PScore: Enhancing Privacy Awareness in Online Social Networks

Georgios Petkos, Symeon Papadopoulos, Yiannis Kompatsiaris
Centre for Research and Technology Hellas, Information Technologies Institute
Thessaloniki, Greece
Email: {gpetkos,papadop,ikom}@iti.gr

***Abstract*—The phenomenal increase in the use of social media in recent years has raised a number of issues related to privacy. In this paper, we propose a framework for raising the awareness of Online Social Network (OSN) users with respect to the information about them that is disclosed and that can be inferred by OSN service operators as well as by third parties that can access their data. This framework takes the form of a semantic, hierarchical scoring structure, that enables users to easily browse over different privacy-related aspects of their presence in a social network. Contrary to previous privacy scoring approaches, the proposed framework provides a finer and more intuitive organization of privacy information. Importantly, it also takes into account both information that is explicitly mentioned in users' shared content, as well as implicit information, that may be inferred from it. We make available an open source implementation of the framework[1].**

***Keywords*-online social networks, privacy, privacy score, privacy control, privacy dimensions**

I. Introduction

Online Social Networks (OSNs) have experienced rapid growth in recent years. Current estimates report that there are almost 2 billion OSN users and, according to projections, this number will increase to 2.5 billion users by 2018[2]. The widespread use of OSNs has nevertheless brought forward the issue of privacy; a number of studies have shown that OSN users face a number of challenges with respect to privacy. For instance, in [17], 65 Facebook users were asked to carefully examine their profiles and it was found that all of them had at least one sharing violation, i.e. they were all sharing content with people that they really would not like to. The goal of the work presented in this paper is to propose a framework that will help OSN users increase their awareness with respect to their presence in the OSNs and make better decisions with respect to online information sharing.

However, privacy is to a certain extent a subjective concept: different users have different attitudes about revealing their personal information. For instance, [14] notes that people can often be classified into three main categories based on the overall degree of information disclosure: a) privacy fundamentalists, b) pragmatists, and c) unconcerned. Moreover, according to a follow-up study [15], information disclosure behaviour is not one-dimensional; instead, people can also be differentiated by what kind of information they tend to reveal. In addition, a major issue about privacy is the fact that information about a user may not only appear in an explicit manner, but it can also appear implicitly and may be obtained using appropriate inference mechanisms. For instance, one might easily guess that a user that is interested in university/educational issues is very likely to be a young adult. Inferred information is particularly important for conveying institutional notions of privacy to the user, i.e. helping the user understand what the OSN itself can infer about the user. The approach that we propose for enhancing privacy awareness in this paper takes the form of a privacy scoring framework and is designed with these considerations in mind. In particular, we identify the following basic requirements for the scoring framework:

1) It must take into account the fact that *privacy concerns are likely to differ between users* and therefore it is important to consider each user's personal preferences in order to compute privacy scores.
2) It must recognize that *different types of information have different significance to the user* and therefore the framework should be structured according to the different types of information.
3) It must also *take into account inferred information* and should be generic enough so that different inference mechanisms can be added to it and extend it in a seamless manner. Also, ideally, it should be able to link the inferences made to the specific OSN presence data that support the particular inference.

The main contribution of this paper is that it presents the first OSN-oriented privacy scoring framework that recognizes the above requirements and that attempts to deal with them. In addition, we provide an open source implementation of the proposed scoring framework. It is important to note that the implementation can easily be extended in order to cater for additional privacy related attributes and inference mechanisms.

The rest of the paper is structured as follows. The next section provides a typology of personal information, focusing on types of information that may be considered by individuals and by law to be sensitive and/or valuable. Then, section III reviews previous approaches on privacy scoring in OSNs. Section IV presents the proposed scoring framework - which we will refer to as PScore - and section V concludes the paper.

II. Typology of private information and data

We start by discussing specific types of information that can be considered private or sensitive. This will allow us

[1] https://github.com/MKLab-ITI/usemp-pscore
[2] http://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

978-1-4673-6590-1/15 $31.00 © 2015 IEEE
DOI 10.1109/ARES.2015.80

to identify the information that we need to maintain and then to organize them in a semantic manner.

It can be argued that there are two high-level notions of privacy. The first is related to what most people *perceive* as private information. This is typically rather fuzzy, as different people can have very different perceptions about what kind of information is considered private. The second high level notion of privacy is related to *legal* definitions of privacy and personal information.

Let us first discuss the general, non-legal, notion of privacy. One might argue that almost any type of personal information may be considered by different individuals to be private, sensitive or valuable under specific circumstances. For instance, different types of information about the profile of an individual may be valuable to a marketing company for promotion reasons. Additionally, there have been many cases in which people became victims of discrimination on the basis of their traits or past actions [1], even in cases where such discrimination was not justified. In general, our study of related work that is presented here suggests that personal information is generally considered of private or sensitive nature in cases that:

- The information can be used for unjustifiable *discrimination* in a variety of social, cultural, professional and other settings. For instance, information about the gender, age, ethnicity, political or religious beliefs, sexual preferences, and financial status of a person were used for unjustified discrimination in the context of personnel selection [1], as well as for loan approval and pricing [2].
- The information may be used for *manipulation* of the opinion/beliefs of the person him/herself or the opinions of others about him/her. In the more common case, this includes information that is often extensively used by third parties for profiling and targeting (in case of ad campaigns).
- The publication of that information may have *detrimental effects* on the mental, physical and economic state of the individual, e.g., threats to their residence privacy, stalking [3], identity theft, etc.

Considering the legal notion of privacy, it is recognized that there are two kinds of laws under different countries' legislations that are related to the type of data that we are considering here. The first is about the protection of personal information. This defines specific rules and requirements regarding the process of personal data processing. The data protection law typically also specifies that users should provide their explicit consent for the collection, storage and processing of sensitive data. Types of information that are considered sensitive by the data protection law include information about a person's racial/ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, health status and sex life. The second kind of related laws concern anti-discrimination. Types of information that are protected by anti-discrimination laws in different countries include the following: sex, gender identity and sexual orientation, age, race and ethnicity, nationality, disability, religion, world-views and political opinions.

In addition, it should be noted that there are other definitions of personal data: for instance, the Privacy Guidelines published by the Organization for Economic Cooperation and Development (OECD) define that the term personal data describes "any information relating to an identified or identifiable individual (data subject)". Characteristic categories of personal data under this definition include user generated content, activity or behavioral data, social data, location data, demographic data, or data of an official nature, e.g., financial information and account numbers, and health information [13].

In the discussion so far we have effectively identified different types of personal information that can be characterized private, personal or sensitive in a variety of contexts. Since our focus is on OSNs, it is useful to also discuss a taxonomy of personal data in relation to OSN services that was proposed by Schneier [4]. This taxonomy considers the source of data about a user, rather than the type of personal information, and will be useful for the development of the scoring framework. Briefly, Schneier identifies the following six categories of OSN data:

- *Service data.* This is the set of data that a user explicitly provides to the OSN. In many cases, this includes the user's legal name, age, gender, etc.
- *Disclosed data.* This includes the content (messages, status updates, photos, etc.) posted by the user to his/her own page.
- *Entrusted data.* This is the content posted by the user to the page of another user. It is similar to disclosed data, with the difference that, in many cases, the user does not have full control of the content, but some other user does.
- *Incidental data.* This is the content posted about the user by some other user (e.g., when a friend of the user posts a picture depicting the user). Again, this is similar to disclosed data, but again, the user does not have full control over such data.
- *Behavioural data.* This type of data includes the actions of the user in the OSN. For instance, this may include information about the profiles the user interacts with, what games he/she plays, what pages he/she likes, etc.
- *Derived data.* This is data about a user that may be derived from all other types of data, typically by means of algorithmic processes. We will also refer to such kind of data as *inferred*.

Clearly, this is not the only way by which data in OSNs can be organized. As mentioned, it primarily focuses on the source of the data; another possible organization could focus on the semantics of the data about a user, this is the type of top-level organization we adopt in the PScore framework. It should be noted though that Schneier's taxonomy identifies that the level of control a user has over the data that concern him/her may vary significantly depending on the above categories, for instance the user typically has full control over service and disclosed data,

while limited control over entrusted and incidental data and almost no control over derived data. This fact has been considered in the development of the scoring framework.

III. Existing privacy scoring methods

Here, some existing work in the field of privacy scoring models is reviewed. The number of relevant papers is rather limited, indicating that the field is still relatively unexplored. Additionally, most of the work is quite recent, indicating the emerging interest in the problem.

One of the first works in the field of privacy scoring comes from Liu and Terzi [5], [6]. The authors introduced the concept of *Privacy Scores*, a function of the *sensitivity* and the *visibility* of different pieces of profile information. Sensitivity (in the sense used in the paper) denotes how important it is to protect this particular piece of information (e.g., age, gender, etc.) and is computed by analysing the results of user studies with respect to the disclosure of such information (the more people are willing to disclose a piece of information, the less sensitive it is considered). Moreover, visibility quantifies the extent to which it is accessible to other users. Formally, assuming a user j has n profile items, the Privacy Score PR is computed using Equation 1.

$$PR(j) = \sum_{i=1}^{n} PR(i,j) = \sum_{i=1}^{n} \beta_i V(i,j) \tag{1}$$

where β_i and $V(i,j)$ are respectively the sensitivity and the visibility of profile item i for user j. In order to compute the scores, Liu and Terzi use as input a $n \times N$ response matrix R (with n being the number of profile items and N the number of considered users), which expresses how willing a user is to disclose some profile item. Given this formulation, a simple statistical model grounded on Item Response Theory and Maximum Likelihood Estimation is used to compute the Privacy Scores.

The work by Liu and Terzi introduced the concepts of sensitivity and visibility. Our own formulation will also use these concepts and will also compute aggregate privacy scores as products of these factors (but not only these factors). It should be noted though that, while the sensitivity of profile item i (β_i) is assumed to be the same across all users in [5], in our framework, as will be shown in the next section, we enable each user to specify a different sensitivity score for each attribute. This way we cover the first requirement that we listed in the introduction, that is to take into account the fact that privacy concerns differ between different people. Also, the model by Liu and Terzi focuses only on information that is explicitly available on the profile of the user, whereas our framework also produces scores in cases where the information is implicitly made available or inferred. This way we cover the third requirement.

An approach that introduces an extension to the Privacy Scores of Terzi and Liu is presented by Srivastava [7]. Apart from information explicitly provided by the user as part of his/her profile, Srivastava also examines textual messages in order to extract explicitly mentioned pieces of sensitive information, such as address, e-mail, location, etc. It should be noted though that this involves simple pattern detection in textual data and it does not involve any sophisticated inference mechanism. Srivastava uses the same sensitivity and visibility computation that Terzi and Liu used to produce a score; however, they call it *Privacy Quotient*. They also introduce a new measure, called *Privacy Leakage*, which is applied to a single message/piece of content and quantifies how much of the privacy exposure for some user is due to that particular message. It is computed by dividing the sensitivity for the message by the sum of sensitivities over all messages.

Another recent study comes from Domingo-Ferrer [8]. This work introduced the *Privacy-Functionality Score* that quantifies how much information a user reveals compared to other users. It utilizes the Privacy Score, as defined in Equation 1, and is defined in Equation 2.

$$PRF(j) = \frac{\sum_{x=1, x \neq j}^{n} PR(x)}{1 + PR(j)} \tag{2}$$

This ratio allows the OSN users to be easily ranked in terms of how much they reveal compared to other users.

Another work that examined the problem of privacy scoring comes from Nepali [9], [10]. Nepali introduces the *Privacy Index*. In particular, considering that there are items that are published and others that are not, it defines the Privacy Index in Equation 3.

$$PIDX = \frac{\sum_{k \in K} S_k}{\sum_{i \in I} S_i} \tag{3}$$

where K is the set of published items, I is the set of all items and S is the sensitivity of the item. The Privacy Index is somewhat similar to the leakage score of [7]. It is clear though that while the leakage score described how much of the information leakage about a user is due to some specific message/piece of content, the Privacy Index describes how much of the information that is sensitive and the user has disclosed to the OSN operator, has been also made public.

Finally, recent research [11] has extended the Privacy Scores of [5] to consider: a) information from multiple OSNs, and b) information posted anywhere on the Web and retrievable via search engines. Table I summarizes the presented privacy scoring approaches.

IV. Proposed Privacy Scoring Model

This section presents the proposed PScore framework. We start by organizing the personal attributes that can be considered as private or sensitive in a number of high-level categories that we refer to as *privacy dimensions*. This organization allows for a semantic and intuitive presentation and handling of the different aspects of a user's personal information. For instance, one of the privacy dimensions that will be considered is demographics, which includes user attributes such as age, gender, etc. Another privacy dimension is about health factors, which includes attributes such as smoking and drinking, etc. Such a grouping has multiple benefits for the end user. In particular, it enables

Table I
OVERVIEW OF STUDIED PRIVACY SCORING APPROACHES

Score	Description	Elements
Privacy Score [5]	A score grounded on the sensitivity and visibility of the items posted by an OSN user.	Profile items, sensitivity per item, visibility per item.
Privacy Quotient and Leakage [7]	Extension of Privacy Score with a focus on text messages and on attributing part of score to specific messages using the Leakage score.	Profile items (text), sensitivity per item, visibility per item, leakage per item.
Privacy Functionality Score [8]	Extends the Privacy Score by normalizing it over the total privacy exposure of all users, i.e. it offers a comparative privacy view.	Privacy score per OSN user, privacy score of other OSN users.
Privacy Index [9], [10]	Captures the portion of sensitive data posted to a social network that has also been made public.	Sensitivity score per item, visibility level (public/private) per item.
Privacy Scores [11]	Extension of Privacy Score to include data from multiple OSNs and information retrievable via search engines.	Privacy score per item per OSN, information retrievable via search engines.

him/her to form a succinct, easy to grasp mental model of his/her private information and to prioritize its different parts. It also helps to cover the second requirement that was mentioned at the introduction. On top of the privacy dimensions framework, we develop the scoring model, by enriching it with privacy scores.

At a glance, the proposed privacy scoring framework consists of multiple scores that reflect quantities such as the sensitivity, visibility, etc. of different privacy dimensions and attributes. In addition to maintaining a number of distinct scores, each of which reflects a distinct aspect of privacy, we also compute aggregate privacy scores in order to end up with concise and simple-to-grasp privacy indicators. Providing different types of scores that reflect distinct aspects of privacy allows users to obtain a more detailed and focused perception of their privacy status.

A. Privacy dimensions

Based on the discussion of Section II we compiled a set of *personal attributes* that can be considered private or sensitive according to either a general perception of privacy or a legal definition. Subsequently, we identified eight key categories of personal attributes, which we name *privacy dimensions*. These include: A) Demographics, B) Psychological Traits, C) Sexual Profile, D) Political Attitudes, E) Religious Beliefs, F) Health Factors and Condition, G) Location and H) Consumer Profile. Table II summarizes the eight identified privacy dimensions, along with the attributes under each of them and a short discussion of potential threats that could be entailed if the particular type of information was accessible by some inappropriate audience.

Each attribute can take a number of values. It is important to note that it may not be possible to detect all possible values using the set of available inference mechanisms. For instance, an inference mechanism that is used for the "Family status" attribute may be able to detect only the values "Single" and "Married", but not values such as "Divorced". To give a more concrete example, Table III presents possible value sets for the the attributes under the Demographics dimension.

It should also be noted that while effort was made to come up with a relatively extensive list of representative attributes for each dimension, this list is very likely to be enriched and amended in the future.

The above effectively creates a hierarchy in which the top level represents the overall OSN personal data profile, at the next level there is a number of privacy dimensions, each privacy dimension has a number of attributes and each attribute can take one or more out of a set of possible values. This formulation will be the basis of our privacy scoring model that is described next.

B. Privacy scoring

The proposed privacy scoring mechanism enriches the privacy dimensions hierarchy with several scores, each reflecting a different aspect of personal information disclosure. Additionally, overall privacy scores are computed at each level of the hierarchy, providing a summary of privacy issues for each dimension, attribute or value. Clearly, the two important characteristics of this framework are the following: a) it is tailored to the hierarchical structure of the privacy dimensions, b) there are multiple scores associated with the elements of each level of the hierarchy. Hence, the framework enables the following two kinds of user awareness: a) navigation through the levels of the hierarchy and understanding of how the scores for some particular value affect or are affected by the levels above and below it, and b) focus on specific aspects of the factors that are related to privacy; e.g., it will be possible to focus on visibility, sensitivity, the overall privacy score, etc.

Computation of scores at the values level is based on a set of information extraction and inference mechanisms. To more explicitly represent this, we consider an additional level at the privacy dimensions framework, below the values level, which contains any type of data that is generated as a result of a user's behaviour and interaction with the services of an OSN operator. This includes posted content (text, images), explicitly declared profile information, user network data, sets of likes, etc. We call this the *OSN presence data layer* and consider it as the primary source for populating the privacy scores for the given user. Naturally, between the privacy values level and the online presence data, there is a layer of modules that perform various mining and inference procedures. For instance, the method in [18] utilizes the likes of users and the method in [19] utilizes user interactions to infer the values of different user attributes.

It should be noted that a number of inference mechanisms will gradually be made available at the repository in which the implementation of the PScore framework is

Table II
OVERVIEW OF PRIVACY DIMENSIONS

#	Dimension	Attributes	Threats-Sensitivity
A	Demographics	1) Age 2) Gender 3) Nationality 4) Racial origin 5) Ethnicity 6) Literacy level 7) Employment status 8) Income level 9) Family status	Discrimination in a variety of settings. The most frequently used type of information.
B	Psychological Traits	1) Emotional stability 2) Agreeableness 3) Extraversion 4) Conscientiousness 5) Openness	Discrimination, e.g. in personnel selection
C	Sexual Profile	1) Sexual preference	Discrimination, e.g. in workplace, education, housing
D	Political Attitudes	1) Supported party 2) Political ideology	Discrimination, e.g. in workplace or personnel selection.
E	Religious Beliefs	1) Supported religion	Discrimination, e.g. in house sale/rental, job selection, workplace.
F	Health Factors and Condition	1) Smoking 2) Alcohol drinking 3) Drug use 4) Chronic diseases 5) Disabilities 6) Other health factors	Discrimination, e.g. health insurance denial and/or discriminatory pricing.
G	Location	1) Home location 2) Work location 3) Favourite places 4) Visited places	Discrimination, e.g. house insurance, stalking.
H	Consumer Profile	1) Favourite brands 2) Hobbies 3) Devices	Ad targeting and discrimination in online price-setting.

Table III
DEMOGRAPHIC ATTRIBUTES

#	Attribute	Example values and range
A.1	Age	Rather than using the absolute number of years, it is reasonable to use age groups, e.g.: 6-12, 12-18, 15-25, 25-35, 35-45, 45-55, 55-65, 65-75, older than 75 years
A.2	Gender	Male, Female
A.3	Nationality	French, Belgian, Italian, etc.
A.4	Racial origin	Asian, African, Caucasian, Latino/Hispanic, Other
A.5	Ethnicity	List of target ethnicities, e.g. Arabic, Easter-European, etc.
A.6	Literacy level	None, Nursery school, High school, Bachelor's degree, Master's degree, Ph.D., Other
A.7	Employment status	Employed, Unemployed, Retired, Other
A.8	Income level	Qualitative ranges of monthly income. e.g., a 5-scale range from low to high.
A.9	Family status	Single, married, divorced, other

made available. In addition, our implementation makes sure that it is straightforward to plug in new inference mechanisms.

The proposed scoring framework is schematically displayed in Figure 1. In short, it has the hierarchical structure of the privacy dimensions framework and assigns a set of scores to each node of the hierarchy. We will now present the various scores that will be assigned to the nodes of the main four layers of the hierarchy (user, dimensions, attributes, values). It should be noted that some of those cannot be considered as scores per se, they may rather be considered as additional fields that enrich the representation of the scoring model. These appear at the values layer; for instance, there is a field called "Declared/Inferred" that simply states whether knowledge about the particular value has been explicitly provided by the user or it has been inferred. Scores are computed in a bottom-up manner, i.e. the OSN presence data are used by the inference mechanisms in order to fill in the scores at the values level, then the scores at the values level are used to compute the scores at the attributes level and so on until eventually scores for the overall user profile are computed. Starting from the level of values, the scores that characterize each value are the following:

Confidence. This is a continuous value in the range from 0 to 1. It represents how confident we are that the corresponding value is true and is typically computed by the inference algorithm along with the produced inference. It needs to be noted that the confidence values under the same attribute should sum to 1. This is due to the fact that the confidence expresses our degree of belief that the value is true and therefore for mutually exclusive values the confidence should be normalized. It should be noted though that there are attributes that can take more that one values simultaneously and that the normalization constraint does not hold for these attributes. An attribute

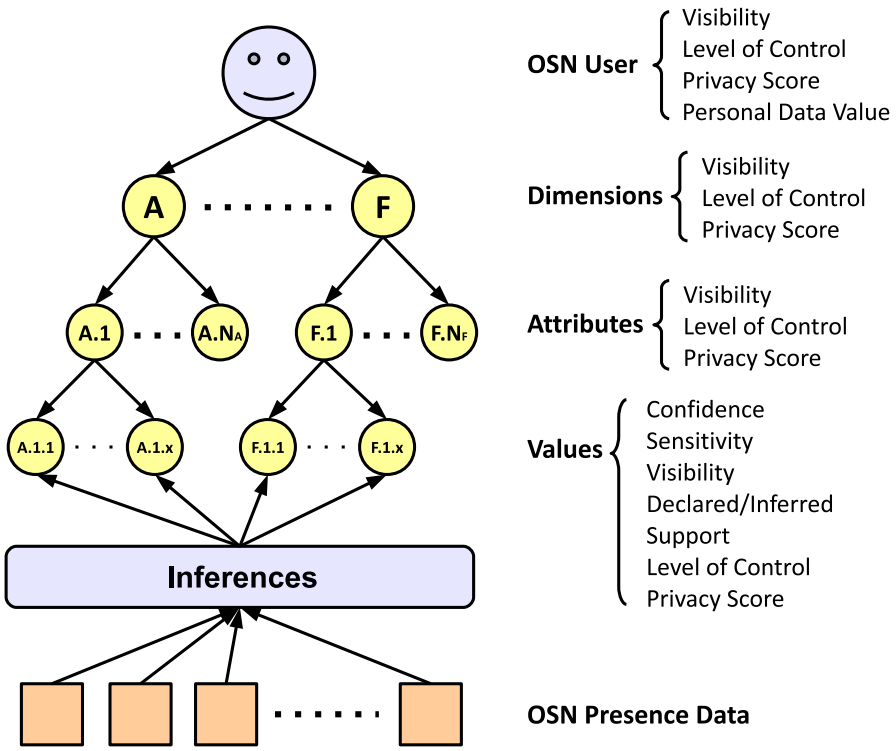

Figure 1. The privacy dimensions framework.

whose possible values are mutually exclusive is "family status", whereas as attribute that can take multiple values is "favourite places".

Sensitivity. A continuous value in the range from 0 to 1, with higher scores corresponding to higher sensitivity levels. It reflects how important it is to protect this piece of information. An important decision that was made was to define sensitivity scores at the values level, rather than at the attributes level. The reason for this is that for many attributes, the sensitivity for some values will be different compared to others; e.g., the sensitivity of the value "homosexual" is expected to be higher for many users than the one of the value "heterosexual".

Visibility. This reflects how accessible this specific piece of information is to other people. This score is expressed with three individual sub-scores. The first is the *overall visibility score*, a continuous value in the range from 0 to 1. A score of 1 denotes that this piece of information is accessible to everyone, whereas a score of 0 denotes a piece of information that is accessible only to the user (private). The overall visibility score depends on the OSN privacy settings on the specific content that has been used to identify the relevant value. The second visibility sub-score is a qualitative label, that we call the *visibility label* and expresses the widest possible audience to which this information is accessible. For instance, a value with an overall visibility score of 0 will have the visibility label "Private", a value with score 1 will have the label "Public" and an intermediate value will denote the widest group of people that have access to the value, e.g. "Friends", "Custom set of friends", etc. The third visibility sub-score expresses an estimate of the actual audience that sees this value and we refer to it as the *actual visibility*. It is an integer number representing the actual number of users that are aware of that value and depends on the estimates of the actual audience of the content that has been used to infer that value. Importantly, as will be explained later, out of these three sub-scores, only the first will be used for computing the overall privacy score and the other two will be used only for providing supplementary information to the end user.

Declared/Inferred. This is a binary value that defines whether our knowledge about the particular value comes from explicitly provided information that the user provided or was inferred by some algorithm. It is not an actual score but reflects information that is important for maintaining a complete view of privacy with respect to some particular value. Declared values will have a confidence of 1; however, there could be inferred values with confidence 1. Additionally, in some cases a value may be both declared and inferred. In such cases, the value will be considered as declared (i.e. declared will override inferred).

Support. Just as Declared/Inferred, this is not a score per se, it is rather a structure that points to the OSN presence data based on which the other score dimensions have been filled. If the value is declared, then this points to a single item; however, if the value is inferred then this may point to a list of items. More particularly, if the value is inferred,

this may point to different types of data, depending on the employed inference mechanisms. For instance, in some cases this may point to textual data, in some other it may point to images or to the network around the user. This field will allow the user to obtain a justification about why the system believes that certain variables apply to him/her.

Level of control. This score represents the ability of a user to control the disclosure of data about him/her. It ranges from 0 to 1; low values denote a limited ability to control the disclosure of this particular value, due to the involvement of entrusted and incidental data (cf. Schneier's taxonomy in section II).

Privacy score. This is a score that reflects the overall privacy exposure of a user with respect to a particular privacy value. It summarizes the other scores (confidence, sensitivity, visibility). It ranges from 0 to 1 and higher values denote higher privacy exposure. Note that although the privacy score essentially summarizes the other scores, the model maintains a separate list of the individual scores (confidence, sensitivity, visibility) in order to allow the user to examine different aspects of privacy.

All three upper levels of the PScore framework, namely the user, the dimensions and the attributes, are associated with the following set of scores: a) Visibility, b) Privacy score and c) Level of control. These have similar meaning to the corresponding scores at the value level.

C. Computation of scores

In the following we examine how these scores are actually computed. As mentioned, we follow a bottom-up aggregation strategy. Thus, in the following we examine how each component of the score at each level is computed from data that is available from the level below it. Let us start with the computation between the raw OSN data and the values level.

Confidence. There are two cases for filling in the confidence value. The first is when we examine a declared value, in which case the confidence is set to 1, regardless of any inferences made with respect to the same value. The second is when we examine an inferred value. Importantly, there may be multiple inference mechanisms each of which may process a subset of the available data: for instance, Natural Language Processing (NLP) techniques may be utilized to process the user's text posts, deep learning approaches may be utilized to process images posted by the user and label propagation techniques may be utilized to process information coming from the network around the user. Moreover, some inference algorithms may be applied on multiple pieces of data. Therefore, we need a mechanism for aggregating the predictions of the multiple inference mechanisms and for different subsets of the data.

It can be observed that different privacy values may be reflected on only a subset of the data. Therefore, if for example only a small subset of the data reflects the fact that the user is homosexual, then the confidence obtained from examining this subset should not be significantly decreased by examining other pictures that do not reflect this. An additional complication arises from the fact that the confidence values for most of the attributes should sum to one. If we had a single inference mechanism working on an individual set of data, this constraint would be straightforward to handle. However, considering that we will typically have many different inference mechanisms and some will operate on multiple data items, we should be careful with respect to how we ensure the normalization constraint when dealing with attributes that take mutually exclusive values. Let us make the issue more clear with a simple example. Consider the attribute "sexual orientation" with possible values "heterosexual", "homosexual" and "bisexual". Some algorithm may provide the confidence scores 0.1, 0.8 and 0.1 for the three possible values, whereas some other algorithm that considers some other piece of data (that possibly reflects something different) may give confidence scores of 0.8, 0.1 and 0.1. If we perform aggregation using the maximum operator at value level, then we will end up with confidence scores 0.8, 0.8 and 0.1. These scores could be normalized by their sum (we would then obtain the scores 0.47, 0.47 and 0.06 respectively); however, this is likely an unrealistic assumption in many cases. The proposed solution is to use only the confidence scores provided by the single inference mechanism or piece of data that maximizes the intermediate privacy score of Equation 4.

$$\sum_i sensitivity_i \cdot confidence_i \tag{4}$$

where i ranges through the different values an attribute may take. The reason is that the confidence scores provided by a single inference mechanism for the values of a particular attribute will be normalized, therefore, it makes sense to select the confidence provided by a single inference mechanism. The score in the previous equation is a sort of temporary privacy score and can be used in order to select the most prominent inference mechanism for some particular attribute.

Sensitivity. Sensitivity scores of different values can be obtained in two ways. The first is to use direct user input. The user is given the possibility to either set explicit sensitivity scores for each privacy value or, for efficiency, to do it in a top-down manner where for example he/she gives a sensitivity score for some privacy dimension or attribute, which is then propagated down the hierarchy. The possibility for users to directly set their sensitivity scores in different parts of the model is an important empowerment tool and turning privacy management into an ongoing and organic process of negotiating the boundaries of disclosure, identity, and time [12]. This effectively covers the first requirement that was listed in the introduction. The second way by which sensitivity scores can be obtained involves prior knowledge about the sensitivity scores of an "average" user. There are several user studies that consider the relative importance of different attributes, and those could be used to set initial values to the attributes contained in the framework. Moreover, the use of an approach similar to that of [5] is considered for adoption in the future. That is, we may opt for estimating the

sensitivity scores automatically from a response matrix R. As mentioned, the response matrix expresses how willing a user is to disclose information about some personal item.

Visibility. Appropriately setting the visibility scores is crucial as a means to enhance users' awareness with respect to privacy exposure risks. Driven by these requirements, the overall visibility score is primarily computed by taking into account the privacy settings of the OSN data that support a particular value. If the support comes from declared or inferred data that is publicly available, then the visibility is set to 1, and if the support comes from data that is private, then it is set to 0. Intermediate scores are computed using a monotonically increasing function that may depend on the size of the neighbourhood of a user to which the data is visible. In our implementation this is a simple ratio but another option is to pass this ratio through an appropriate non-linear function, such as a sigmoid. The visibility label is computed by considering the size of the audiences to which the data used to infer the value is accessible. For instance, if an inference relies on the use of two pieces of content: a piece that is accessible only to friends and another piece that is publicly available, then the visibility label will be "Friends", since only friends have access to both pieces of content in order to perform the same inference. The computation of the actual visibility score, that quantifies the size of the audience that actually is aware of that value, entails high uncertainty and is still work in progress. Different proxies based on the number of likes and comments that a piece of content received are considered.

Declared/Inferred. This field is directly filled simply by checking whether the value of specific attributes is provided directly by the user or not.

Support. This field links the value with specific OSN presence data that indicate that the value is true for the user. In the case of a declared value, this field directly points to the relevant field in the user profile. Otherwise, it points to the data that was used to draw the inference. It is reminded though that multiple inference mechanisms may be used and that some inference mechanisms will work on multiple data items. Therefore, the support field may include multiple records, each of which is related to a specific inference mechanism and data. Each such support record contains a pointer to the data, the identification of the inference mechanism used and a distinct confidence score (used in order to compute the value's confidence score as described previously).

Level of control. As mentioned, this score expresses the ability of the user to control the disclosure of data about him/her. Referring to Schneier's taxonomy, a user does not have full control over entrusted and incidental data. Thus, the level of control will be computed as one minus the ratio of the number of support items that are entrusted or incidental over the number of all support items. More formally, it is given by Equation 5.

$$control = 1 - \frac{entrusted + incidental}{all} \qquad (5)$$

Privacy score. The overall privacy score summarizes the other scores. It is a monotonically increasing function of aggregated sensitivity, visibility and confidence. In our implementation we consider different alternatives that the user can select. In its simplest form it is the product of the value's sensitivity, visibility and confidence. Alternatively, we modulate it with the help of an appropriately shaped logistic function (whose parameters can also be adjusted by the user).

Once scores are computed at the values level, computation at the three upper levels is straightforward. In particular, only the overall privacy score, the level of control and visibility are considered for the three upper levels. For the privacy and visibility scores, the same strategy is employed for computations between any pair of levels. In particular, these scores are computed as the averages of the corresponding scores at the level below. Another option is to pass the averages through an appropriately shaped non-linear function, possibly with the goal to boost the privacy score, and thus to increase privacy awareness. The visibility labels are aggregated in a different manner. In particular, they are aggregated using a `max` operator. Finally, the level of control is handled by a `min` operator, since we are interested in highlighting attributes and dimensions, where the user has reduced (minimum) control.

V. Conclusions

This paper presented PScore, a scoring framework designed to raise the awareness of OSN users with respect to privacy. All in all, the proposed scoring framework organizes privacy related information in a semantic manner and is associated with a number of scores, each focusing on a different aspect of privacy. The framework has three distinct characteristics that cover the requirements that were listed in the introduction and that differentiate it to other scoring frameworks. First, it considers the user's personal preferences by allowing them to define their own sensitivity scores. Second, it is based on a hierarchical structure that provides a semantic organization of information in a number of dimensions, attributes and values. Third, it is able to deal with information that is not explicitly present in the OSN, but may be inferred or extracted using data-driven mechanisms. Inferred information is particularly useful to help the user understand what the OSN operators may know about him/her from analysing the user's OSN presence data. In fact, the contribution of the proposed framework, as compared to previous approaches to privacy scoring, is the fact that it recognizes these three requirements and it attempts to address them. Additionally, an open source implementation of the framework is made available, to which it is straightforward to add new inference mechanisms.

Ongoing work focuses on developing a number of detection and inference mechanisms that will complete the computational aspects of the framework. In fact, it should be stressed that the used inference mechanisms are crucial for the performance of the framework and that more

accurate inference mechanisms are likely to significantly improve its usefulness. In this paper, we focused on the scoring framework and its desired characteristics, but the development of appropriate privacy related inference mechanisms is equally important. Yet, the implementation of the scoring framework is independent of particular inference mechanisms and can therefore be easily linked to new inference algorithms. In the future we also intend to develop user interface and interaction mechanisms that will support the communication of the scoring framework to end users with the vision of eventually creating an integrated platform that will provide a complete OSN privacy assistance tool. User-based evaluations of the framework are also foreseen once an end-to-end system is available.

ACKNOWLEDGEMENTS

This work is supported by the USEMP FP7 project, partially funded by the EC under contract number 611596.

REFERENCES

[1] Acquisti, A., Fong, C. M.: An Experiment in Hiring Discrimination Via Online Social Networks. Social Science Research Network Working Paper Series, Apr. (2012)

[2] Raman, A. S., Barloon, J. L., Welch, D. M.: Social media: Emerging fair lending issues. The Review of Banking and Financial Services, 28(7) (2012)

[3] Haron, H., Yusof, F.B.M.: Cyber stalking: The social impact of social networking technology. International Conference on Education and Management Technology. pp. 237-241. (2010)

[4] Schneier, B.: A Taxonomy of Social Networking Data, Security & Privacy, IEEE , vol.8, no.4, pp.88,88, July-Aug. (2010)

[5] Liu, K., Terzi, E.: A framework for computing the privacy scores of users in online social networks. In ICMD (2009)

[6] Liu, K., Terzi, E.: A framework for computing the privacy scores of users in online social networks. In ACM Transactions on Knowledge Discovery from Data. Vol. 5. No. 1. Article 6. (2010)

[7] Srivastava, A., Geethakumari, G.: Measuring privacy leaks in online social networks. ICACCI 2013. (2013)

[8] Domingo-Ferrer, J.: Rational privacy disclosure in social networks. Modeling decisions for artificial intelligence. LNCS. Volume 6408 (2010)

[9] Nepali, R.K., Wang, Y.: Sonet: A social network model for privacy monitoring and ranking. ICDCS 2013 (2013)

[10] Wang, Y., Nepali, R., Nicolai, J.: Social network privacy measurement and simulation. In ICNC 2014. (2014)

[11] Sramka, M.: Evaluating privacy risks in social networks from the user's perspective. Advanced Research in Data privacy. Studies in computational intelligence 517 (2015)

[12] Hong, J. I., Ng, J. D., Lederer, S., Landay, J. A.: Privacy risk models for designing privacy-sensitive ubiquitous computing systems. In Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques (pp. 91-100). ACM. (2004)

[13] OECD: Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value. OECD Digital Economy Papers, No. 220, OECD Publishing. http://dx.doi.org/10.1787/5k486qtxldmq-en (2013)

[14] Knijneburg, B.P., Koba, A., Jin, H.: Dimensionality of information disclosure behavior. International Journal of Human-Computer Studies. 71(12), 1144-1162. (2013).

[15] Knijneburg, B.P.: Information Disclosure Profiles for Segmentation and Recommendation. In Symposium on Usable Privacy and Security (SOUPS). (2014)

[16] Johnson, M., Egelman, S., Bellovin, S.: Facebook and privacy: it's complicated. Proceedings of the 8th Symposium on Usable Privacy and Security. (2012)

[17] Madejski, M., Johnson, M., Bellovin, S.: A study of privacy settings errors in an online social network. Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on. IEEE. (2012)

[18] Kosinski, M., Stillwell, D.J., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences (PNAS). (2013)

[19] McPherson, M., Smith-Lovin, L., Cook, J. M.: Birds of a feather: Homophily in social networks. Annual review of sociology, pp. 415-444. (2001)