Privacy Scoring of Social Network User Profiles through Risk Analysis

Sourya Joyee De, Abdessamad Imine

Privacy Scoring of Social Network User Profiles through Risk Analysis*

Sourya Joyee De[2] and Abdessamad Imine[1,2]

[1] Lorraine University, France
abdessamad.imine@loria.fr
[2] LORIA-INRIA Nancy Grand-Est, France
sourya-joyee.de@inria.fr

Abstract. The social benefit derived from online social networks (OSNs) can lure users to reveal unprecedented volumes of personal data to a social graph that is much less trustworthy than the offline social circle. Although OSNs provide users privacy configuration settings to protect their data, these settings are not sufficient to prevent all situations of sensitive information disclosure. Indeed, users can become the victims of harms such as identity theft, stalking or discrimination. In this work, we design a privacy scoring mechanism inspired by privacy risk analysis (PRA) to guide users to understand the various privacy problems they may face. Concepts, derived from existing works in PRA, such as privacy harms, risk sources and harm trees are adapted in our mechanism to compute privacy scores. However, unlike existing PRA methodologies, our mechanism is user-centric. More precisely, it analyzes only OSN user profiles taking into account the choices made by the user and his vicinity regarding the visibility of their profile attributes to potential risk sources within their social graphs. To our best knowledge, our work is the first effort in adopting PRA approach for user-centric analysis of OSN privacy risks.

Keywords: Online Social Networks (OSN), privacy harms, privacy score, harm trees, Privacy Risk Analysis (PRA).

1 Introduction

Users reveal personal data, build their social graphs and affiliate to groups to derive various social benefits (such as connecting to offline friends, establishing new connections) from their online social network (OSN) profiles. It is possible to infer various personal data of a user not only from the values of the OSN profile attributes (such as birth year, home address, work place, education) revealed by the user himself, but also from those revealed by his friends and from group affiliations [24,2,14]. Moreover, members of the social graph may be complete strangers, future employers, colleagues, relatives, etc., from whom various privacy risks may arise. For example, in his workplace, an employee may withhold some information about himself and maintain an image that is different from his

* This work is partially funded by MAIF Foundation.

personal life [11]. An OSN profile may reveal these otherwise hidden information to colleagues leading to poor impression or hurting professional growth. Users can also become the victims of harms such as identity theft, stalking, discrimination, or sexual predation. In the absence of additional support, the privacy settings provided by OSNs are not enough to mitigate these privacy problems. So, there is a need to guide users to: 1) understand the privacy problems they may face due to their actions on OSNs (such as the personal data they reveal, the social circle they build) and 2) adopt suitable preventive measures. Designing such a guidance tool is our broad aim. In this work, we focus on the first step, i.e., design a privacy scoring mechanism to compute for the users the privacy risks of their OSN profiles and social graphs.

Computation of the privacy level of an OSN user's profile in terms of privacy metrics has recently drawn the attention of researchers [13,18,20,22,15,17]. In contrast to these works, our privacy scoring mechanism is inspired by privacy risk analysis (PRA) [5,3,4,9]. A PRA methodology helps service providers to assess the privacy risks of information systems that process personal data. Such methodologies are gaining focus as the EU General Data Protection Regulation (GDPR) mandates the conduction of a data protection impact assessment[3] for service providers with certain categories of personal data processing.

In this work, we adopt the PRA approach in designing our privacy scoring mechanism to assist users (instead of the service provider), borrowing concepts like privacy harms, risk sources and harm trees from [5,6,7]. Unlike existing PRA methodologies, we do not consider the entire OSN system or risk sources like hackers or the service provider and ignore privacy weaknesses [5] introduced by the service provider's choices during system design and implementation. Instead, we focus on the choices made by the user and his friends regarding the visibility of their profile attributes to potential risk sources already in their social graph. To the best of our knowledge, our work is the first effort in utilizing PRA concepts for user-centric analysis of OSN privacy risks based on the visibility of attribute values.

We introduce the main ingredients of our privacy scoring mechanism in Section 2 and discuss attribute visibility from an OSN user profile in Section 3. In Section 4 we present our privacy scoring mechanism. Finally, in Section 5 we discuss related works and conclude with future directions in Section 6.

2 Model Ingredients

Users may publish various personal data in their OSN profiles. Various actors in the OSN may become *risk sources* processing the revealed personal data to cause a variety of *threats* that ultimately lead to *privacy harms* for the user. In what follows, we define these concepts, which form the building blocks of our privacy scoring mechanism, more formally and provide appropriate examples. We represent the OSN as a graph $G = (V, E)$, where V is the set of nodes

[3] The technical details of a privacy impact assessment (PIA) are referred to as privacy risk analysis (PRA) [5,6].

representing the users of the OSN and E is the set of edges representing the friendship links among the users. $e_{i,j} \in E$ represents a friendship link between the nodes v_i and v_j. The *target user*, denoted by v_T, represents the OSN user for whom the privacy score is being computed. We also assume that the target user has at least one friend.

Attributes and Other Personal Data. Some personal data are made available by the target user and his friends in their OSN profiles. We call these personal data *user attributes* that can be defined as:

Definition 1. *A **user attribute** is a personal data*[4] *item considered as a part of the user profile information. It helps to present this user to other users of the same OSN.*

Each user has a set A of profile attributes. We consider the following elements of set A: 1. Birth year (B.Yr); 2. Birthday (B.Dt); 3. Gender (Gen); 4. Phone number (Ph); 5. Gender interests (G.Int); 6. Home address (H.Add); 7. Workplace (W.Pl); 8. Work designation (W.desig); 9. Political views (Pol); 10. Religious views (Rel); 11. Relationship status (RStat); 12. Interests (Int). Each user attribute may assume different values. Other personal data such as work locality (W.Loc) can be obtained by inference from these attributes. Other attributes may also be revealed in different OSNs, but we consider only this set for the current discussion. We also assume that providing a name is mandatory and can be seen by everyone on the OSN. So we do not consider it as an attribute.

Privacy Harms. We adapt the definition of privacy harm from [5,6,7] in the context of an OSN.

Definition 2. *A **privacy harm** is the negative impact of the use of an OSN on the target user as a result of one or more privacy breaches.*

Over the years, many types of privacy harms have been observed in real life as well as found to be possible by different research works [12,11,21,16,10] from the data revealed from OSNs. In this work, we consider two harms: 1) stalkers use the target user's profile to assess him as a potential victim (H.1) and 2) identity fraud/theft (H.2). Of course, the harms presented here are not exhaustive and only involve a subset of the user attributes provided above. Other harms, involving different user attributes, are possible and can be analyzed in the same way as we will show in the next sections for these representative harms.

Risk Sources. We adapt the definition of risk sources from [5,6,7] in the context of an OSN.

Definition 3. *A **risk source** is any entity (individual or organization) that may process (legally or illegally) data belonging to the target user and whose actions may directly or indirectly, intentionally or unintentionally lead to privacy harms.*

[4] according to the GDPR (General Data Protection Regulation) of European Union.

In this work, we focus on the user's social graph to find out the relevant risk sources which include: 1) friends of the target user ($A.1$); 2) the friends of friends of the target user ($A.2$); 3) the friends of friends of friends of the target user ($A.3$); 4) the strangers to the target user (degrees of relationships higher than 3) ($A.4$). These risk sources only process data already made visible to them by the user leading to various harms. For example, the colleagues of the user who are his friends in the OSN ($A.1$) can form a negative impression about him based on his political and/or religious views or based on his interests, sexual orientation, etc., which may negatively affect him at his work-place. We ignore risk sources such as the OSN service provider, the government and hackers.

Threats. We define threats in the context of an OSN as:

Definition 4. *A **threat** is an action of a risk source with respect to one or more pieces of personal data resulting in a privacy harm.*

In the context of an OSN, threats include unintended inference of data (FE.1) (e.g., strangers infer the gender of the target user from the genders of his friends), direct access to data by unintended audiences due to similar attributes revealed by the user (FE.2) (e.g., friends of friends come to know the user's phone number), and the undesirable reactions from intended audiences (FE.3) (e.g., colleagues respond negatively to the target user's political views) [21,11,16]. We only consider threats resulting from inappropriate privacy settings used by the target user and his friends for their attributes and ignore threats originating from the service provider's design and/or implementation choices (e.g., lack of anonymization, poor protection of data stores) as we only focus on the analysis of the OSN user profile and not the entire system.

Inference of Personal Data. The attributes revealed by the target user or his friends can reveal other personal data of the target user. The attributes used for the inference could be of the same type. For example, the gender (Gen) of the target user's friends can be used to infer the gender (Gen) of the target user. It is also possible to use other types of attributes to reveal a particular personal data. For example, the work place (W.Pl), a data about the user's profession, is an indicator of the target user's work location (W.Loc), which is a location data. Sometimes, multiple attributes can be used to infer a personal data item. For example, the sexual orientation (SO) of a target user can be inferred from his gender interests (G.Int) and gender (Gen). These different types of inference methods can thus be categorized based on three criteria as follows: 1) whether the personal data is inferred *directly*, i.e., from attribute(s) revealed by the target user himself or *indirectly*, i.e., from attribute(s) revealed by the friends of the target user; 2) whether a *single* or *multiple* attribute(s) are used for the inference; 3) whether the attribute(s) used for the inference constitutes a *similar* type of personal data as the one that is being inferred or are completely *different*. Here, we only consider direct/indirect, single and similar attribute inference for user attributes and direct/indirect, single/multiple and similar/different attribute inference for other personal data not included as user attributes.

Table 1 presents the attributes that can be used to infer various types of personal data through some of the above inference methods[5]. The types of personal data (such as contact data, location data, identification data) we use are inspired from [6]. A particular personal data can be inferred using one or more inference methods. The choice of inference method depends on the availability of attribute values and the desired accuracy of inference.

Personal data type	Code	User Attribute or Other Personal Data	User Attribute (Target User)	User Attribute (Friends)	Inference Types
Contact data	M.1	Phone No. (Ph.)	Phone No. (Ph.)	×	Direct, single, similar attribute
	M.2	Home Address (H.Add)	Home Address (H.Add)	×	Direct, single, similar attribute
Location data	M.3	Home Locality (H.Loc)	Home Address (H.Add)	Home Address (H.Add)	Direct/indirect, single, different attribute
	M.4	Work Locality (W.Loc)	Workplace (W.Pl)	Workplace (W.Pl)	Direct/indirect, single, different attribute
Identification data	M.5	Gender (Gen)	Gender (Gen)	Gender (Gen)	Direct/indirect, single, similar attribute
	M.6	Age (Age)	Birth year (B.Yr)	Birth year (B.Yr)	Direct/indirect, single, similar attribute
	M.7	Date of birth (DoB)	Birth year (B.Yr), Birth day (B.Dt)	×	Direct, multiple, similar attribute

Table 1: Inferring user attributes and other personal data

3 Attribute Visibility

After assigning values to the attributes in their OSN profiles, users can select from a range of privacy settings to ensure that the attribute values are visible to desirable audiences in their social graph. Here, we consider that the user can choose from the following privacy settings, inspired from those used in Facebook:

1. *"private"*: makes an attribute value visible to no one;
2. *"friends"*: makes an attribute value visible to friends only;
3. *"friends of friends"*: makes an attribute value visible to friends and friends of friends;
4. *"public"*: makes an attribute value visible to all users of the OSN.

The visibility matrix $\mathbf{M}$ of a target user v_T displays the visibility values of all the attributes in A (the set of user attributes, see Definition 1) as given by their privacy settings chosen by v_T and his friends. Each element of the matrix is a set that denotes the members of the OSN to whom the jth attribute a_j is

[5] In Table 1, neither the list of inference methods nor the personal data that can be inferred from the given set of attributes nor the personal data types that must be considered is exhaustive. Other inferred personal data, personal data types and inference methods can be easily incorportated in our framework.

visible. These members are assigned based on the privacy setting of the attribute selected either by v_T or a friend of v_T. Entry **M(1,j)** represents the visibility of the jth attribute, $v_T.a_j$, as set by v_T. As for **M(i,j)**, with $i > 1$, it represents the visibility of the jth attribute, $v_i.a_j$, as set by the ith friend ($i \neq 1$) of v_T (but, with respect to v_T and not themselves)[6]. Other types of privacy settings used in other OSNs can also be used to fill in **M**.

For $i = 1$, i.e., for v_T himself, **M(i,j)** is assigned values as follows:

1. **M(i,j)** = {}, if the privacy setting of $v_T.a_j$ is "private";
2. **M(i,j)** = $\{A.1\}$, if the privacy setting of $v_T.a_j$ is "friends";
3. **M(i,j)** = $\{A.1, A.2\}$, if the privacy setting of $v_T.a_j$ is "friends of friends";
4. **M(i,j)** = $\{A.1, A.2, A.3, A.4\}$, if the privacy setting of $v_T.a_j$ is "public".

For $i > 1$, i.e., for the friends v_i of v_T, **M(i,j)** is assigned values as follows:

1. **M(i,j)** = {}, if the privacy setting of $v_i.a_j$ is "private";
2. **M(i,j)** = $\{A.1, A.2\}$, if the privacy setting of $v_i.a_j$ is "friends"[7];
3. **M(i,j)** = $\{A.1, A.2, A.3\}$, if the privacy setting of $v_i.a_j$ is "friends of friends";
4. **M(i,j)** = $\{A.1, A.2, A.3, A.4\}$, if the privacy setting of $v_i.a_j$ is "public".

The true visibility $Vis_{true}(v_T.a_j)$ of a target user's attribute is the same as **M(1,j)**. However, its observed visibility $Vis_{obs}(v_T.a_j)$ depends on the values of **M(i,j)**, for all i. For our purpose, we assume that $Vis_{obs}(v_T.a_j)$ is the set **M(i,j)** that has the maximum number of risk sources for a given attribute a_j over all i, i.e., the observed visibility is the same as the weakest privacy setting among all the privacy settings assigned to the attribute by the target user and his friends. For some attributes whose value cannot be inferred from the attribute values of the friends due to the nature of the attribute (for example, birth day (B.Dt), phone no. (Ph), etc.), $Vis_{obs}(.) = Vis_{true}(.)$.

We now show how the visibility matrix and the true and observed visibility values are computed for a target user Ana, for the attribute B.Yr, given her friendship network and the disclosure of this attribute by her and her friends in Figure 1.

Figure 2 presents Ana's visibility matrix. The first row of the matrix, **M(1,B.Yr)**, corresponds to Ana's privacy setting for B.Yr. The subsequent rows represent the privacy settings of her friends (but, with respect to her) for B.Yr. For example, Figure 1 shows that Ana's friend Emma reveals her B.Yr to her friends. Thus, apart from Ana herself and her mutual friends with Emma, Emma's B.Yr is visible to Emma's friends who are friends of friends with respect to Ana. Therefore, in the visibility matrix, we fill up the row corresponding to Emma for B.Yr with the value $\{A.1, A.2\}$ (and not $\{A.1\}$, because it is filled up from Ana's point of view). Ana's friend Bob reveals his B.Yr to his friends of friends. From Ana's point of view, Bob's B.Yr is visible to Ana's friends of friends of friend. So we fill up the corresponding cell in the visibility matrix with

[6] Notation wise, for simplicity, we assume that the target user is the first friend for himself, i.e., when $i = 1$, $v_i = v_T$.

[7] $A.2$ is included because a friend of v_i ($i \neq 1$) is a friend of friend of the target user.

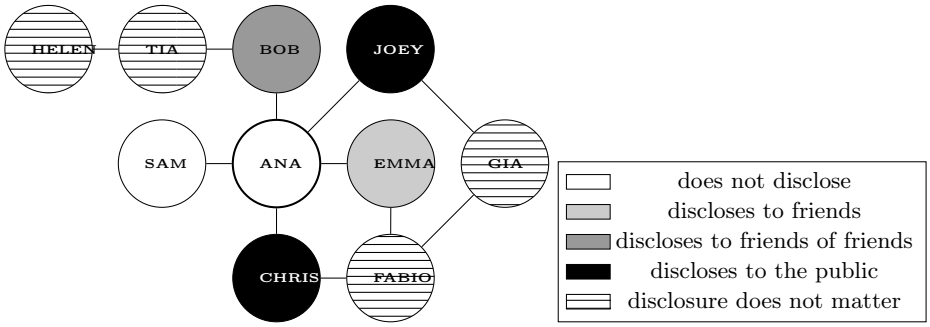

Fig. 1: The target user and its vicinty for the revelation of the attribute B.Yr

the value $\{A.1, A.2, A.3\}$. Ana's friend Joey reveals his B.Yr to the public (i.e., beyond friend of friend), i.e., $\{A.1, A.2, A.3, A.4\}$ with respect to Ana. The true visibility of Ana's B.Yr is given by $Vis_{true}(v_{Ana}.B.Yr) = \{\}$ and the observed visibility of Ana's B.Yr is given by $Vis_{obs}(v_{Ana}.B.Yr) = \{A.1, A.2, A.3, A.4\}$.

$$\begin{array}{c} \\ \text{Ana} \\ \text{Bob} \\ \text{Chris} \\ \text{Emma} \\ \text{Joey} \\ \text{Sam} \end{array} \begin{array}{c} \text{B.Yr} \\ \left(\begin{array}{c} \{\} \\ \{A.1, A.2, A.3\} \\ \{A.1, A.2, A.3, A.4\} \\ \{A.1, A.2\} \\ \{A.1, A.2, A.3, A.4\} \\ \{\} \end{array} \right) \end{array}$$

Fig. 2: Visibility matrix for the target user Ana for B.Yr

4 Privacy Scoring Mechanism

The discussions in Section 2 and Section 3 form the basis of the privacy scoring mechanism that we describe in this section. As discussed in the Introduction, the mechanism ultimately informs users of an OSN about the privacy risks of their profiles and social graphs. In brief, the privacy scoring mechanism consists of the following steps, each of which we discuss in details with appropriate examples in the rest of this section:

1. Construction of a harm tree for each privacy harm.
2. Pruning harm trees based on attribute visibilities.
3. Computation of the accuracy values for each attribute value.
4. Pruning harm trees based on the accuracy values.
5. Evaluation of the likelihood of each harm.

4.1 Construction of harm trees

The first step in deriving the privacy score is to construct the harm tree for each privacy harm. A harm tree [5,6,7] describes the relationship among the privacy harms, threats, risk sources and the personal data/ attributes of the target user. The root node of a harm tree denotes a privacy harm. Leaf nodes represent the exploitation of personal data (user attributes or other personal data) by risk sources. Intermediate nodes represent the threats caused by the risk sources. Child nodes can be connected by: 1) an AND node if all of them are necessary to give rise to the parent; 2) an OR node if any one of them is sufficient to give rise to the parent and 3) a k-out-of-n node if any k of the n child nodes are sufficient to give rise to the parent node.

In case of some harms, the personal data that can be exploited may vary from risk source to risk source or a particular occurrence of the harm to another one. For example, a potential employer may assess the target user's profile based on political views, religious views, sexual orientation, interests and relationship status or a subset of these data. In such cases, we present n of the most probable attributes leading to the harm in the harm tree. Out of these n attributes, any k may be used by the risk source leading to the harm.

The harm tree for H.1 in Figure 3 represents that a target user's profile can be assessed for suitability for stalking by a friend of a friend of a friend ($A.3$) or a stranger ($A.4$). The stalker can use either the gender (Gen) or the age (Age derived from the attribute B.Yr) or both of a target user to assess the profile. The risk source also needs to know a more or less precise location data for the user given by the home locality (H.Loc derived from H.Add) or the work locality (W.Loc derived from W.Pl). These data can be either accessed directly (FE.2) or can be inferred (FE.1). Figure 4 presents the harm tree for H.2.

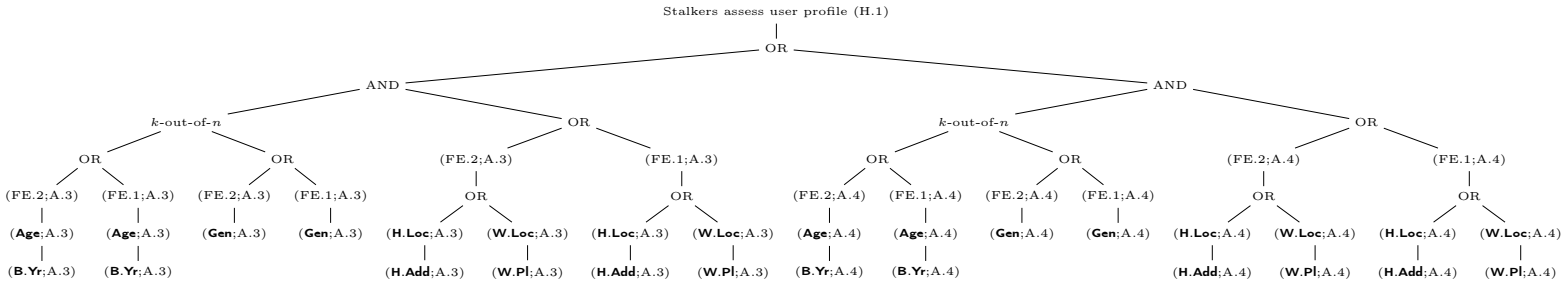

Fig. 3: Harm tree for H.1

The harm trees can be constructed by privacy experts beforehand and stored in a database. The latter can be updated when new harms are discovered. Existing harm trees can also be modified based on new information. This step can be performed once (and the database can be updated once in a while) and can be reused for each target user.

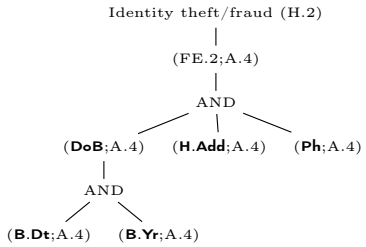

Fig. 4: Harm tree for H.2

4.2 Pruning harm trees based on attribute visibility

The observed visibilities $Vis_{obs}(.)$ of the target user's attributes are derived from the visibility matrix $\mathbf{M}(i, j)$. Table 2 represents the true and the observed visibilities (derived from the visibility matrix of the corresponding user) of an example target user T (accuracy is discussed in Section 4.3 and the column for accuracy is used in Section 4.5). The branches of the harm trees using the attributes for which $|Vis_{obs}(.)| = 0$ can be pruned as these attributes or personal data are neither disclosed by the user nor can they be inferred from his friends. So, for the target user T, the branches in the harm tree for H.2 (see Figure 4) corresponding to DoB (since $|Vis_{obs}(T.\mathsf{B.Dt})| = 0$ for B.Dt and both B.Dt and B.Yr are required to obtain DoB), H.Add (since $|Vis_{obs}(T.\mathsf{H.Add})| = 0$) and Ph (since $|Vis_{obs}(T.\mathsf{Ph})| = 0$) can be pruned (pruning shown by $\times$ in Figure 5).

Next, a second level of pruning can be carried out based on whether a harm tree uses the exploitation of personal data by a risk source who does not have access to it. For example, suppose that for the attributes B.Dt, H.Add and Ph of another target user T', $Vis_{obs}(.) = \{A.1\}$, implying that the risk sources $A.2$, $A.3$ and $A.4$ do not have access to these attribute values nor can they infer the required personal data (e.g. DoB) to cause the harm. In the harm tree for H.2 (see Figure 4), the risk source $A.4$ must have access to DoB, Ph. and H.Loc. So, for T', the corresponding branches are pruned in the harm tree for H.2. In contrast, if the observed visibility values of B.Dt, B.Yr, H.Add and Ph for a target user T'' are given by $Vis_{obs}(.) = \{A.1, A.2, A.3, A.4\}$, the corresponding branches of the harm tree for H.2 cannot be pruned.

The harm tree for H.2 becomes non-existent for the target users T and T' as the personal data necessary to cause H.2 are not available to the risk source $A.4$. So the privacy settings of T and T' and those of their friends protect them from H.2 but the privacy settings of T'' and his friends do not. The harm tree for H.1 (given in Figure 3) can be pruned similarly (see Figure 6).

4.3 Accuracy of attribute values

The accuracy of an attribute in having a particular value depends on the true and the observed visibility of the attribute(s) from which it can be derived. If for an attribute, $|Vis_{true}| > 0$, the target user has himself revealed the attribute. So,

Attribute ($v_T.a_j$)	**True Visibility** $Vis_{true}(v_T.a_j)$	**Observed Visibility** $Vis_{obs}(v_T.a_j)$	**Accuracy**
B.Dt	{}	{}	$A.1, A.2, A.3, A.4 : 0$
B.Yr	{}	$\{A.1, A.2, A.3, A.4\}$	$A.1, A.2 : 0.45; A.3 : 0.4; A.4 : 0.4$
Gen	{}	$\{A.1, A.2, A.3, A.4\}$	$A.1, A.2 : 0.8; A.3 : 0.7; A.4 : 0.6$
Ph	{}	{}	$A.1, A.2, A.3, A.4 : 0$
H.Add	{}	{}	$A.1, A.2, A.3, A.4 : 0$
W.Pl	{}	$\{A.1, A.2, A.3, A.4\}$	$A.1, A.2 : 0.45; A.3 : 0.4; A.4 : 0.3$

Table 2: True and observed visibility sets and the accuracy values for T

when the target user reveals an attribute to his friends (i.e., $Vis_{true} = \{A.1\}$), to his friends of friends (i.e., $Vis_{true} = \{A.1, A.2\}$) and to strangers (i.e., $Vis_{true} = \{A.1, A.2, A.3, A.4\}$), then the corresponding risk sources know the value of the corresponding attribute with full accuracy[8]. When there is a difference in the observed and the true visibility sets, then at least some risk sources do not know the value with full accuracy and therefore infer the value with some accuracy. We consider a simple measure of accuracy for the jth attribute of the target user v_T as derived by the kth risk source $A.k$ given as:

$$Acc(v_T.a_j)_{A.k} = Max_l(Pr[v_T.a_j = s_l | \forall i, i > 1, v_i.a_j = s_l, e_{T,i} \in E, A.k \in \mathbf{M(i,j)}])$$

$$= Max_l(\frac{|v_i.a_j|_{v_i.a_j = s_l, i>1, A_k \in \mathbf{M(i,j)}}}{|v_i.a_j|_{i>1}})$$

where, s_l is the lth value that can be assumed by the attribute a_j of the target user v_T or his friend v_i, $\mathbf{M}$ is the visibility matrix and A_k is the kth risk source. $|v_i.a_j|_{i>1}$ denotes the total number of friends v_i (we assume that the target user has at least one friend) and $|v_i.a_j|_{v_i.a_j = s_l, i>1, A_k \in \mathbf{M(i,j)}}$ denotes the number of friends v_i for whom $v_i.a_j = s_l$ and $A_k \in \mathbf{M(i,j)}$. The range of values (s_l for all l) assumed by an attribute can be obtained from the values assigned to the attribute by friends of v_T or from an accepted set of values (e.g., cities in France).

The above formula can be used to compute the accuracy value for attributes that assume a categorical value. For example, Gen can assume a value from {Male, Female}, RStat can assume a value from {Single, Married, Divorced} etc. For some attributes such as B.Yr, instead of inferring the exact value, the risk source may infer the range within which the value lies.

We illustrate the computation of accuracy values with an example. Suppose the target user T' does not reveal his B.Yr. He has a 100 friends and 60 of those friends reveal their B.Yr to strangers (i.e., $\mathbf{M(i,B.Yr)} = \{A.1, A.2, A.3, A.4\}$, $1 < i \leq 61$), 5 of them reveal it to their friends of friends (i.e., $\mathbf{M(i,B.Yr)} = \{A.1, A.2, A.3\}$, $61 < i \leq 66$) and 10 reveal it to their friends (i.e., $\mathbf{M(i,B.Yr)} = \{A.1, A.2\}$, $66 < i \leq 76$). The rest, i.e., 25 do not reveal it at all (i.e., $\mathbf{M(i,B.Yr)} = \{\}$, $76 < i \leq 101$). We further assume that of the first 60 friends, 70% are in the range of 1980 to 1990, 20% are earlier than 1980 and remaining later than 1990. For all the other groups, 20% are in the range of 1980 to 1990,

[8] The accuracy values lie between 0 (no accuracy) and 1 (full accuracy).

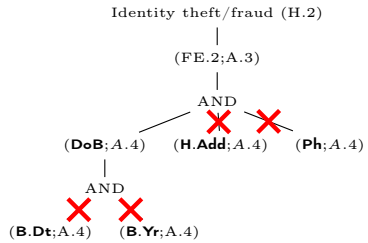

Fig. 5: Pruning of harm tree for H.2 for T and T' based on visibility

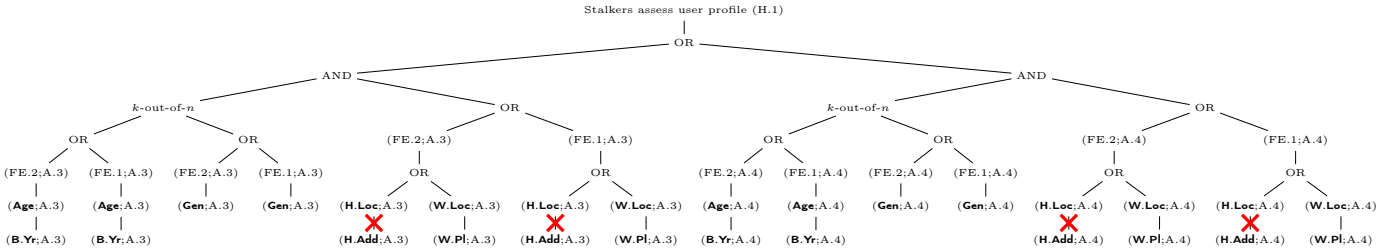

Fig. 6: Pruning of harm tree for H.1 for T based on visibility

40% earlier than 1980 and remaining later than 1990. Then the accuracy with which $A.1$ (mutual friends) can infer about the B.Yr of T' is:

$Acc(v_{T'}.\mathsf{B.Yr})_{A.1} = Max(Pr[1980 \leq v_{T'}.\mathsf{B.Yr} \leq 1990 | \forall i, 1980 \leq v_i.\mathsf{B.Yr} \leq 1990, e_{T',i} \in E, A.1 \in \mathbf{M(i,B.Yr)}], Pr[v_{T'}.\mathsf{B.Yr} < 1980 | \forall i, v_i.\mathsf{B.Yr} < 1980, e_{T',i} \in E, A.1 \in \mathbf{M(i,B.Yr)}], Pr[v_{T'}.\mathsf{B.Yr} > 1990 | \forall i, v_i.\mathsf{B.Yr} > 1990, e_{T',i} \in E, A.1 \in \mathbf{M(i,B.Yr)}]) = Max(0.45, 0.18, 0.12) = 0.45.$

The computation of the accuracy value is inspired by the friend-aggregated model in [24]. However, as discussed in [24], other types of computations of the accuracy value are also possible, depending upon the inference method being used. Different risk sources may choose different inference methods based on their capabilities. The computation method presented above provides a lower bound to the achievable accuracy values – risk sources, using better inference methods, can achieve better accuracy. Our aim is to provide the user with a base level for the score (improving the inference method is not our focus), implying that the privacy risk is at least equal to the privacy score that we present.

4.4 Pruning the harm trees based on accuracy

Once the accuracy values are known, a third stage of pruning can be carried out based on which attributes in the harm trees are known with full accuracy and which ones are to be inferred. We show this step for T and the harm H.1 in Figure 7. For T, the attributes B.Yr and Gen have to be inferred from what T's friends reveal by the risk sources $A.3$ and $A.4$ (FE.1) as similar attributes have not been disclosed by T himself (FE.2). So the branches of this harm tree for

these attributes and FE.2 are pruned. Similarly, H.Add and W.Pl must be inferred from what T's friends have disclosed (FE.1) as similar attributes have not been disclosed by T himself (FE.2), by both risk source $A.3$ and $A.4$. So whenever an attribute value is known with full accuracy[9] by a risk source, the corresponding branch (FE.2) in the tree is left untouched while the branch for inferring the value of the attribute (FE.1) by that risk source is pruned. Otherwise branches with FE.1 are retained and those with FE.2 pruned. We also fix the values for k and n. In the worst case (for the user), each risk source uses only the attribute having the maximum accuracy for the harm. Then, we substitute all nodes with k-out-of-n by OR nodes [23]. In the best case (for the user), each risk source uses all the attributes for the harm. In this case, we substitute all k-out-of-n nodes by AND nodes. There may be intermediate cases, where risk sources use different number and combinations of attributes. For example, one intermediate scenario is where the attributes with the top k accuracy values are used by the risk source.

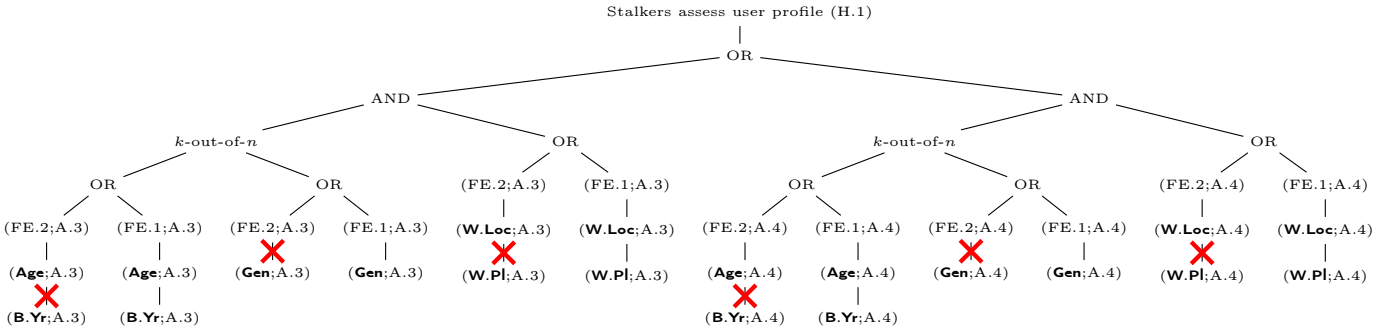

Fig. 7: Pruning of the harm tree for H.1 for T based on accuracy

4.5 Evaluation of harm likelihoods

Once accuracy values are assigned to all the leaf nodes in a harm tree, they must be combined to obtain the overall likelihood of the harm. The combination uses the following rules, inspired from [23], where Acc_i is the accuracy value of the ith attribute (i.e., ith child node):[R1.] AND node: $\prod_i Acc_i$, $i = 1, \ldots, n$ (assuming independence of child nodes); [R2.] OR node: $Max_i(Acc_i)$; [R3.] k-out-of-n node: $\prod_i Acc_i$, $i = 1, \ldots, k$, where the k attributes are the ones with the top k accuracy values (assuming independence of child nodes). The above rules are applied bottom-up on the harm tree. We illustrate the computation of the

[9] The value of an attribute is known with full accuracy only when the value is disclosed by the target user himself, i.e., only for some cases of direct, similar attribute inferences (e.g., a risk source comes to know v_T's gender because v_T reveals it).

likelihood of H.1 for T using the example accuracy values in Table 2 for the worst case in Figure 8. The accuracy values and the likelihood value for the relevant nodes are presented inside curly brackets beside each node. The likelihood of H.1 is 0.28. The likelihoods for other harms can be similarly computed.

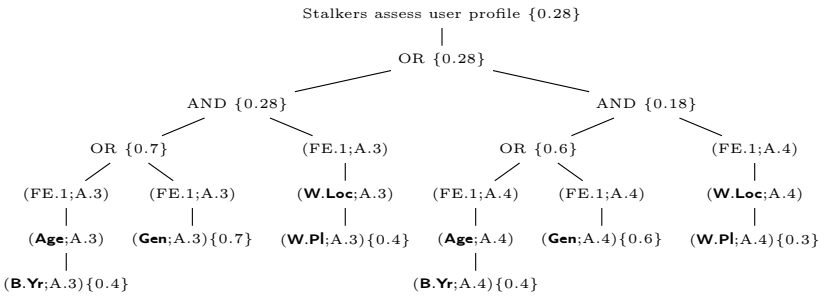

Fig. 8: Likelihood computation based on worst case harm tree for H.1 for T

5 Related Works

One of the earliest privacy scoring models is the one by Liu and Terzi [13]. In their work, privacy score is a monotonically increasing function of the visibility of attribute values and their sensitivity. It has been assumed that the privacy settings assigned to an attribute depend on its sensitivity and hence a response matrix that records the privacy settings of different attributes by a number of users has been used to estimate the value of sensitivity of each attribute. The visibility of the attribute value is influenced by the privacy setting of the user and his position in the network. The probability that an attribute is truly visible is estimated using the observed visibility values (i.e., the privacy settings) recorded in the response matrix using the Item Response Theory. In contrast, we do not assume that users consider the "sensitivity" of personal data when they specify their privacy settings, nor do we use sample data to compute the privacy scores.

Wang and Nepali [20] introduce the privacy index as a measurement of the exposure of the privacy of a participant in an OSN based on known attributes. In [22,15], they use it for their social network model for privacy monitoring and ranking. Both sensitivity and visibility of attributes are taken into account in the computation of the privacy index. We only consider visibility of attribute values as a contributor to the computation of privacy scores. The sensitivity of the attributes are implicitly revealed by their popularity in the harm trees. In the recent PScore framework [18], the scoring mechanism can be linked to any inference algorithm. Any inference algorithm could also be plugged into our method and the only adjustment required while doing so is to update the calculation of the accuracy value. However, in contrast to [18], our mechanism is concrete yet simple.

Some works [19,1,17] also focus on the rating of the user's OSN friends based on their attitudes towards privacy, helping him to make an informed decision of sharing information with them. We do not consider the ranking of the user's friends or the active disclosure of the target user's data by the risk sources, but rather focus on privacy risks that arise from what the target user or his friends willingly disclose about themselves. In our approach, the user does not need to provide any input that may require any awareness about privacy problems.

In most works, the implicit assumption is that if the user prefers to disclose or has no problems in allowing the propagation of some data then it is less sensitive to him than if he prefers otherwise. We assume that the user is not a privacy expert and may end up disclosing data that may cause him a lot of harm. Therefore, our privacy scores serve to warn the user about the imminent dangers of revealing personal data on the OSN. None of the previous works on privacy scores draw inspiration from privacy risk analysis.

Privacy harms, threats and risk sources specific to OSNs and their relationship with various personal data must be obtained from previous research. Information disclosed in OSNs can significantly affect others' impression of the user [16] and hiring decisions [16,12]. Other harms include thieves or sexual predators tracking, monitoring, locating and identifying a user as a potential victim, political parties targetting a user through ads and data mining [12] and identity theft[10]. OSN users often regret sharing information on alcohol and drug use, sex, religious and political opinions, personal and family issues, work etc., chiefly due to undesirable reactions from other users and unintended audience [21].

Our work is inspired by privacy risk analysis (PRA), a review of which can be found in [5,6]. Harm trees linking privacy weaknesses and risk sources to harms, via feared events have been introduced and widely used in [5,6,7]. Here, we adopt these concepts to our setting. PRA methodologies help the service provider to evaluate systems processing personal data for privacy risks, thus helping to design and implement these systems in the least privacy invasive way. Deng et al. [8] provide an example of using their LINDDUN risk analysis framework [9] for analyzing social networks. Our mechanism differs from these PRA methodologies in a number of ways: 1) our aim is to guide users instead of service providers; 2) we analyze each user's OSN profile and social graph to uncover the privacy risks, instead of the entire OSN system; 3) we consider risk sources that are already within the user's social graph and who process personal data that are already made visible to them by the user and do not consider hackers, OSN service providers, the government etc.; 4) we consider only the choices made by the user and his friends regarding the visibility of their profile attributes, but not privacy weaknesses [5,6] originating from the service provider's choices during system design and implementation (such as insufficient protection of data store, lack of anonymization techniques) 5) since OSN profiles are user-specific, counter-measures suggested based on the privacy scores will differ from user to user, based on privacy risks of their profiles and their requirements regarding social benefit. In addition, unlike [7], the harm trees do not consider system components (generic or specific) but only the data elements and the risk sources

and the pruning of harm trees takes place based on attribute visibility and the accuracy of the inferred attribute values rather than system architectures and the implementation context. To our best knowledge, our work is the first effort in utilizing PRA concepts for user-centric analysis of privacy risks of OSN profiles.

6 Conclusion and Future Works

We designed a privacy scoring mechanism for OSN profiles inspired by privacy risk analysis (PRA). The privacy scores can be used to inform the user about the privacy risks of his OSN profile. Our model can form the basis of designing a user interface to effectively communicate privacy scores and conduct a usability study to understand their effect on the user's privacy awareness. Based on the scores, we can also suggest counter-measures to users, taking into account the trade-off between the privacy risks and the social benefits of using OSNs. Such counter-measures include: 1) the selection of the right privacy setting for each profile attribute; 2) a decision on which friendships to continue based on their effects on the user's privacy scores and/or the negotiation of a privacy setting allowing both the user and his friends to maintain privacy and derive the social benefits of using an OSN. We leave these as future work.

References

1. Cuneyt Akcora, Barbara Carminati, and Elena Ferrari. Privacy in Social Networks: How Risky is Your Social Graph? In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*, pages 9–19. IEEE, 2012.
2. Faiyaz Al Zamal, Wendy Liu, and Derek Ruths. Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors. *ICWSM*, 270, 2012.
3. Commission Nationale de l'Informatique et des Libertes (CNIL). Privacy Impact Assessment (PIA) Methodology (How to Carry Out a PIA), 2015.
4. Commission Nationale de l'Informatique et des Libertes (CNIL). Privacy Impact Assessment (PIA) Tools (templates and knowledge bases), 2015.
5. Sourya Joyee De and Daniel Le Métayer. PRIAM: A Privacy Risk Analysis Methodology. In *11th International Workshop on Data Privacy Management (DPM)*. IEEE, 2016.
6. Sourya Joyee De and Daniel Le Métayer. Privacy Risk Analysis. In *Synthesis Series*. Morgan & Claypool Publishers, 2016.
7. Sourya Joyee De and Daniel Le Métayer. A Risk-based Approach to Privacy by Design (Extended Version). Number RR-9001, December, 2016.
8. Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. LINDDUN: running example-Social Network 2.0.
9. Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. A Privacy Threat Analysis Framework: Supporting the Elicitation and Fulfilment of Privacy Requirements. *Requirements Engineering*, 16(1):3–32, 2011.
10. Ralph Gross and Alessandro Acquisti. Information Revelation and Privacy in Online Social Networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80. ACM, 2005.

11. Lei Huang and Dan Wang. What a Surprise: Initial Connection with Coworkers on Facebook and Expectancy Violations. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, pages 293–296. ACM, 2016.
12. Maritza Johnson, Serge Egelman, and Steven M Bellovin. Facebook and Privacy: It's Complicated. In *Proceedings of the eighth symposium on usable privacy and security*, page 9. ACM, 2012.
13. Kun Liu and Evimaria Terzi. A Framework for Computing the Privacy Scores of Users in Online Social Networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(1):6, 2010.
14. Alan Mislove, Bimal Viswanath, Krishna P Gummadi, and Peter Druschel. You are Who You Know: Inferring User Profiles in Online Social Networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 251–260. ACM, 2010.
15. Raj Kumar Nepali and Yong Wang. SONET: A Social Network Model for Privacy Monitoring and Ranking. In *Distributed Computing Systems Workshops (ICDCSW), 2013 IEEE 33rd International Conference on*, pages 162–166. IEEE, 2013.
16. Ariane Ollier-Malaterre, Nancy P Rothbard, and Justin M Berg. When Worlds Collide in Cyberspace: How Boundary Work in Online Social Networks Impacts Professional Relationships. *Academy of Management Review*, 38(4):645–669, 2013.
17. David Pergament, Armen Aghasaryan, Jean-Gabriel Ganascia, and Stéphane Betgé-Brezetz. FORPS: Friends-Oriented Reputation Privacy Score. In *Proceedings of the First International Workshop on Security and Privacy Preserving in e-Societies*, pages 19–25. ACM, 2011.
18. Georgios Petkos, Symeon Papadopoulos, and Yiannis Kompatsiaris. PScore: A Framework for Enhancing Privacy Awareness in Online Social Networks. In *Availability, Reliability and Security (ARES), 2015 10th International Conference on*, pages 592–600. IEEE, 2015.
19. BS Vidyalakshmi, Raymond K Wong, and Chi-Hung Chi. Privacy Scoring of Social Network Users as a Service. In *Services Computing (SCC), 2015 IEEE International Conference on*, pages 218–225. IEEE, 2015.
20. Wenye Wang and Zhuo Lu. Cyber security in the Smart Grid: Survey and Challenges. *Computer Networks*, 57(5):1344–1371, 2013.
21. Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. I Regretted the Minute I pressed Share: A Qualitative Study of Regrets on Facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, page 10. ACM, 2011.
22. Yong Wang, Raj Kumar Nepali, and Jason Nikolai. Social Network Privacy Measurement and Simulation. In *Computing, Networking and Communications (ICNC), 2014 International Conference on*, pages 802–806. IEEE, 2014.
23. Ronald R Yager. OWA Trees and Their Role in Security Modeling Using Attack Trees. *Information Sciences*, 176(20):2933–2959, 2006.
24. Elena Zheleva and Lise Getoor. To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles. In *Proceedings of the 18th international conference on World wide web*, pages 531–540. ACM, 2009.