collateral information collection and data fusion (see Fig. 19). The two most popular groups affected by collateral information collection are location-related attributes (including work history, education, hometown and location) and photos. Almost, half of the 207 apps collected the photos of the user's friends (97), with location (39) and work history (35) being the second and third most popular attributes. It is also easy to see that apps requesting these three attributes are clustered to a few app providers (between 2 to 4 app per app provider, mean value over all 3 attributes). In case of

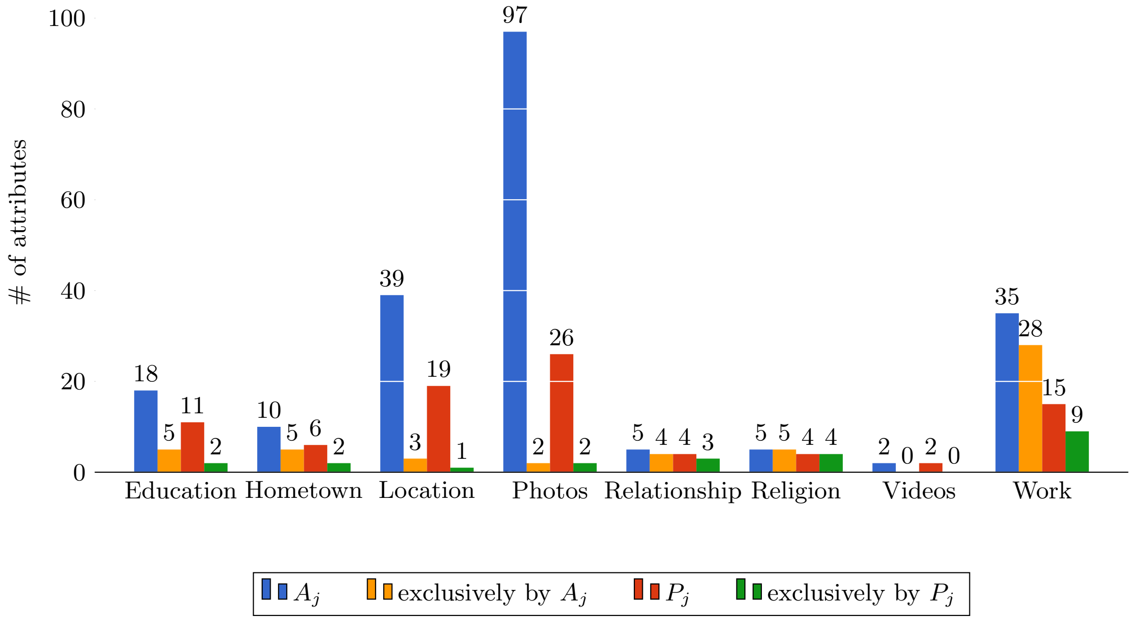


Fig. 19: Total number of apps and appPs requesting collateral information collection of sensitive attributes (per attribute)

less popular collaterally collected attributes, the requesting apps are more uniformly distributed across app providers (between 1 to 1.6 app per app provider, mean value over the remaining 5 attributes). Note that exclusive collateral information collection also happens as evidenced by the yellow and green bars in Fig. 19.

5 Legal analysis

In this section, we investigate the research question: *Under the data protection legislation, is collateral information collection considered a risk for the protection of the personal data of Facebook users?* To answer this question, we use the new General Data Protection Regulation [18] as point of reference. We investigate the responsibility of collateral information collection, meaning who the data controller and data processor are. We discuss the relevance of transparency and consent for data protection. Furthermore, we review the necessity of implementing concepts of data protection by design and default. Based our legal analysis, we identify who is responsible for such information collection, and argue that the lack of transparency and consent makes collateral information collection dubious from a data protection point of view.

5.1 General Data Protection Regulation

After 20 years and many technological innovations, the Data Protection Directive (Directive 95/45/EC) [50] was due for an update. A Regulation was chosen over a Directive, as a Regulation applies directly and does not need to be transposed into national law. After long negotiations on 24 May 2016 the European General Data Protection Regulation (GDPR) [18] entered into force and it will start to apply from 25 May 2018. The GDPR constitutes a single set of rules for all Member States which regulates the processing of personal data if it is done in the context of an establishment of an actor (controller/processor) within the EU, or if personal data of people who are in the EU is processed in the context of offering goods or services or the monitoring of their behaviour. Besides this extended territorial scope, the GDPR also brought some additional changes. Especially interesting for the question of this paper is the specification of (joint) controllership and the introduction of data protection by design and by default.

5.2 Controllers and processors

The GDPR, as the Directive 95/46/EC before, still uses the distinction between controllers and processors to associate certain duties on actors in the case of data processing. A controller is anybody who, alone or with others, determines the purposes and means of the processing of personal data [18]. A processor is any party which processes the personal data on behalf of the controller. This distinction is important, as the controller is the main responsible party for implementing and complying with the provisions of the GDPR.

The controller should normally be, as the name already says, the entity who has the control. The control can be stemming from 1) explicit or implicit legal competence (the controller might be appointed by law or the capacity to determine is not explicitly laid down by law but stems from common legal provisions or established legal practice) or b) factual influence: in this case an assessment of the factual circumstances is made [51]. It should be noted that therefore, the determination of who is controller and processor is a factual one and cannot be contractually allocated to another person/entity which, considering the factual circumstances, has neither legal nor factual influence on how the personal data are processed. However, the content of contracts can be useful to assess who factually has the control. The control relates to the determination of the purposes and means of the data processing. Essentially, this means what level of influence the entity has on the "why" and the "how" of the processing [51].

5.3 Joint control

It is possible that for a single data processing several controllers are involved. The Directive was not explicit on this and the concept of joint controllers was mainly provided by the Article 29 Working Party, however, the GDPR now includes explicitly the concept of joint controllers in Article 26 GDPR. The Regulation specifies that where two or more controllers jointly determine the purposes and means of processing, they are joint controllers and they need to determine their respective responsibilities for compliance with the Regulation in a transparent manner. This includes an arrangement the essence of which shall be made available to the data subject, and which may designate a contact point for data subjects. Nonetheless, the data subject may exercise his or her rights in respect of and against each of the controllers.

Just the fact that different parties co-operate when processing personal data does not mean that they are joint controllers. If they do not share common purpose or means with regard to the specific processing, then it might only be a transfer of data between separate controllers. It will often not be clear-cut and depend on the factual circumstances whether several controllers are considered to be separate controllers or in case they are considered joint controllers, to which extent they share the control. The same data set can be processed by several separate controllers or by joint controllers, e.g. when a travel agency books a journey and hotel every controller (travel agency, airline, and hotel) is a separate controller, however, if the controllers set up a common platform for the booking and thereby determine the essential elements of the means, they are joint controllers for this data processing [51].

5.4 Collateral information collection: who is controller?

The Article 29 Working Party in their opinion 5/2009 on online social networking specified that social networking service providers, such as in the case of this paper Facebook, are data controllers under the Data Protection Directive [52]. As the concept of controller and processor has not much changed under the Regulation, it can be understood that Facebook is still a controller under the Regulation. Furthermore, application providers might be data controllers, if they develop applications which run in addition to the ones from social networking service providers, and users decide to use such an application [53].

Whether Facebook and the app provider are separate or joint controllers will depend on in how far they share the purposes and means of the data processing. In general, the app provider will be a separate controller when it determines its own purposes and means for the processing of personal data. With regards to the processing activities of the app provider the social networking service provider will generally not be a controller as it only provides access to the data (and thereby transfers it to the next controller), except of course the app provider is acting on behalf of the social networking service provider [52]. In that case, it could be either that the app provider is a processor whereby the purpose is fully decided upon by the social network service provider, or, if they determine the means, e.g. the platform together, there could be a case of joint control.

Is your Facebook friend a controller? According to Helberger et al. [54], "users are in many cases the primary perpetrators of privacy infringements on social networking sites". The Regulation, like the Directive, does not apply in certain cases. One of these exemptions is referred to as the household exemption. Article 2 of GDPR, provides that the Regulation does not apply to the processing of personal data by a natural person in the course of a purely personal or household activity. In their opinion on social networking service providers, the Article 29 Working Party explained that the users of social networks will be data subjects in most cases, or might fall under the household exemption [53]. However, there can be cases in which the user might not be covered by the household exemption. This is especially the case if the user is acting on behalf of a company or association, or uses Facebook for commercial, political or charitable goals [53]. Also, the exemption does not apply if the information is made accessible to an indefinite number of people [55]. This is, for instance, the case when the profile information is accessible beyond self-selected contacts or is index-able by search engines [53]. In such cases, it is beyond the personal or household sphere and the user might be a controller [53].

If users qualify as controllers, they have to comply in principle with the data protection obligations. However, European data protection law was not intended for *amateur controllers* [54]. Some provisions can be applied without many problems. For instance, amateur controllers should also ask for their friends consent before they process their friends' data and in principle provide information to their friends such as, why they publish their data and where they do it [54]. However, in many cases "compliance with many provisions in data protection law probably exceeds the knowledge and experience of most users" [54]. Therefore, some provisions are not possible to be complied with by amateur controllers [54].

5.5 Transparency and information

While the principle of transparency was not as such mentioned in Directive 95/46/EC, it can be found very explicitly in the GDPR [18]. Article 5 specifically requires that data shall be processed in a transparent manner in relation to the data subject. Moreover, in Section 1 of GDPR, the rights of the data subject includes specific requirements on transparent information provision. Even Article 26 GDPR on joint controllers includes the requirement that the determination of the respective responsibilities should be done in a transparent manner. The Recitals 39 and 58 GDPR clarify that the principle of transparency requires that any information and communication relating to the processing of personal data needs to be concise, easily accessible, easily understandable, in clear and plain language and that visualisation can be used. In particular, information should be provided on the identity of the controller, the purposes of the processing and "further information to ensure fair and transparent processing in respect of the natural persons concerned and their right to obtain confirmation and communication of personal data concerning them which are being processed" [56]. However, users do not only need information when they are the data subjects: if they are controllers they will need information in order to comply with their duties as data controllers [54]. To be more specific, they will need practical information about the potential security and privacy risks on the social networking services and they will need legal information [54]. The Article 29 Working Party recommends social network providers to provide adequate warnings to users about the privacy risks related to themselves and to others when they upload information to the social networking service.

Even though in principle every controller must ensure the transparency of its processing, the social networking service provider faces higher expectations regarding information provisioning [54].

5.6 Consent and information

From a legal perspective, one of the main challenges of data protection attached to app permissions, as described above, is the fact that personal data processing may lack legitimacy. Article 6 of the GDPR [18] provides a limited number of legal grounds legitimising personal data processing, such as the consent of the data subject. Consent, as stated in Article 4 (11) GDPR, is defined as "any freely given, specific, informed and unambiguous indication for the data subject's wishes". As enshrined in Article 6 (1) (a), 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, which are specified in Article 7 GDPR, including that the controller needs to be able to demonstrate that consent has been given, the request for consent needs to be clearly distinguishable from other matters, understandable for the user and the user can withdraw consent at any time.

5.7 Data protection by design and default

An important change by the GDPR that is relevant to this paper is the introduction of data protection by design and default in Article 25 of the GDPR. This provision obliges the controller to take appropriate technical and organisational measures to implement data protection principles to ensure that by default only personal data which are necessary for each specific purpose of the processing are processed. Such measures should ensure that by default personal data are not made accessible to an indefinite number of persons without the individuals intervention. This provision is important, as the main problem in the Facebook case described in this paper is the possibility to access user A's data via user B since the default settings allow this data sharing while user A is not aware and never actively consented to it. Article 25 GDPR will require that the default settings are set in the most data protection friendly way and therefore the user typically will have to actively opt-in if she wants to give apps access to data of her friends.

Furthermore, the GDPR stipulates that appropriate technical and organisational measures should be implemented which meet in particular the principles of data protection by design and by default [57]. These measures could provide "transparency with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing, enabling the controller to create and improve security features" [57].

5.8 Profiling

Profiling is defined in Article 4 GDPR as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements" [18]. As stated in Sect. 1, app providers offering multiple apps could utilise *data* *fusion* to construct a more complete representation of users. Hence, data fusion implements certain sub-cases of profiling (e.g., directly extracting personal preferences and interests from Facebook attributes) while enabling other sub-cases of profiling by virtue of expanding the knowledge of the app provider (i.e., the data controller) on a given data subject. Regulating profiling plays a significant role in the GDPR, as it is mentioned in several articles and recitals. Article 22 GDPR declares that data subjects shall have the right not to be profiled if this profiling results in legal or other significant effects detrimental to them. The only exception relevant to our case is when the data subjects give explicitly consent. As per Recital 71 GDPR, in any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision. Also, such measure should not concern a child.

Article 13 GDPR moreover states that in the case of profiling, the controller is obliged to provide the data subject with information necessary to ensure fair and transparent processing, including the profiling logic and the significance and the envisaged consequences of such processing for the data subject. On top of this, Article 15 GDPR states that the data subjects have the right to access their personal data and the above information anytime (with a reasonable frequency) after the profiling took place. Article 21 GDPR also defines the right to object if profiling is done to enable direct marketing at any time and free of charge. Last, Article 35 GDPR makes a Data Protection Impact Assessment (DPIA) mandatory, in particular, if the processing is performed in order to take decisions relating to natural persons, and includes a systematic evaluation with respect to natural persons based on automated processing, in general, and profiling, in particular. Another focal case for a DPIA is if the processing is on a large-scale and involves special categories of data (as defined in Article 9 GDPR).

5.9 Is collateral information collection a risk for the protection of the personal data?

It is clear from the above that the new GDPR poses some well-defined requirements to the social networking service providers, i.e., Facebook in our case, with regards to the protection of users' personal data. Several points, mostly concerning the legal obligation of users as amateur data controllers, on the one hand, are murky at best. On the other hand the problem clearly converges when the personal data of users can be transferred from one controller to the other, as from Facebook to an app provider, without notifying the users (i.e., Article 5 GDPR) and without obtaining consent from the user (i.e., Articles 6, 7 GDPR).

With respect to the obligations of the data controller and processor to transparency, app providers can become data controllers and processors of a user's personal data, without the user becoming aware of such data transfer (i.e., *collateral information collection*). It should be noted that a Facebook user and their friends have insufficient information on both the amount of data that will be collected and the purposes their data will be used for by an app and its provider [58]. In other words, data collection and processing go far beyond the user's and their friends' legitimate expectations, and interferes with the principle of transparency as per Article 5 GDPR.

Furthermore, considering the obligations of data controller and processor with regard to consent, third party apps on Facebook may collect and proceed to process a user's personal data without prior, informed and direct consent from the user themselves, that is operating exclusively based on the consent provided by one of her friends. In other words, consent can be given only by the user who installed the app (i.e., friend) and not by the data subject (i.e., user) whose data is collected and processed. One might say that Facebook app settings give the users control over their personal data to be handed to apps by their friends via ticking the appropriate checkboxes in the sub-menu "Apps Others Use". Therefore, one could claim that consent has been theoretically given. However, it should not be considered as valid as it is not informed. In fact, owing to the default, pro-sharing privacy setting on Facebook (see Fig. 17), users are generally unaware of the fact that they have to uncheck the boxes (actively opt out) in order to prevent such data processing [19, ?]. This also goes against the concept of privacy by default. It is worth mentioning that in a relevant case in the U.S., the Federal Trade Commission required that such apps cannot imply consent, but this should be rather affirmatively expressed by users [26].

Finally, *profiling* is a separate concern on its own behalf; nevertheless, data fusion (a technique partly constituting and greatly enabling profiling) has an amplifying effect on collateral information collection (see Tab. 1). Moreover, this amplifying effect is mutual. Therefore, data protection obligations of app providers offering multiple apps are even more pronounced. As for all automated processing, especially if the outcome of automated processing are decisions relating to data subjects (natural persons) by the data processor, details about how and why the processing is done and its potential effects on the data subject should be given to the data subject. We have no examples of this happening in the case of app providers or Facebook: we are not aware of any systematic means (potentially with the technical help of the Facebook App Platform) where this information flow can be realised. To sum up, data fusion potentially resulting in profiling may happen without any kind of transparency and consent: the data subject is simply not aware of such processing taking place. Another, especially intriguing aspect of profiling is the obligation for the profiler to conduct a DPIA if the decision based on profiling affect data subjects significantly and/or profiling is large-scale and involves special categories of data. Given the popularity of certain apps and the type of attributes collected a DPIA could be obligatory. Finally, the Article 29 Working Party identifies any "datasets that have been matched or combined, for example originating from two or more data processing operations performed for different purposes and/or by different data controllers in a way that would exceed the reasonable expectations of the data subject" likely to result in a high risk; therefore any operations resulting in such datasets should be covered by a DPIA [59].

6 Solutions to collateral information collection

In this section, we investigate the research question: *how can we mitigate collateral information collection?* To answer this question, we propose a privacy dashboard extension that aims at improving transparency and helping users to make informed decisions about such information collection, while introducing a scoring computation for evaluating the collateral information collection over the apps and app providers. Moreover, we discuss alternative solutions focusing on notification and access control mechanisms, cryptographic solutions and on app auditing and in particular on a Data Protection Impact Assessment driven by the GDPR.

6.1 Enhancing transparency with dashboard

Transparency Enhancing Technologies (TET) [60–62], can support decision-making processes by helping users to understand the problem and foster users' control by assisting them to make informed decisions [63, 64]. For instance, to take action about the amount of information that may be affected by collateral information collection. The increase of awareness on personal data collection is in line with the legal principle of data protection by design and default (Articles 39 and 78 of the GDPR [18]). Furthermore, driven by our questionnaire responses, there are shreds of evidence that participants are not only concerned about the collateral information collection, but they also want to be notified and restrict access to their personal data on Facebook. They also consider removing the apps that can cause the collateral information collection.

In order to design a TET with regard to collateral information collection, we need two main ingredients: 1) a quantitative measure characterising the added exposure of profile attributes, and 2) a usable way to present this information to the user. Fortunately, established methods for constructing both ingredients are available in the literature [65, 66]. However, both the quantitative metrics such as a privacy score and usable presentation schemes such as dashboards are not yet tailored to the specifics of collateral information collection. In the following section we demonstrate how to integrate our quantification mechanisms for collateral information collection into these frameworks.

Privacy Score In a nutshell, we describe the premises to compute the Privacy Score ($\mathtt{PS}$) for an app (i.e., A_j) and an app provider (i.e., P_j), as an indicator of the level of collateral information collection. The higher the $\mathtt{PS}$ the more significant the threat for a user, meaning that more personal data of the profile of a user (i.e., u) can be collected by an A_j and P_j. To compute the $\mathtt{PS}$, Liu and Terzi [65] proposed a formula consisting of the product of the visibility of a user's profile attribute (i.e., a_i) in an OSN graph with its sensitivity which represents its perceived importance (weight). Our $\mathtt{PS}$ computation formula represents the case of collateral information collection from A_js and P_js in an OSN.

Sensitivity We denote by $\mathcal{S}$ the set of different attribute weights in u's profile, i.e., $\mathcal{S} = \{\sigma_1, \ldots, \sigma_n\}$. We consider σ_i for $1 \leq i \leq n$ the number of different attribute weights in u's profile where $\{\sigma_i \in \mathbb{Q} | 0 \leq \sigma_i \leq 1]\}$, i.e., a_i is more sensitive than $a_{i'}$ iff $\sigma_i > \sigma_{i'}$, therefore,

$$\sigma_i = \begin{cases} 1 & \text{if } \textit{very sensitive}, \\ 0 & \text{if } \textit{not sensitive}, \\ 0 < \sigma_i < 1 & \text{if } \textit{in between}, \end{cases} \tag{15}$$

for $1 \leq i \leq n$.

Privacy Score of u for A_js We denote by $\mathtt{PS}_A^{u,\mathsf{F}^u}$ the Privacy Score ($\mathtt{PS}$) of u for all A_js, where $\{\mathtt{PS}_A^{u,\mathsf{F}^u} \in \mathbb{Q} |\ 0 \leq \mathtt{PS}_A^{u,\mathsf{F}^u} \leq 1\}$, when A_js can collect attributes from u's profile in F^u, i.e.,

$$\mathtt{PS}_A^{u,\mathsf{F}^u} = \frac{\sum_{i=1}^{n} (\sigma_i \bullet \vec{\mathit{\Pi}}_{A_j}^{u,\mathsf{F}^u})}{|\mathcal{T}|} = \frac{\sum_{i=1}^{n} (\sigma_i \bullet \bigvee_{A \in \mathsf{L}^{u,\mathsf{F}^u}} (\vec{A}_j^{u,\mathsf{F}^u}[i] \wedge \vec{C}_{A_j}^{u,\mathsf{F}^u}[i]))}{|\mathcal{T}|} \ . \tag{16}$$

- Remark: $a \ \bullet \ b = \begin{cases} 1 & \text{if } a = b = 1, \\ a \times b & \textit{Otherwise}, \end{cases}$

The complexity of evaluating Eqn. (16) for all f in F^u is $\mathcal{O}(n^5 \times |\mathsf{L}^{u,\mathsf{F}^u}|)$.

Privacy Score of u for all P_j We denote by $\mathtt{PS}_P^{u,\mathsf{F}^u}$ the Privacy Score ($\mathtt{PS}$) of u for all P_js, where $\{\mathtt{PS}_P^{u,\mathsf{F}^u} \in \mathbb{Q} | 0 \leq \mathtt{PS}_P^{u,\mathsf{F}^u} \leq 1]\}$, when P_js can collect attributes from the u's profile in F^u, i.e.,

$$\mathtt{PS}_P^{u} = \frac{\sum_{i=1}^{n} \left(\sigma_i \bullet \vec{\mathit{\Pi}}_{P_j}^{u,\mathsf{F}^u}\right)}{|\mathcal{T}|} = \frac{\sum_{i=1}^{n} \left(\sigma_i \bullet \bigvee_{P \in \mathcal{AP}^{u,\mathsf{F}^u}} \left(\bigvee_{A_j \in \mathsf{P}_j^{u,\mathsf{F}^u}} (\vec{A}_j^{u,\mathsf{F}^u}[i] \wedge \vec{C}_{A_j}^{u,\mathsf{F}^u}[i])\right)\right)}{|\mathcal{T}|} \ . \tag{17}$$

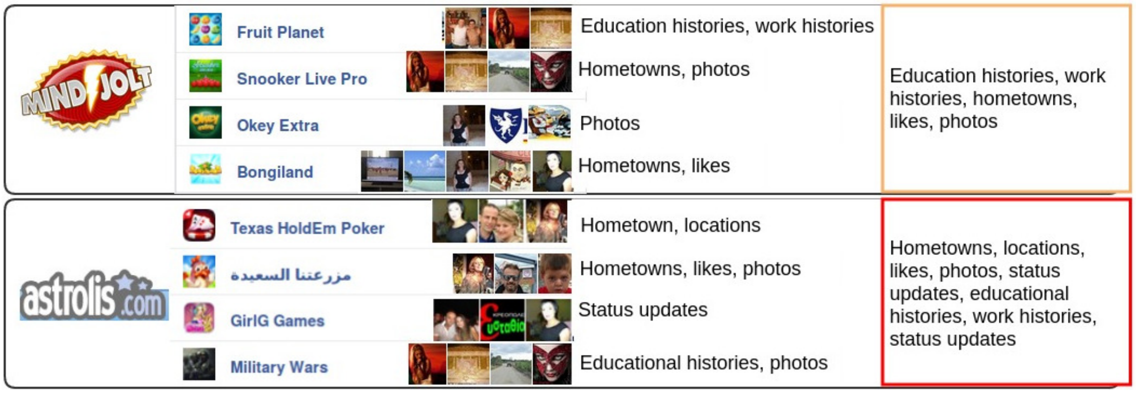
The complexity of evaluating Eqn. (17) for all f in F^u is $\mathcal{O}(n^5 \times |\mathsf{P}_j^{u,\mathsf{F}^u}| \times |\mathcal{AP}^{u,\mathsf{F}^u}|)$.

Quantifying sensitivity One way to measure and quantify the perceived sensitivity of users is by running survey studies. Minkus et al. [35] measured the sensitivity attributed to different privacy settings by Facebook users from a survey of 189 participants. They estimated how users perceive the importance of each privacy setting. They identified variations of how sensitive each privacy permission is. For instance, the permission "What personal information goes into apps others use?” is considered more sensitive (i.e., $\sigma_i = 2.82$) than “Who can send you friend requests?” (i.e., $\sigma_i = 1.09$). Their sensitivity estimations, however, is slightly different from quantifying the sensitivity of the attributes in our case. Although there is a relation between privacy settings and attributes, there are no studies to identify the correlation between the perceived sensitivity from a privacy setting to the set of attributes. Moreover, there are no studies yet for measuring the sensitivity of the attributes collected by app / app provider and neither for *collateral information collection*. Nevertheless, Minkus et al. [35] provide us with sensible default sensitivity values; these can be then updated by running a new survey (this is left for future work). Privacy-conscious users can then adjust the sensitivity values on demand and dynamically.

A different promising direction for quantifying the sensitivity of attributes is by demonstrating which attributes an app can collect before, during or after the installation. This can be achieved through a privacy dashboard demonstrating all possible attributes an app can collect during app installation with the help of authorisation dialogues [28].

6.2 Damage Control: Privacy Dashboard

To enhance transparency, several tools have been proposed [67, 68]; among these wa privacy dashboards is a well established instrument [66, 69–71]. For privacy dashboards, several designs have been demonstrated [69, 72], with some of them specifically tailored for Facebook [66, 70]. Within our work, we propose components that can be used in existing dashboard designs [66, 69, 70], aiming to extend and enhance the minimisation of data disclosure by the users for the *collateral information collection* (see Fig. 20 for an initial user interface design).



MIND JOLT	Fruit Planet	Education histories, work histories	Education histories, work histories, hometowns, likes, photos
	Snooker Live Pro	Hometowns, photos	
	Okey Extra	Photos	
	Bongiland	Hometowns, likes	
astrolis.com	Texas HoldEm Poker	Hometown, locations	Hometowns, locations, likes, photos, status updates, educational histories, work histories, status updates
	مزرعتنا السعيدة	Hometowns, likes, photos	
	GirlG Games	Status updates	
	Military Wars	Educational histories, photos	

Fig. 20: Privacy dashboard: user interface concept

Requirements Transparency should be reflected by visualising the personal data that can be disclosed to apps and app providers by the friends of a user on Facebook. Specifically, the dashboard extension should demonstrate which app is collecting which of the personal data of a user and through which friend. Moreover, the dashboard extension should demonstrate the impact of *data fusion* via multiple apps offered by a single app provider. The dashboard extension should provide all the quantitative representations in a readable and understandable format. Orthogonal to these issues, a user should be able to adjust the weights of sensitivity for her personal profile attributes, as this should be reflected in the computation of the privacy score. However, she should also be able to use the dashboard right away with reasonable default weights.

Proposed design Technically speaking, the proposed dashboard tools illustrate how the data disclosure of a user takes place through the acquisition of the user's personal data via apps (and respective app providers) when installed by their Facebook friends. It displays the nature and proportion of the user's personal data that can be collected by apps and, more importantly, by app providers. To better help users to understand, the privacy metrics are visualised and the level of potential information dissemination is represented by colour scale to indicate the level of collateral information collection [73].

From our user opinion study, we have concluded that Facebook users are more concerned about certain types of personal information such as photos, videos, location, and relationships [35]. Our dashboard can accommodate the visualisation of *data fusion* and the degree of collateral information collection by assigning different weights (sensitivity) to each attribute in the user's profile [35]. These weights can then be manually fine-tuned by the user for an app and app provider (see Fig 20).

Dashboard technical limitations on Facebook Implementing a privacy dashboard on Facebook has several limitations. To identify which of the installed apps for a user enable the collateral information collection, the list of apps needs to be retrieved from Facebook. Currently, as the Facebook API does not allow for inspecting the installed apps of a user, one way is by scraping the app center page. However, Facebook has designed their web interface to resist high volume content retrieval. It is possible for a skilled developer to circumvent this protection and to collect and identify which installed apps enable the collateral information collection for a user; this can be automated but the task is far from trivial (e.g., using Selenium [74]). The list of apps can be retrieved while a user is logged in to their Facebook account. To identify which friends have installed also similar apps, a scraping operation needs to be performed for each page of an app. To identify apps that enable collateral information collection and have been installed only by friends of a user, it needs all possible pages of apps to be scraped and downloaded from the Facebook app center; assuming that a developer has a compilation of the majority of popular apps that enable collateral information collection. Making the problem more complex, Facebook is susceptible to constant changes: it updates the interface, the permissions, and the privacy settings regularly, making the development of a dashboard cumbersome and with a need for continuous modification.

A detailed design and implementation of the dashboard remains the next step in our future work. Additional information such as claimed purpose of collection by the apps can be added to the dashboard. Moreover, further functionality can be added to the dashboard such as leading the users from the dashboard to uninstall the app, which this would strengthen compliance with the GDPR [18].

6.3 Alternative countermeasures

We shown in this work that transparency-enhancing solutions can support notification and facilitate direct consent in order to comply with the EU legal regulation [18]. Notification can be enhanced using *privacy nudges* [4, 75] while direct consent with *privacy authorization dialogues* for apps [28, 76], and *access control mechanisms* [77–79] tailored to collateral information collection. Considering *privacy nudges*, such an enhanced notification can be provided in the following manner: a notice about the additional attributes that the app provider will collect, a timer interface before the Facebook users clicks the "app installation", or even a "sentimental nudge" about collateral information collection as bad practice. Furthermore, as standalone consent mechanisms, *access controls* and *authentication dialogues* should be designed in the following manner: to provide permissions for collateral information collection for both the friends and the users, and to explicitly highlight such permissions while Facebook users install an app which enables such information collection. Moreover, permissions should have a direct, one-to-one mapping to the attributes of the profile of a Facebook user, which is not the case for the $user\_friends$ permission.

Additional to transparency solutions, other countermeasures to collateral information collection can use cryptographic tools to minimise or counterfeit such indirect information collection. Such solutions can help to better control the dissemination of the information of Facebook users, providing strong mathematical guarantees for the users' privacy. For instance, flyByNight [80] and Scramble! [81] are proposing cryptographic schemes to ensure confidentiality and integrity of messages exchanged among Facebook users. Extending the functionality of such solutions, cryptographic tools such as Multi-Party Computation (MPC) [82] can be used by a user, to allow access only to selected apps and app providers for their data that are shared through their friends. Encryption solutions are in the contrary to business models of Facebook and other OSNs and are commonly detected and blocked from the system, though.

Coming from Facebook itself, v2.x of the API has the potential to decrease both the likelihood and the impact of collateral information collection. Apps using these API versions can only see a shortened friend list for the installing user: those who have also installed the app *and* granted the $u$ser_$f$riends permission. The app can gather profile attributes only from users appearing in this friend list. While this is an API change we applaud, it does not constitute a total solution. First, although it may considerably reduce the risk of getting exposed to collateral information collection, those affected may still suffer from *exclusive* collateral information collection: there could be some attributes exclusively collected through friends and not directly. Second, since there are a plethora of affected apps, the beneficial effect of the API change is not so pronounced for a user: the probability of installing at least one app enabling exclusive collateral information collection is still high. Third, this API change does not have any effect on multi-app data fusion. Illustrating this through Table 1, the new API makes sure that the top cells are nonempty sets, should the lower cells be nonempty sets as well.

Finally, a Data Protection Impact Assessment can be viewed more than just an obligation for high-risk data processing operation: it can be also viewed as a legally inspired countermeasure against privacy threats with regard to collateral information collection. As DPIA is a process for building and demonstrating compliance to GDPR, it can provide meaningful information both to data protection authorities and concerned data subjects, thereby promoting transparency and also trust. Moreover, for Facebook apps and app providers, a qualitative DPIA can be augmented by a more quantitative privacy risk assessment based on our proposed *Privacy Score*, which captures all traditional aspects of risk computation (see Sect. 6.1). Such an enhanced DPIA promotes comparability over different apps and app providers, and conveys more tangible information towards users.

7 Related work

This section describes the related work on privacy issues that can arise from the use of apps in the Facebook ecosystem. It describes the existing work including the relevant user surveys, the app adoption models and network topologies, the computation of the privacy score and finally the dashboard solutions.

Interdependent privacy and the app-related privacy issues in OSNs Chaabane et al. [83] showed that apps can have tracking capabilities and disseminate the collected information to "fourth party" organisations [84]. Similarly, Huber et al. [17] developed an automatic evaluation tool, AppInspect [9], and demonstrated that personal identifiable information of Facebook users was leaked by a large set of Facebook apps to analytic and advertisement companies [17]. Frank et al. [85] showed that low-reputation apps often deviate from the permission request patterns, while Chia et al. [75] showed that certain apps collect more information than necessary. Moreover, Wang et al. [5] identified third-party app bad practices for privacy notice and consent on Facebook, while studying $1,800$ most popular third-party apps. For instance, using the Facebook API, an app can overwrite the users' and their friends' privacy settings in a calendar, displaying the birthdays of the user and her friends while the privacy settings are set to "Only me" in both sides.

Adding to these concerns, Thomas et al. [86] examined the lack of privacy controls over the shared content of friends on Facebook. Alice, Bob's friend, can share information and unintentionally violate Bob's privacy leading to *privacy conflicts*. Using a formal representation of the privacy conflict concept, they estimated the aggregated information of Bob under the presence of privacy conflicts with Alice which, can lead to uncover sensitive information about Bob. To the best of our knowledge, Weng et al. [5] were the first to report the fact that Facebook API allows third-party apps to collect users' profile information through their friends on Facebook, and Biczók and Chia [6] were the first to introduce the notion of *interdependent privacy*. From their work, Biczók and Chia modelled the impact of the *interdependent privacy* problem performing a game theoretic study (2-player, 1-app). Sharing a user's information without her direct consent can lead to the emergence of externalities; positive externalities e.g., personalised experience for social-networking apps and negative externalities e.g., exposed profile items. Pu and Grossklags [21], extended the work of Biczók and Chia [6] and developed a formula to simulate the app adoption behaviour of users and their friends and the monetary impact of the *interdependent privacy* problem [87]. Currently, Harkous and Aberer [88] extended the *interdependent privacy* problem to cloud services ecosystem such as Dropbox and Google Drive. They studied the privacy loss of a user, when the user and her collaborators grant access to a shared file for a new cloud vendor.

Facebook survey studies Liu and Gummadi [3] studied the discrepancy between the desired privacy in contrast to reality on Facebook. Through a Facebook app, they collected responses of 200 participants and they compared them with their actual privacy settings each user had. They identified that almost 50% of the content was shared with the default privacy settings exposing their information to all Facebook users, in contrast to 20% of the users which had the actual desired setting. Our survey study extends the work of Wang et al. [5]. While they identified third-party app bad practices for privacy notice and consent on Facebook, they proposed enhanced app authorisation dialogues. To validate their authorisation designs they performed a survey study. Interestingly, a minor set of participants reported the unfair third-party app practices that can "gather information about ones friends" without their notice or consent. Our work goes beyond such a minor set of users' opinion (i.e., four participants) in an open question. We have designed a questionnaire and constructed a study of 114 participants to investigate in more detail the users' opinions on the collateral information collection problem.

Network topologies and app adoption models Regarding (Facebook) network topologies, there are three relevant models, which should be considered. The Barabási-Albert model [46] (`BA`) generates networks with a power-law degree distribution. Several studies about real-world social networks [45, 43] have reported that their degree distributions follow a power law. The Watts-Strogatz model [47] (`WS`) generates small world networks with high clustering and small diameter, properties which have also been found in many real world OSNs [45, 44]. Note that Watts-Strogatz is able to generate graphs with high clustering without exhibiting a power-law degree distribution, whereas Barabási-Albert presents a power-law degree distribution without high clustering. The Erdős-Rényi model [42] (`ER`) generates a uniform degree distribution and is therefore used as baseline. Regarding app adoption, the baseline is a uniform model (`unif`), where all users install an app with the same probability, and each installation is independent from other installations and the underlying network topology. However, owing to local network effects prevalent in OSNs [48], a preferential model (`prop`) is more realistic, where the probability of a user installing an app is proportional to the number of friends that have already installed the app.

Quantifying the collateral information collection in OSNs Maximilien et al. [89] proposed a formula to estimate the privacy risk for a user. They computed the privacy risk as the product of *sensitivity* and *visibility* of personal data. Liu and Terzi [65] extended this work of [89] and proposed a framework for computing the privacy risk using a probabilistic model based on the Item Response Theory (IRT). Although, IRT presents an interesting approach to compute the *sensitivity* of the user's personal data, *visibility* is not properly addressed. Moreover, Sánchez and Viejo [90] 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 [91]. Minkus et al. [35] estimated the *sensitivity* and *visibility* of the privacy settings based on a survey of 189 participants. Finally, Nepali and Wang [92] proposed a privacy index to evaluate the inference attacks as described by Sweeney [93], while Ngoc et al. [94] introduced a metric to estimate the potential leakage of private information from public posts in OSNs. To the best of our knowledge, there is no related work quantifying the collateral information collection.

Transparency enhancing technologies and privacy dashboards For enhancing transparency, a wide selection of tools are listed in the surveys of Hedbom [67] and Janic et al. [68]. Along with privacy icons [95] and privacy nudges [96], privacy dashboards [66, 70] are a well studied concept for notification and enhancing awareness of users. For instance, Data Track privacy dashboard [71] was a design of the successful European projects PRIME (FP6) and PrimeLife (FP7) [62].

Concerning privacy dashboards, several designs have been proposed to enhance transparency in OSNs while users can disclose information to their friends and other recipients [4, 75]. However, none of these dashboard solutions are tailored towards collateral information collection. Bier et al. [69] proposed and implemented a privacy dashboard namely PrivacyInsight. PrivacyInsight has been designed under a subset of the legal and usability requirements defined by the GDPR and the ISO standard 9241-11 [97]. However, the authors do not investigate a scenario with either joint control of personal data or multi-app *data fusion.* (see Sect. 5). Talukder et al. [66] proposed a privacy dashboard called Privometer that merely aims to measure the amount of sensitive information that can be leaked from a single user's profile on Facebook. Buchmann et al. [70] designed the Personal Information Dashboard (PID), aiming to provide transparency representing the information that users disclose across multiple OSN domains. Although neither one of the solutions treats the analysed problem directly, their designs can be extended to implement an enhanced privacy dashboard, and include the collateral information collection case coupled with multi-app *data fusion.* Such an information collection of apps can be given by the users or through their friends, while a dashboard can foster better notification and consent by both the users and their friends; a friend might be willing to uninstall an app if it enables collateral information collection (see Sect. 2).

Alternative countermeasures to dashboard Wang et al. [96] proposed techniques to nudge users helping them to avoid online disclosure on Facebook that they may regret later. To investigate the efficiency of privacy nudges, they developed a Facebook app and complemented the study with a survey questionnaire of 21 participants. They identified that "Stop and Think" nudge helped participants to better avoid regrettable postings. The "Pay attention to the audience" nudge helped them to better identify the audience and be more cautions about their posts. Whereas, the "Content feedback" nudge was perceived as needles and not very positive nudge, as participants were feeling being judged by the tool. Paul et al. [98] studied an enhanced interface for the Facebook privacy settings. They proposed a coloured representation of the privacy settings (*C4PS - Colours for Privacy Settings*), to demonstrate the visibility of the profile items of users, such as photos. To verify their assumptions of the effectiveness of such a design, they created a mock-up of the Facebook privacy settings and performed a survey were they gathered responses of 40 participants (students). They identified that such an enhancement, can help users to better employ the desired privacy settings.

Moreover, FSEO [99], FaceCloak [100] and NOYB [101] are privacy schemes focused on OSNs and particularly on Facebook. Their main goal is to achieve privacy by providing fake information, considering both the app provider and the OSN user as adversaries. Scramble! [81] proposes an *access control mechanism* over a user's data, making the use of encryption techniques. According to this model, authorised users have partial access to the data, depending on the access control lists. flyByNight [80] is another privacy solution for OSNs, that makes use of symmetric-key cryptography. This approach tries to overcome the limitations of Facebook by introducing a privacy platform through a proxy server. FaceVPSN [102] introduces a distributed platform for storing information, providing fake information to the OSN. Furthermore, there exist other solutions that propose privacy-friendly architectures such as, Safebook [103], EASiER [104] and more [105–107].

8 Conclusion and future work

We conclude that collateral information collection poses a privacy problem for Facebook users. In a nutshell, we presented a multi-faceted study concerning the collateral information collection caused by applications installed by the friends of a user on Facebook. Our main findings were the following. Using a questionnaire, we show that the vast majority of our participants are very concerned and would like proper notification and control mechanisms regarding collateral information collection. Also, they would like to restrict the applications accessing the profile data of both the user and their friends, when their friends' enable the collateral information collection and vice versa. Running simulations for various network topologies and application adoption models, we quantified the probability that a Facebook user can be affected by the collateral information collection. Assuming an application with more than 1M users (such as TripAdvisor), there is 80% probability for this problem to appear. Employing real data from the Facebook third-party application ecosystem, we quantified the significance of collateral information collection by computing the proportion of attributes collected by applications installed by the friends of a user. Based on popular applications, we identified that almost half of the 207 applications that enable the collateral information collection collected the photos of the friends of a user, with location and work history being the second and third most popular attributes. Through the prism of the new General Data Protection Regulation (GDPR), we investigate and conclude that collateral information collection is likely to result in a risk for the protection of the personal data of the Facebook users. The cause is the lack of notification and consent, the non-existence of privacy by default with regard to Facebook privacy settings, and the amplifying effect of *data fusion* and, potentially, profiling.

To mitigate the collateral information collection, we proposed solutions aiming at enhancing transparency and increasing control of the information dissemination of Facebook users through third-party applications. In particular and for enhancing transparency, we proposed a *privacy dashboard extension* to enhance the existing privacy dashboard designs when collateral information collection needs to be instantiated and quantified. Such an enhancement can also help to empower the users' decisions and enforce restrictive actions if necessary. Moreover, we discuss alternative solutions focusing on notification and access control mechanisms, cryptographic countermeasures and application auditing and in particular on Data Protection Impact Assessment (DPIA) driven by the GDPR.

To the best of our knowledge, our work is the first one to report the potential user *profiling* threat that could be posed by application providers. They can gain access to complementary subsets of personal data from user profiles by offering multiple applications. Moreover, our study can serve as a guide for future applications, software platforms, or permission systems such that, the interdependent aspects of privacy can be directly taken into consideration in the design phase.

Future work We identified four directions for potential future work on collateral information collection. First, conducting a fully rounded quantitative survey would pay immediate dividends by (1) extending the sample size and demographic coverage compared to our questionnaire, (2) collecting answers from social networking platforms other than Facebook and (3) quantifying the perceived sensitivity of users concerning various profile attributes. Second, we are particularly interested in using DPIA as a countermeasure to collateral information collection and other social and mobile applications related to privacy threats, especially by augmenting the current DPIA state-of-the-art with a quantitative privacy risk assessment. It is likely that application providers (in particular those with multiple applications and/or many users) are obliged to perform a DPIA, both for individual applications and for their complete application suite, to demonstrate compliance to the GDPR. Third, the *data fusion* aspect of application providers offering multiple applications may have another dimension: when application providers offer applications on multiple platforms, such as Facebook, Android or Google Drive. We would like to investigate, whether such cross-platform data fusion (and thus, profiling) is happening and/or at least feasible. For example, imagine if social profile items from Facebook could be combined with the precise location of users from a mobile app. Furthermore, Facebook Login, a widely used single sign-on mechanism, may transfer the collateral information collection issue to other websites and services. Fourth and last, an interesting but unexplored aspect of collateral information collection is whether such data gathering could be of no harm to users' privacy, or even can be beneficial to users under some special circumstances.

Acknowledgements

We notably want to thank Markus Huber and SBA Research Center for providing us with the necessary material for our study. We also want to express our gratitude to Faruk Gölogłu for providing us feedback for the mathematical model and Seda Gürses for giving us valuable input to our questionnaire study. Moreover, we would like to thank all the anonymous reviewers who helped for better shaping the idea and the quality of this text. This work was supported in part by the Research Council KU Leuven (C16/15/058), the Spanish Government (TIN2014-55243-P and FPU-AP2010-0078), the Catalan Government (AGAUR 2014SGR-691) and by Microsoft Research through its PhD Scholarship Programme. G. Biczók has been supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

1. Statista, Number of daily active Facebook users worldwide as of 2nd quarter 2017 (in millions), https://www.statista.com/statistics/346167/facebook-global-dau/, accessed Aug, 2017.
2. A. Acquisti, R. Gross, Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook, in: Privacy Enhancing Technologies, 6th International Workshop, PET 2006, Cambridge, UK, June 28-30, 2006, Revised Selected Papers, 2006, pp. 36–58. doi:10.1007/11957454_3.
URL `https://doi.org/10.1007/11957454\_3`
3. Y. Liu, P. K. Gummadi, B. Krishnamurthy, A. Mislove, Analyzing facebook privacy settings: user expectations vs. reality, in: Proceedings of the 11th ACM SIGCOMM Internet Measurement Conference, IMC '11, Berlin, Germany, November 2-, 2011, 2011, pp. 61–70. doi:10.1145/2068816.2068823.
URL `http://doi.acm.org/10.1145/2068816.2068823`
4. Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, L. F. Cranor, "I regretted the minute I pressed share": a qualitative study of regrets on Facebook, in: Symposium On Usable Privacy and Security, SOUPS'11, Pittsburgh, PA, USA - July 20 - 22, 2011, 2011, p. 10. doi:10.1145/2078827.2078841.
URL `http://doi.acm.org/10.1145/2078827.2078841`
5. N. Wang, H. Xu, J. Grossklags, Third-party apps on Facebook: privacy and the illusion of control, in: Proceedings of the 5th ACM symposium on computer human interaction for management of information technology, ACM, 2011, p. 4.
6. G. Biczók, P. H. Chia, Interdependent Privacy: Let Me Share Your Data, in: Financial Cryptography and Data Security - 17th International Conference, FC 2013, Okinawa, Japan, April 1-5, 2013, Revised Selected Papers, 2013, pp. 338–353. doi:10.1007/978-3-642-39884-1_29.
URL `https://doi.org/10.1007/978-3-642-39884-1\_29`

7. I. Symeonidis, F. Shirazi, G. Biczók, C. Pérez-Solà, B. Preneel, Collateral Damage of Facebook Apps: Friends, Providers, and Privacy Interdependence, in: ICT Systems Security and Privacy Protection - 31st IFIP TC 11 International Conference, SEC 2016, Ghent, Belgium, May 30 - June 1, 2016, Proceedings, 2016, pp. 194–208. doi:10.1007/978-3-319-33630-5_14.
URL `https://doi.org/10.1007/978-3-319-33630-5\_14`
8. Facebook, Facebook privacy settings and 3rd parties, https://developers.facebook.com/docs/graph-api/reference/v2.10/user, accessed Aug, 2017.
9. AppInspect, AppInspect: A framework for automated security and privacy analysis of OSN application ecosystems, http://ai.sba-research.org/, accessed Aug, 2017.
10. Facebook, Criminal case, https://www.facebook.com/CriminalCaseGame/, accessed Sep, 2017.
11. Facebook, Candy crush saga, https://www.facebook.com/games/candycrush, accessed Sep, 2017.
12. Rovio, Angry birds, https://www.facebook.com/angrybirds/, accessed Sep, 2017.
13. TripAdvisor, Tripadvisor, https://www.facebook.com/games/tripadvisor, accessed Jan, 2016.
14. MailOnline, TripAdvisor links to Facebook to show reviews from your friends... and their friends too, http://www.dailymail.co.uk/travel/article-2128713/TripAdvisor-links-Facebook-reviews-friends.html, accessed Nov, 2017.
15. Facebook, Add Facebook Login to Your App or Website, https://developers.facebook.com/docs/facebook-login/, accessed Oct, 2017.
16. I. E. T. F. (IETF), The OAuth 2.0 Authorization Framework, https://www.rfc-editor.org/rfc/pdfrfc/rfc6749.txt.pdf, accessed Oct, 2017.
17. M. Huber, M. Mulazzani, S. Schrittwieser, 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, 2013, pp. 143–154. doi:10.1145/2512938.2512942.
URL `http://doi.acm.org/10.1145/2512938.2512942`
18. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L119, 4.5.2016, http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679, accessed Aug, 2017.
19. Federal Trade Commission, FTC and Facebook agreement for 3rd parties wrt privacy settings, http://www.ftc.gov/sites/default/files/ documents/cases/2011/11/111129facebookagree.pdf (Accessed Aug, 2017).
20. W. Enck, P. Gilbert, B. Chun, L. P. Cox, J. Jung, P. D. McDaniel, A. Sheth, TaintDroid: an information flow tracking system for real-time privacy monitoring on smartphones, Commun. ACM 57 (3) (2014) 99–106. doi:10.1145/2494522.
URL `http://doi.acm.org/10.1145/2494522`
21. Y. Pu, J. Grossklags, An Economic Model and Simulation Results of App Adoption Decisions on Networks with Interdependent Privacy Consequences, in: Decision and Game Theory for Security - 5th International Conference, GameSec 2014, Los Angeles, CA, USA, November 6-7, 2014. Proceedings, 2014, pp. 246–265. doi:10.1007/978-3-319-12601-2_14.
URL `https://doi.org/10.1007/978-3-319-12601-2\_14`
22. Y. Pu, J. Grossklags, Towards a Model on the Factors Influencing Social App Users' Valuation of Interdependent Privacy, PoPETs 2016 (2) (2016) 61–81.
URL `http://www.degruyter.com/view/j/popets.2015.2016.issue-2/popets-2016-0005/popets-2016-0005.xml`
23. D. L. Hall, J. Llinas, An introduction to multisensor data fusion, Proceedings of the IEEE 85 (1) (1997) 6–23. doi:10.1109/5.554205.
24. D. Jobber, Principles and Practice of Marketing, Principles and Practice of Marketing, McGraw-Hill, 2010.
URL `https://books.google.be/books?id=is67PwAACAAJ`
25. E. Bloemendaal, I. M. H. Øveråsen, Interdependent Privacy on Facebook, https://www.dropbox.com/s/ci9ur2231ykle6i, accessed Aug, 2017.
26. Federal Trade Commission, Facebook, Inc., http://www.ftc.gov/enforcement/cases-proceedings/092-3184/facebook-inc, accessed Aug, 2017.

27. J. Buchmann, R. Capurro, M. Löw, G. Müller, A. Pretschner, A. Roßnagel, M. Waidner, K.-I. Eiermann, M. Eldred, F. Kelbert, et al., Internet Privacy: Options for adequate realisation, Springer Science & Business Media, 2014.
28. N. Wang, J. Grossklags, H. Xu, An online experiment of privacy authorization dialogues for social applications, in: Computer Supported Cooperative Work, CSCW 2013, San Antonio, TX, USA, February 23-27, 2013, 2013, pp. 261–272. doi:10.1145/2441776.2441807.
URL http://doi.acm.org/10.1145/2441776.2441807
29. B. Krishnamurthy, I know what you will do next summer, Computer Communication Review 40 (5) (2010) 65–70. doi:10.1145/1880153.1880164.
URL http://doi.acm.org/10.1145/1880153.1880164
30. O. O'Neill, Some limits of informed consent, Journal of Medical Ethics 29 (1) (2003) 4–7. arXiv:http://jme.bmj.com/content/29/1/4.full.pdf, doi:10.1136/jme.29.1.4.
URL http://jme.bmj.com/content/29/1/4
31. D. Boyd, E. Hargittai, Facebook privacy settings: Who cares?, First Monday 15 (8).
URL http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3086
32. B. Debatin, J. P. Lovejoy, A. Horn, B. N. Hughes, Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences, J. Computer-Mediated Communication 15 (1) (2009) 83–108. doi:10.1111/j.1083-6101.2009.01494.x.
URL https://doi.org/10.1111/j.1083-6101.2009.01494.x
33. S. Kokolakis, Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon, Computers & Security 64 (2017) 122–134. doi:10.1016/j.cose.2015.07.002.
URL https://doi.org/10.1016/j.cose.2015.07.002
34. M. Madejski, M. L. Johnson, S. M. Bellovin, A study of privacy settings errors in an online social network, in: Tenth Annual IEEE International Conference on Pervasive Computing and Communications, PerCom 2012, March 19-23, 2012, Lugano, Switzerland, Workshop Proceedings, 2012, pp. 340–345. doi:10.1109/PerComW.2012.6197507.
URL https://doi.org/10.1109/PerComW.2012.6197507
35. T. Minkus, N. Memon, On a scale from 1 to 10, how private are you? scoring facebook privacy settings, in: Proceedings of the Workshop on Usable Security (USEC 2014). Internet Society, 2014.
36. M. S. Matell, J. Jacoby, Is there an optimal number of alternatives for likert scale items? study i: Reliability and validity, Educational and Psychological Measurement 31 (3) (1971) 657–674. arXiv:http://dx.doi.org/10.1177/001316447103100307, doi:10.1177/001316447103100307.
URL http://dx.doi.org/10.1177/001316447103100307
37. J. Golbeck, M. L. Mauriello, User Perception of Facebook App Data Access: A Comparison of Methods and Privacy Concerns, Future Internet 8 (2) (2016) 9. doi:10.3390/fi8020009.
URL https://doi.org/10.3390/fi8020009
38. Federal Trade Commission, Android Flashlight App Developer Settles FTC Charges It Deceived Consumers, https://www.ftc.gov/news-events/press-releases/2013/12/android-flashlight-app-developer-settles-ftc-charges-it-deceived, accessed Aug, 2017.
39. Y. Pu, J. Grossklags, Valuating Friends' Privacy: Does Anonymity of Sharing Personal Data Matter?, in: Thirteenth Symposium on Usable Privacy and Security, SOUPS 2017, Santa Clara, CA, USA, July 12-14, 2017., 2017, pp. 339–355.
URL https://www.usenix.org/conference/soups2017/technical-sessions/presentation/pu
40. D. Cooper, J. H. Kagel, Other regarding preferences: a selective survey of experimental results, Handbook of experimental economics 2.
41. J. Ugander, B. Karrer, L. Backstrom, C. Marlow, The Anatomy of the Facebook Social Graph, CoRR abs/1111.4503.
URL http://arxiv.org/abs/1111.4503
42. P. Erdos, A. Rényi, On the Evolution of Random Graphs, Vol. 5, 1960, pp. 17–60.
43. E. Ferrara, G. Fiumara, Topological Features of Online Social Networks, CoRR abs/1202.0331.
URL http://arxiv.org/abs/1202.0331
44. C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, B. Y. Zhao, User interactions in social networks and their implications, in: Proceedings of the 2009 EuroSys Conference, Nuremberg, Germany, April 1-3, 2009, 2009, pp. 205–218. doi:10.1145/1519065.1519089.
URL http://doi.acm.org/10.1145/1519065.1519089

45. A. Mislove, M. Marcon, P. K. Gummadi, P. Druschel, B. Bhattacharjee, Measurement and analysis of online social networks, in: Proceedings of the 7th ACM SIGCOMM Internet Measurement Conference, IMC 2007, San Diego, California, USA, October 24-26, 2007, 2007, pp. 29–42. doi:10.1145/1298306.1298311.
URL `http://doi.acm.org/10.1145/1298306.1298311`
46. R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Reviews of modern physics 74 (1) (2002) 47.
47. D. J. Watts, S. H. Strogatz, Collective dynamics of "small-world" networks, Nature 393 (6684) (1998) 409–10.
48. A. Sundararajan, Local network effects and complex network structure, The BE Journal of Theoretical Economics 7 (1).
49. D. Boyd, N. B. Ellison, Social Network Sites: Definition, History, and Scholarship, J. Computer-Mediated Communication 13 (1) (2007) 210–230.
50. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281, 23.11.1995, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31995L0046, accessed Aug, 2017.
51. Article 29 Working Party, Opinion 1/2010 on the concepts of "controller' and "processor", http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp169_en.pdf, accessed Aug, 2017.
52. B. Van Alsenoy, Regulating data protection: the allocation of responsibility and risk among actors involved in personal data processing.
53. Article 29 Working Party, Opinion 5/2009 on online social networking, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2009/wp163_en.pdf, accessed Aug, 2017.
54. N. Helberger, J. Van Hoboken, Little brother is tagging you legal and policy implications of amateur data controllers, in: Computer Law Review International (CRi), 4/2010, 11(4), pp. 101-109, 2010.
55. Court of Justice of the European Union, Case C-101/01, Bodil Lindqvist, OJ 2004 C7/3, ECLI:EU:C:2003:596, http://curia.europa.eu/juris/liste.jsf?num=C-101/01, accessed Aug, 2017.
56. Recital 39 of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L119, 4.5.2016, http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679, accessed Aug, 2017.
57. Recital 78 of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L119, 4.5.2016, http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679, accessed Aug, 2017.
58. I. Symeonidis, P. Tsormpatzoudi, B. Preneel, Collateral Damage of Online Social Network Applications, in: Proceedings of the 2nd International Conference on Information Systems Security and Privacy, ICISSP 2016, Rome, Italy, February 19-21, 2016., 2016, pp. 536–541. doi:10.5220/0005806705360541.
URL `https://doi.org/10.5220/0005806705360541`
59. Article 29 Working Party, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, http://ec.europa.eu/newsroom/document.cfm?doc_id=44137, accessed Aug, 2017.
60. S. Sackmann, J. Strüker, R. Accorsi, Personalization in privacy-aware highly dynamic systems, Commun. ACM 49 (9) (2006) 32–38. doi:10.1145/1151052.
URL `http://doi.acm.org/10.1145/1151052`
61. D. J. Weitzner, H. Abelson, T. Berners-Lee, C. Hanson, J. A. Hendler, L. Kagal, D. L. McGuinness, G. J. Sussman, K. K. Waterman, Transparent Accountable Data Mining: New Strategies for Privacy Protection, in: Semantic Web Meets eGovernment, Papers from the 2006 AAAI Spring Symposium, Technical Report SS-06-06, Stanford, California, USA, March 27-29, 2006, 2006, p. 141.
URL `http://www.aaai.org/Library/Symposia/Spring/2006/ss06-06-025.php`
62. S. Fischer-Hübner, P. Duquenoy, M. Hansen, R. Leenes, G. Zhang (Eds.), Privacy and Identity Management for Life - 6th IFIP WG 9.2, 9.6/11.7, 11.4, 11.6/ PrimeLife International Summer School,

Helsingborg, Sweden, August 2-6, 2010, Revised Selected Papers, Vol. 352 of IFIP Advances in Information and Communication Technology, Springer, 2011. doi:10.1007/978-3-642-20769-3.
URL https://doi.org/10.1007/978-3-642-20769-3

63. M. Hansen, Marrying Transparency Tools with User-Controlled Identity Management, in: The Future of Identity in the Information Society - Proceedings of the Third IFIP WG 9.2, 9.6/ 11.6, 11.7/ FIDIS International Summer School on The Future of Identity in the Information Society, Karlstad University, Sweden, August 4-10, 2007, 2007, pp. 199–220. doi:10.1007/978-0-387-79026-8_14.
URL https://doi.org/10.1007/978-0-387-79026-8_14

64. N. McDonnel, C. Troncoso, P. Tsormpatzoudi, F. Coudert, L. Métayer, Deliverable 5.1: State-of-play: Current Practices and Solutions. FP7 PRIPARE Project, http://pripareproject.eu/research/#wp5-gaps-and-recommendations, accessed Aug, 2017.

65. K. Liu, E. Terzi, A Framework for Computing the Privacy Scores of Users in Online Social Networks, TKDD 5 (1) (2010) 6:1–6:30. doi:10.1145/1870096.1870102.
URL http://doi.acm.org/10.1145/1870096.1870102

66. N. Talukder, M. Ouzzani, A. K. Elmagarmid, H. Elmeleegy, M. Yakout, Privometer: Privacy protection in social networks, in: Workshops Proceedings of the 26th International Conference on Data Engineering, ICDE 2010, March 1-6, 2010, Long Beach, California, USA, 2010, pp. 266–269. doi:10.1109/ICDEW.2010.5452715.
URL https://doi.org/10.1109/ICDEW.2010.5452715

67. H. Hedbom, A Survey on Transparency Tools for Enhancing Privacy, in: The Future of Identity in the Information Society - 4th IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS International Summer School, Brno, Czech Republic, September 1-7, 2008, Revised Selected Papers, 2008, pp. 67–82. doi:10.1007/978-3-642-03315-5_5.
URL https://doi.org/10.1007/978-3-642-03315-5_5

68. M. Janic, J. P. Wijbenga, T. Veugen, Transparency Enhancing Tools (TETs): An Overview, in: Third Workshop on Socio-Technical Aspects in Security and Trust, STAST 2013, New Orleans, LA, USA, June 29, 2013, 2013, pp. 18–25. doi:10.1109/STAST.2013.11.
URL https://doi.org/10.1109/STAST.2013.11

69. C. Bier, K. Kühne, J. Beyerer, PrivacyInsight: The Next Generation Privacy Dashboard, in: Privacy Technologies and Policy - 4th Annual Privacy Forum, APF 2016, Frankfurt/Main, Germany, September 7-8, 2016, Proceedings, 2016, pp. 135–152. doi:10.1007/978-3-319-44760-5_9.
URL https://doi.org/10.1007/978-3-319-44760-5_9

70. J. Buchmann, M. Nebel, A. Rossnagel, F. Shirazi, H. Simo, M. Waidner, Personal information dashboard: Putting the individual back in control, in: Digital Enlightenment Yearbook 2013: The Value of Personal Data, no. http://ebooks.iospress.nl/publication/35, IOS Press, 2013, pp. 139–164.

71. E. Wästlund, S. Fischer-Hübner, End user transparency tools: UI prototypes.

72. S. Few, Information dashboard design - the effective visual communication of data, O'Reilly, 2006.

73. E. Hamilton, M. Kriens, H. Karapandžic, K. Yaici, M. Main, S. Schniffer, Report on trust and reputation models, ENISA Report.

74. Selenium HQ, Browser Automation, http://docs.seleniumhq.org/, accessed Aug, 2017.

75. P. H. Chia, Y. Yamamoto, N. Asokan, Is this app safe?: a large scale study on application permissions and risk signals, in: Proceedings of the 21st World Wide Web Conference 2012, WWW 2012, Lyon, France, April 16-20, 2012, 2012, pp. 311–320. doi:10.1145/2187836.2187879.
URL http://doi.acm.org/10.1145/2187836.2187879

76. H. Xu, N. Wang, J. Grossklags, Privacy by ReDesign: Alleviating Privacy Concerns for Third-Party Apps, in: Proceedings of the International Conference on Information Systems, ICIS 2012, Orlando, Florida, USA, December 16-19, 2012, 2012.
URL http://aisel.aisnet.org/icis2012/proceedings/ResearchInProgress/102

77. T. Wang, M. Srivatsa, L. Liu, Fine-grained access control of personal data, in: 17th ACM Symposium on Access Control Models and Technologies, SACMAT '12, Newark, NJ, USA - June 20 - 22, 2012, 2012, pp. 145–156. doi:10.1145/2295136.2295165.
URL http://doi.acm.org/10.1145/2295136.2295165

78. H. Hu, G. Ahn, J. Jorgensen, Multiparty Access Control for Online Social Networks: Model and Mechanisms, IEEE Trans. Knowl. Data Eng. 25 (7) (2013) 1614–1627. doi:10.1109/TKDE.2012.97.
URL https://doi.org/10.1109/TKDE.2012.97

79. J. Pang, Y. Zhang, A new access control scheme for Facebook-style social networks, Computers & Security 54 (2015) 44–59. doi:10.1016/j.cose.2015.04.013.
URL `https://doi.org/10.1016/j.cose.2015.04.013`

80. M. M. Lucas, N. Borisov, FlyByNight: mitigating the privacy risks of social networking, in: Proceedings of the 2008 ACM Workshop on Privacy in the Electronic Society, WPES 2008, Alexandria, VA, USA, October 27, 2008, 2008, pp. 1–8. doi:10.1145/1456403.1456405.
URL `http://doi.acm.org/10.1145/1456403.1456405`

81. F. Beato, M. Kohlweiss, K. Wouters, Scramble! Your Social Network Data, in: Privacy Enhancing Technologies - 11th International Symposium, PETS 2011, Waterloo, ON, Canada, July 27-29, 2011. Proceedings, 2011, pp. 211–225. doi:10.1007/978-3-642-22263-4_12.
URL `https://doi.org/10.1007/978-3-642-22263-4\_12`

82. R. Cramer, I. Damgård, J. B. Nielsen, Secure Multiparty Computation and Secret Sharing, Cambridge University Press, 2015.
URL `http://www.cambridge.org/de/academic/subjects/computer-science/cryptography-cryptology-and-coding/secure-multiparty-computation-and-secret-sharing?format=HB&isbn=9781107043053`

83. C. Abdelberi, M. A. Kâafar, R. Boreli, Big friend is watching you: analyzing online social networks tracking capabilities, in: Proceedings of the 2012 ACM workshop on Workshop on Online Social Networks, WOSN 2012, Helsinki, Finland, August 17, 2012, 2012, pp. 7–12. doi:10.1145/2342549.2342552.
URL `http://doi.acm.org/10.1145/2342549.2342552`

84. C. Abdelberi, Y. Ding, R. Dey, M. A. Kâafar, K. W. Ross, A Closer Look at Third-Party OSN Applications: Are They Leaking Your Personal Information?, in: Passive and Active Measurement - 15th International Conference, PAM 2014, Los Angeles, CA, USA, March 10-11, 2014, Proceedings, 2014, pp. 235–246. doi:10.1007/978-3-319-04918-2_23.
URL `https://doi.org/10.1007/978-3-319-04918-2{\_}23`

85. M. Frank, B. Dong, A. P. Felt, D. Song, Mining Permission Request Patterns from Android and Facebook Applications, in: 12th IEEE International Conference on Data Mining, ICDM 2012, Brussels, Belgium, December 10-13, 2012, 2012, pp. 870–875. doi:10.1109/ICDM.2012.86.
URL `https://doi.org/10.1109/ICDM.2012.86`

86. K. Thomas, C. Grier, D. M. Nicol, unFriendly: Multi-party Privacy Risks in Social Networks, in: Privacy Enhancing Technologies, 10th International Symposium, PETS 2010, Berlin, Germany, July 21-23, 2010. Proceedings, 2010, pp. 236–252. doi:10.1007/978-3-642-14527-8_14.
URL `https://doi.org/10.1007/978-3-642-14527-8\_14`

87. Y. Pu, J. Grossklags, Using Conjoint Analysis to Investigate the Value of Interdependent Privacy in Social App Adoption Scenarios, in: Proceedings of the International Conference on Information Systems - Exploring the Information Frontier, ICIS 2015, Fort Worth, Texas, USA, December 13-16, 2015, 2015.
URL `http://aisel.aisnet.org/icis2015/proceedings/SecurityIS/12`

88. H. Harkous, K. Aberer, "If You Can't Beat them, Join them": A Usability Approach to Interdependent Privacy in Cloud Apps, CoRR abs/1702.08234.
URL `http://arxiv.org/abs/1702.08234`

89. E. M. Maximilien, T. Grandison, K. Liu, T. Sun, D. Richardson, S. Guo, Enabling Privacy as a Fundamental Construct for Social Networks, in: Proceedings of the 12th IEEE International Conference on Computational Science and Engineering, CSE 2009, Vancouver, BC, Canada, August 29-31, 2009, 2009, pp. 1015–1020. doi:10.1109/CSE.2009.431.
URL `https://doi.org/10.1109/CSE.2009.431`

90. D. Sánchez, A. Viejo, Privacy Risk Assessment of Textual Publications in Social Networks, in: ICAART 2015 - Proceedings of the International Conference on Agents and Artificial Intelligence, Volume 1, Lisbon, Portugal, 10-12 January, 2015., 2015, pp. 236–241.

91. A. Viejo, D. Sánchez, Enforcing transparent access to private content in social networks by means of automatic sanitization, Expert Syst. Appl. 62 (2016) 148–160. doi:10.1016/j.eswa.2016.06.026.
URL `https://doi.org/10.1016/j.eswa.2016.06.026`

92. R. K. Nepali, Y. Wang, SONET: A SOcial NETwork Model for Privacy Monitoring and Ranking, in: 33rd International Conference on Distributed Computing Systems Workshops (ICDCS 2013 Workshops), Philadelphia, PA, USA, 8-11 July, 2013, 2013, pp. 162–166. doi:10.1109/ICDCSW.2013.49.
URL `https://doi.org/10.1109/ICDCSW.2013.49`

93. L. Sweeney, Simple demographics often identify people uniquely, Health (San Francisco) 671 (2000) 1–34.
94. T. H. Ngoc, I. Echizen, K. Kamiyama, H. Yoshiura, New Approach to Quantification of Privacy on Social Network Sites, in: 24th IEEE International Conference on Advanced Information Networking and Applications, AINA 2010, Perth, Australia, 20-13 April 2010, 2010, pp. 556–564. doi:10.1109/AINA.2010.118.
URL `https://doi.org/10.1109/AINA.2010.118`
95. L. Holtz, H. Zwingelberg, M. Hansen, Privacy Policy Icons, in: Privacy and Identity Management for Life, 2011, pp. 279–285. doi:10.1007/978-3-642-20317-6_15.
URL `https://doi.org/10.1007/978-3-642-20317-6\_15`
96. Y. Wang, P. G. Leon, K. Scott, X. Chen, A. Acquisti, L. F. Cranor, Privacy nudges for social media: an exploratory Facebook study, in: 22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume, 2013, pp. 763–770.
URL `http://dl.acm.org/citation.cfm?id=2488038`
97. ISO 9241-11:1998, Ergonomic requirements for office work with visual display terminals (VDTs) Part 11: Guidance on usability., Technical report, International Organization for Standardization, Geneva, Switzerland.
98. T. Paul, M. Stopczynski, D. Puscher, M. Volkamer, T. Strufe, C4PS - Helping Facebookers Manage Their Privacy Settings, in: Social Informatics - 4th International Conference, SocInfo 2012, Lausanne, Switzerland, December 5-7, 2012. Proceedings, 2012, pp. 188–201. doi:10.1007/978-3-642-35386-4_15.
URL `https://doi.org/10.1007/978-3-642-35386-4\_15`
99. F. Beato, I. Ion, S. Capkun, B. Preneel, M. Langheinrich, For some eyes only: protecting online information sharing, in: Third ACM Conference on Data and Application Security and Privacy, CODASPY'13, San Antonio, TX, USA, February 18-20, 2013, 2013, pp. 1–12. doi:10.1145/2435349.2435351.
URL `http://doi.acm.org/10.1145/2435349.2435351`
100. W. Luo, Q. Xie, U. Hengartner, FaceCloak: An Architecture for User Privacy on Social Networking Sites, in: Proceedings of the 12th IEEE International Conference on Computational Science and Engineering, CSE 2009, Vancouver, BC, Canada, August 29-31, 2009, 2009, pp. 26–33. doi:10.1109/CSE.2009.387.
URL `https://doi.org/10.1109/CSE.2009.387`
101. S. Guha, K. Tang, P. Francis, NOYB: privacy in online social networks, in: Proceedings of the first Workshop on Online Social Networks, WOSN 2008, Seattle, WA, USA, August 17-22, 2008, 2008, pp. 49–54. doi:10.1145/1397735.1397747.
URL `http://doi.acm.org/10.1145/1397735.1397747`
102. M. Conti, A. Hasani, B. Crispo, Virtual private social networks, in: First ACM Conference on Data and Application Security and Privacy, CODASPY 2011, San Antonio, TX, USA, February 21-23, 2011, Proceedings, 2011, pp. 39–50. doi:10.1145/1943513.1943521.
URL `http://doi.acm.org/10.1145/1943513.1943521`
103. L. A. Cutillo, R. Molva, M. Önen, Safebook: A distributed privacy preserving Online Social Network, in: 12th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WOWMOM 2011, Lucca, Italy, 20-24 June, 2011, 2011, pp. 1–3. doi:10.1109/WoWMoM.2011.5986118.
URL `https://doi.org/10.1109/WoWMoM.2011.5986118`
104. S. Jahid, P. Mittal, N. Borisov, EASiER: encryption-based access control in social networks with efficient revocation, in: Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, ASIACCS 2011, Hong Kong, China, March 22-24, 2011, 2011, pp. 411–415. doi:10.1145/1966913.1966970.
URL `http://doi.acm.org/10.1145/1966913.1966970`
105. S. Jahid, S. Nilizadeh, P. Mittal, N. Borisov, A. Kapadia, DECENT: A decentralized architecture for enforcing privacy in online social networks, in: Tenth Annual IEEE International Conference on Pervasive Computing and Communications, PerCom 2012, March 19-23, 2012, Lugano, Switzerland, Workshop Proceedings, 2012, pp. 326–332. doi:10.1109/PerComW.2012.6197504.
URL `https://doi.org/10.1109/PerComW.2012.6197504`
106. L. Vu, K. Aberer, S. Buchegger, A. Datta, Enabling Secure Secret Sharing in Distributed Online Social Networks, in: Twenty-Fifth Annual Computer Security Applications Conference, ACSAC 2009,

Honolulu, Hawaii, 7-11 December 2009, 2009, pp. 419–428. doi:10.1109/ACSAC.2009.46.
URL `https://doi.org/10.1109/ACSAC.2009.46`

107. E. D. Cristofaro, C. Soriente, G. Tsudik, A. Williams, Hummingbird: Privacy at the Time of Twitter, in: IEEE Symposium on Security and Privacy, SP 2012, 21-23 May 2012, San Francisco, California, USA, 2012, pp. 285–299. doi:10.1109/SP.2012.26.
URL `https://doi.org/10.1109/SP.2012.26`

A Questionnaire: Whether users are concerned about the collateral information collection

Facebook offers entertainment applications such as Crime scene, Candy Crash saga, and Angry birds. When you install an application (App) from the Facebook App store, the App may collect your information on Facebook. For instance, Candy Crash saga collects your name, profile picture, country and email address. Other Apps may collect other types of information.

When your friends install Apps on Facebook, these Apps not only collect their information but may also collect your information.

Example 1 If your friends install Travelling Apps, they may collect your current location to notify your friends if you are close by.

Example 2 If your friends install a Dating App (finding potential dating matches for you) the App may collect your birthday, gender, and sexual preferences to find out whether you are attracted by the same physical preferences (e.g., short, tall, brunette, blond) as your friend.

By default, Facebook allows Apps that your friends install to collect information about you. Note that Apps that your friends install collect your information without notifying you in advance or asking for your approval. However, Facebook does have App settings to manually restrict the collection of your information. The types of information that a friend's Facebook App can collect about you are listed in Table 1.

Bellow you will find a questionnaire regarding user preferences about Facebook Apps. Your answers will only be used for scientific purposes. No personal information will be used and responses will be aggregated if published.

On a scale from 1 to 5 where 1 means not concerned at all and 5 means extremely concerned, how concerned are you about the following: (Not concerned at all, Slightly concerned, Moderately concerned, Very concerned, Extremely concerned.)

- The default privacy settings on Facebook allow my friend's Apps to collect my information.
- Facebook does not notify me in advance of the possibility that one of my friend's Apps is going to collect information about me.
- Facebook does not notify me in advance of the possibility that one of my Apps is going to collect information about my friends.
- Facebook does not ask for my approval in advance of the possibility that one of my friend's Apps is going to collect information about me.

On a scale from 1 to 5 where 1 means not concerned at all and 5 means extremely concerned, how concerned are you about the following pieces of your information: (Not concerned at all, Slightly concerned, Moderately concerned, Very concerned, Extremely concerned.)

- Bio
- Birthday
- Family and relationships

- Interested in Religious and political views
- If I'm online
- My status updates
- My photos
- My videos
- My links
- My notes
- Home Town
- Current location
- Education and work
- Activities, interests, things I like
- My app activity

Would you like to be notified in the following cases? (Never, Depending on the type of information that is collected, Always, I dont know.)

- When one of my friend's Apps is going to collect pieces of my information.
- When one of my Apps is going to collect pieces of my friend's information.

Can you please elaborate on your answer to the previous question? (open answer)

Which actions would you take if you are notified that one of your friends Apps is going to collect your information? (Check all that apply)

- I would take no action.
- I would restrict access to my personal information for Apps that collect my information.
- I would restrict access to my information for this friend.
- I would request my friend to remove the App.
- I would un-friend my friend.
- I dont know.
- Other:

Which actions would you take if you are notified that one of your Apps is going to collect your friends information? (Check all that apply)

- I would take no action.
- I would ask my friend for her/his approval.
- I would restrict access to my friends personal information for this App.
- I would remove this App from my list.
- I dont know.
- Other:

For how long have you had a Facebook account? (Approximately)

How many Facebook friends do you have? (Approximately)

Have you ever installed an App on Facebook?

- Yes
- No

If yes, how many Facebook Apps have you installed in the past six months?

- 1 - 2
- 3 - 4
- 5 - 6
- 7+

What kind of Apps do you often use on Facebook? (For instance games, lifestyle, navigation, weather, etc.)

Have you ever changed the Apps privacy settings?

- Yes
- No

If yes, do you remember which permission you changed? (Check all that apply)

- I restricted who can see my profile information.
- I restricted who can see me in searches.
- I restricted who can collect my information through my friends Apps.
- Other:

In which country were you born?

In which year were you born?

What is your gender?

- Male
- Female
- Prefer not to answer

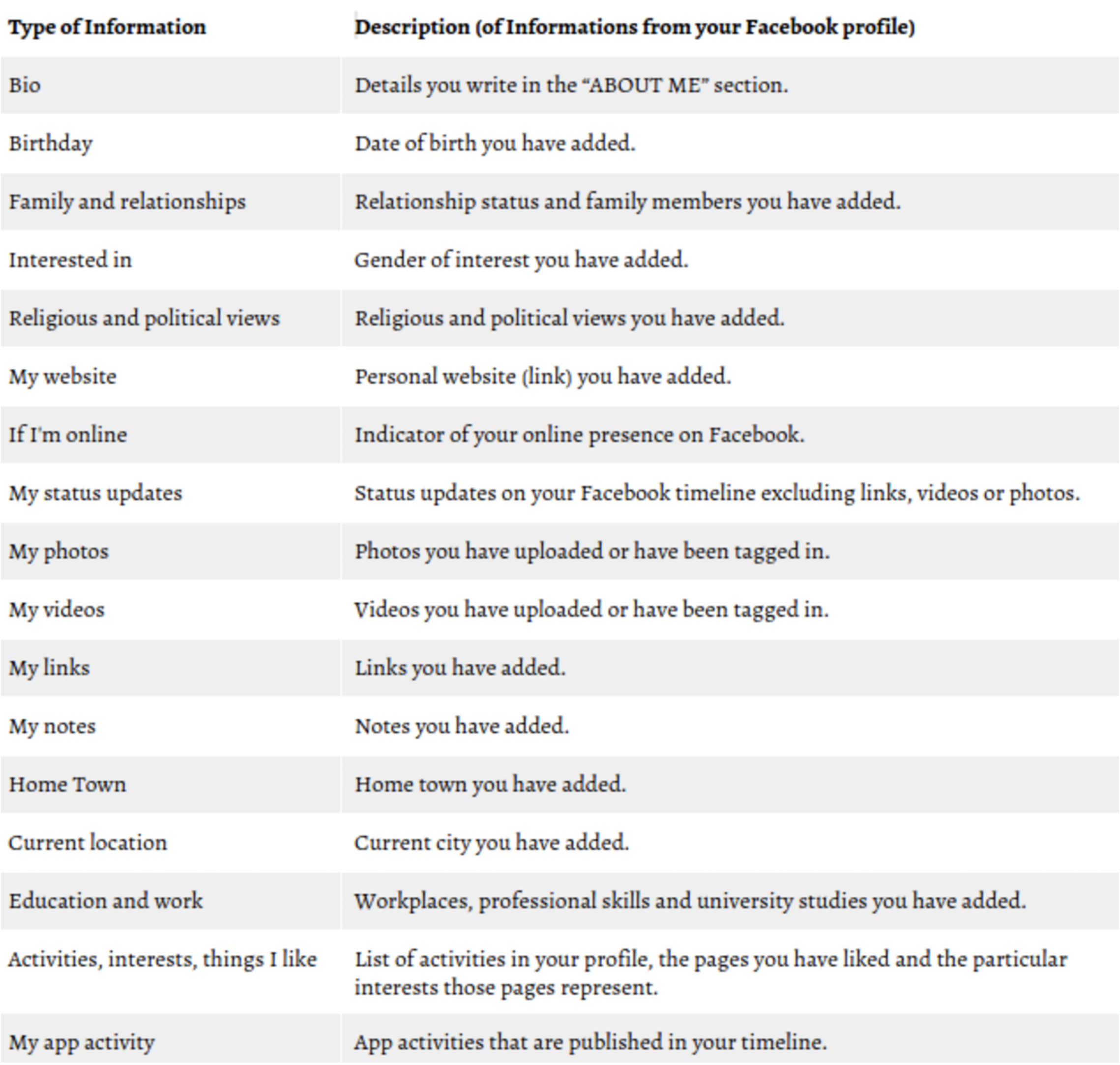
What is the highest level of education you have completed?

- No degree or up to high school
- Bachelors degree or equivalent
- Masters degree and above
- Other:

Do you have an IT background? (Check all that apply)

- Online courses or seminars
- Higher education
- Personal interest
- Other:

Remarks and Questions



Type of Information	Description (of Informations from your Facebook profile)
Bio	Details you write in the “ABOUT ME” section.
Birthday	Date of birth you have added.
Family and relationships	Relationship status and family members you have added.
Interested in	Gender of interest you have added.
Religious and political views	Religious and political views you have added.
My website	Personal website (link) you have added.
If I'm online	Indicator of your online presence on Facebook.
My status updates	Status updates on your Facebook timeline excluding links, videos or photos.
My photos	Photos you have uploaded or have been tagged in.
My videos	Videos you have uploaded or have been tagged in.
My links	Links you have added.
My notes	Notes you have added.
Home Town	Home town you have added.
Current location	Current city you have added.
Education and work	Workplaces, professional skills and university studies you have added.
Activities, interests, things I like	List of activities in your profile, the pages you have liked and the particular interests those pages represent.
My app activity	App activities that are published in your timeline.

Fig. 21: The types of information that a Facebook App can collect