This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2720187, IEEE Access

# Scoring Users' Privacy Disclosure Across Multiple Online Social Networks

Erfan Aghasian, Saurabh Garg, *Member, IEEE*, Longxiang Gao, *Member, IEEE*, Shui Yu, *Senior Member, IEEE* and James Montgomery *Member, IEEE*

**Abstract**—Users in online social networking sites unknowingly disclose their sensitive information that aggravate the social and financial risks. Hence, to prevent the information loss and privacy exposure, users need find ways to quantify their privacy level based on their online social network data. Current studies that focus on measuring the privacy risk and disclosure, consider only a single source of data, neglecting the fact that users in general can have multiple social network accounts disclosing different sensitive information. In this paper, we investigate an approach that can help social media users to measure their Privacy Disclosure Score (PDS) based on information shared across multiple social networking sites. In particular, we identify the main factors that have impact on users privacy, namely, sensitivity and visibility, to obtain the final disclosure score for each user. By applying the statistical and fuzzy systems, we can specify the potential information loss for a user by using obtained PDS. Our evaluation results with real social media data show that our method can provide a better estimation of privacy disclosure score for users having presence in multiple online social networks.

**Index Terms**—Privacy, Social networks, Measurement, Fuzzy logic.

## 1 Introduction

Online social network sites have changed from a niche phenomenon to mass acceptance [1]. While the distribution of information in real world is almost local, the publically shared information in online social media can be retrieved on the internet anytime, anywhere and by anyone [2]. Individuals are able to make connections, exchange information, express their feelings and form and preserve relationships with other individuals on the Internet [3]. Facebook, LinkedIn, Google+, Twitter and other online social networks all have different advantages, both professional (such as sharing one's employment record in LinkedIn) and social (such as connecting with distant friends via Facebook) [4]. This presence of individuals in online social networks creates a trade-off between the possibility of expanding their social and professional circles, and privacy risks. Users provide a variety of information (sensitive and non-sensitive) that may be disclosed to other individuals. Information related to individuals that is shared can include spatial-temporal items such as their location and time-stamp, and personal characteristics such as personal background, hobbies, contacts, personal views and so on. This information sharing can be a cause of potential risks for individuals in online social networks, including identity theft, sexual abuse, stalking, employment, online victimization, surveillance and unintentional fame and even deceptive advertising [5], [6], [7], [8].

In this regard, every social network provides customisation of privacy settings to protect their users from such privacy risks [9] [10]. However, privacy preference settings in online services are often complex and time-consuming to adjust; most users feel confused about them and typically ignore or skip them [4]. Though social-network users experience advantage for their online presence, they are often incapable of estimating the privacy risks posed by information-sharing activities. They should have adequate awareness of their privacy and know the risks they may encounter by sharing their information online. Thus, users should be able to protect their sensitive information from their relatives, neighbors and anyone else who have shared their information with and maintain their secrecy [11]. Hence, there is a need to have a model for quantifying and compute privacy risks to create a better view of information revelation for users. By applying a scoring framework and privacy awareness enhancing models, individuals can have a better scheme of their privacy and apply security procedures to increase their level of privacy in case of necessity.

Several attempts [12] [13] have been made to quantify the privacy of a user, although most of them are designed considering one source of information. This may not be a sufficient method as each user may have multiple social network accounts for different purposes. One source of data may not disclose a wide range of information of a user that can pose a privacy risk, but when these information combined from different sources, it can be risky and dangerous. Veiga [14] had shown that there is an increase in privacy leakage due to multiple online social network platforms compared with a single source. For example, a user normally share his/her personal information in Facebook which may pose a privacy risk. This user may share his/her occupation history and background in another site such as LinkedIn. His/her job information has again its own privacy risk, but a combination of the information from two social media accounts can pose the user to higher risk as more information is revealed. Consequently, by considering the overall information from multiple source,

- *Erfan Aghasian, Saurabh Garg and James Montgomery are with the School of Engineering and ICT, University of Tasmania, Hobart, Australia.*
- *Longxiang Gao and Shui Yu are with the School of Information Technology, Deakin University, Melbourne, Australia.*

a more accurate quantification of the privacy disclosure score can be obtained.

The purpose of this paper is to quantify users' privacy in online social networks to inform them about their effective privacy level from their involvement in multiple networks. To quantify the privacy risk of a user, a scoring function is proposed. The inputs to this scoring function consider a set of common personal attributes that may be discovered through social networking sites. The explicit privacy settings for each item, their frequency of occurrence, both within and across social networking sites, are all considered as inputs to the privacy scoring computation in this model. In this work we analyze the factors that have impact on the privacy of the user (sensitivity and visibility of information). For each factor, we provided a comprehensive explanation on how to calculate that factor and finally describe the way to measure the final privacy disclosure score that is related to these two factors. If more than one source of online social network data set is being considered, each attribute of a user has different states of visibility. Hence, due to the complexity of dependency between these inputs, formulating a single formula is not trivial. Thus, we proposed fuzzy-based methods to design the model.

The next section discusses related works. Section 3 specify the design and the mathematical formulation for calculating the privacy disclosure score. Section 4 presents the evaluation of the model using real social network data. The final section presents conclusions and future direction for this work.

## 2 Related Works

The word privacy has numerous subtly definitions. This can be vary from 'personal privacy' to 'Information Privacy', around which privacy on the Internet in all-purpose revolves. Several authors [15] [16] [17] [18] [19] [20] [21] provide various definitions for privacy and information privacy. Meanwhile, the concept of privacy is varied, no particular description of privacy covers all aspects of the term. Accordingly, this study is concerned principally with the information privacy of users. There is a variety of research on privacy concerns in online social networks that deal with data publishing without revealing the identity of the user. Yet there has not been much attention towards privacy from the users' perspective (risks that arise due to information sharing on online networks) [22].

In measuring privacy in online social networks, it is not inherently clear which information can result in a significant loss such as identity theft. Other risks are even harder to measure: comments about and pictures of a user, which are risk-free for some individuals, can be detrimental to others. One likely case is a criticism against a religion or government. In some countries and cultures such criticism is broadly accepted whereas, in other countries, an individual can get in severe difficulties for performing such an action [23] [24]. Another risk of using online social networks is posting vacation information when users are abroad, when intruders to decide when to rob the house based on the information they gather.

Several techniques and methods have been proposed to calculate and compute the privacy and information sharing in a public manner, including algorithmic and statistical approaches. The recommended model by Maximilien et al. [25] assisted users to have a clear picture about their privacy in comparison to other users and overall privacy risk. Renner [23] represented a common approach for defining privacy risk by multiplying the negative consequence of information leakage in the likelihood of disclosure occurrence. Becker [22] stated the significance of quantifying privacy in the online social network. This issue becomes even more critical in case of protecting the huge volume of corresponding personal information, especially in large-scale online social networks. Liu et al. [13] proposed privacy score to model their data sets considering a response matrix with a varying range of items and users. Srivastava et al. [12] dealt with response matrix considering the measurement of text messages in a single source of data. They developed a naïve quotient model for calculating the privacy by assigning binary values for shared and not sharing information about the profile, respectively. Their privacy model measures two factors: the sensitivity of the information and the visibility of the information. Anderson [26] had computed numerous privacy problems in online social networks and described his system – Footlight – to address the problems. Table 1 shows the comparison of key previous methods for measuring and calculating the privacy for online social networks and public data. All these studies consider measuring privacy risks and information leakage for the user from only a single source of data. Beside the mentioned studies, several authors proposed tools for privacy settings configuration in specific online social networks.

The goal of this paper is to present a scoring model that can calculate a privacy disclosure score for multiple data sources, that provides a different approach to this field. So we have formulated our work as a working formula that will calculate the privacy disclosure score of each user while their data have gathered from multiple sources of online social networks (such as Facebook, Twitter, LinkedIn, ResearchGate, and Google+). We do not limit the solution for a specific social network site but the proposed model can be deployed in all social networks. Finally, users can be informed about their privacy level and how much data they have been shared in such networks.

As shown from Table 1, previous methods had considered only a single source of online social networking sites, while in our proposed model we have considered multiple online social networking sites to calculate the privacy disclosure score of the users. Moreover, except the Liu & Terzi method, all other methods and models applied the dichotomous approach (data is publicly available or private) for computing the privacy score, while our model is a polytomous-based one. The other point of uniqueness in our proposed model is its independence to data-type and data structure. In other words, based on information extracted from social networking site, we can run and apply our model to that data.

## 3 Problem Definition

In measuring users' privacy in online social networks, two general factors can be treated as inputs for measuring the privacy disclosure score of users, which are visibility and sensitivity of information. While calculating each factor is a difficult task, this issue becomes more significant where there are multiple sources of data and users revealing their attributes and information in different sites. These attributes and information can be either structured or unstructured

Table 1: Comparison of historical studies

| Authors | Focus | No. of Sources | Data Type | Approach |
|---|---|---|---|---|
| Liu & Terzi [13] | Privacy risk from individual perspective | 1 | Structured | Dichotomous & Polytomous |
| Srivastava et al. [12] | Privacy risk of text messages | 1 | Unstructured | Dichotomous |
| Domingo-Ferrer [27] | Obtained utility by sharing information | 1 | Structured | Dichotomous |
| Nepali et al. [28] | Privacy exposure based on known parameters | - | Web data | Dichotomous |
| Becker & Chen [22] | Attribute inference | 1 | Structured | Polytomous |
| Talukder et al. [29] | Sensitive information leakage of a profile | 1 | Structured | Dichotomous |

data. At the first step of this research, we focus of analysis by these questions:

- What factors have influence on users' privacy in online social networks?
- How to measure the privacy disclosure score of each user in multiple sources?

Since the privacy disclosure score of a user can be measured, users can understand what are their privacy level compared with other users. Thus, users can take more consideration to their privacy to bring it to an acceptable level of privacy. For measuring the privacy disclosure score, we have considered that users' attributes (such as contact number, email, address, job details, hobbies and interests) can be gained from n different sources. For calculating privacy disclosure score, we measure the sensitivity and visibility of information as the inputs for our system. Function ($F_{Sen}$) indicates the sensitivity of each attribute for the users from sources. Beside calculating the sensitivity, we need to provide formulas to calculate the factors that have impact on users' visibility. These factors are known as accessibility to information ($F_{acc}$), difficulty of data extraction of users' information ($F_{dif}$) and the data reliability for each attribute ($F_{rel}$). For a user, we calculate the score of privacy as a combination of the sensitivity score and visibility score of the user combining several attributes such as name, age, gender, email, hometown, job details and interest from different data sources. Here we assume that each user involves in multiple social networks and each attribute is disclosed to the other users in different manners, based on the usage of the social network. For example, the job details on a social network site like LinkedIn is probably more visible than another social network site compared with Facebook.

## 4 Privacy Scoring Framework

Figure 1 presents the overview of privacy disclosure score framework. At the first phase, we consider the attributes for calculating the privacy. These attributes can be extracted from structured data (such as username, family, Age) or obtained from unstructured data (such as blogs, messages and images). It should be noted that this research does not concern with the technologies that can extract these attributes or methods that can be used to collect the data. After obtaining the framework attributes, we compute sensitivity and visibility of users. At first, we measure the sensitivity of the information. We should take into account that some attributes like religious and political views are more sensitive than others. These factors are to be considered in computation of the sensitivity. Next, we calculate the visibility based on three factors that have a direct impact on visibility (accessibility to information, the difficulty of data extraction and reliability of data). The overall privacy disclosure score is finally obtained from the combination of sensitivity and visibility scores. Finally, we analyze the result and allow the users know how strength are they privacy level in comparison with other users.

### 4.1 Calculation of Sensitivity

Sensitivity shows the risk associated with the attributes of the user. when the sensitivity of an attribute increases, the risk posed by information disclosure of the individuals also increases. Srivastava et al [12] calculated the sensitivity score for 11 attributes for measuring the privacy score based on the quotient model. His results indicated that the most sensitive attributes are related to political view, religious view, contact number and relationship status. In contrast, birthdate and current town details are not that sensitive for the users. For sensitivity score ($F_{Sen}$), we are using sensitive values derived by Srivastava eat al. [12], which is shown in table2 for each profile item in their sample.

Table 2: Sensitivity score for users' attributes

| Attributes | Sensitivity |
|---|---|
| Contact Number | 0.6 |
| E-Mail | 0.1833 |
| Address | 0.85 |
| Birthdate | 0.1166 |
| Hometown | 0.15 |
| Current town | 0.1166 |
| Job Details | 0.2 |
| Relationship Status | 0.4166 |
| Interests | 0.3 |
| Religious Views | 0.5666 |
| Political Views | 0.6833 |

### 4.2 Calculation of Visibility

Visibility determines how widely accessible the attributes of a user are in an online social network. For calculating the visibility, we considered three factors that influence of the visibility

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2720187, IEEE Access

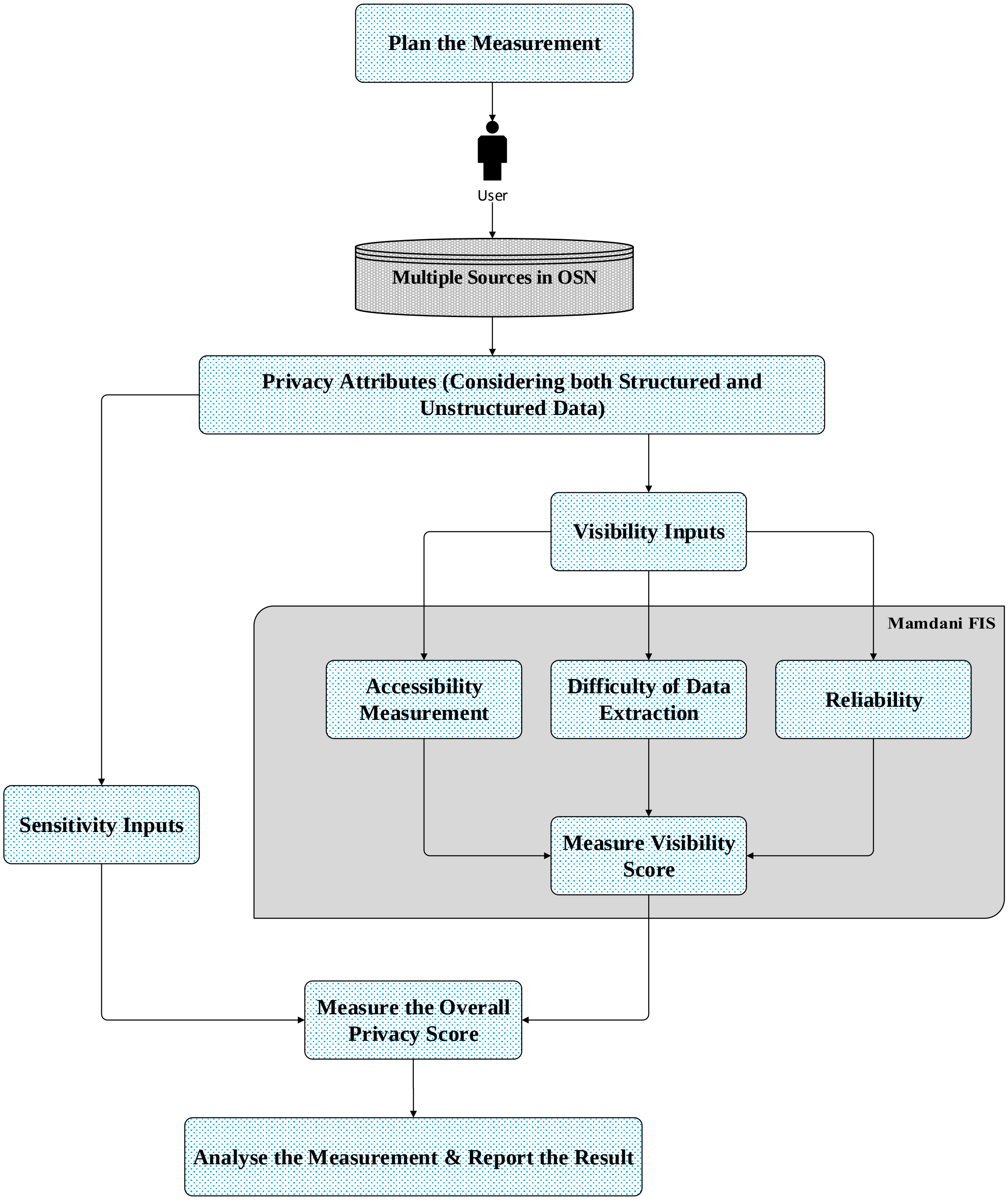

Plan the Measurement
User
Multiple Sources in OSN
Privacy Attributes (Considering both Structured and Unstructured Data)
Visibility Inputs
Mamdani FIS
Accessibility Measurement
Difficulty of Data Extraction
Reliability
Sensitivity Inputs
Measure Visibility Score
Measure the Overall Privacy Score
Analyse the Measurement & Report the Result

Figure 1: Overview of Privacy Score Framework

of user information in online social networks. These factors are ease of accessibility, the difficulty of data extracting and frequency of occurrence of information disclosure. The current predefined permissions for attributes satisfy the visibility of each item for each user. While some information of a number of users is publicly available, other attributes can be private or semi-private.

#### 4.2.1 Accessibility Calculation

We define accessibility as a measure of permissions that are given for their sharing information with others. In other words, accessibility indicates how many people can have access to a specific piece of information and to what level. There are four different level for users' information accessibility. The information can be (a) accessible only by the owner of the information,(b) can be accessible by his/her friends, (c) accessible by his/her friends of friends and finally (d) can be publicly available.

An Accessibility Value ($AV$) between 1 and 4 is given to each attribute (1 → not accessible except data owner, 2→ accessible by friends, 3→ accessible by friends of friends, 4→ publicly available). For calculating the accessibility to each profile item ($F_{acc}$), we assume that each user is participating in $n$ different online social network site while they having different sensitive attributes. The sources indicate in which

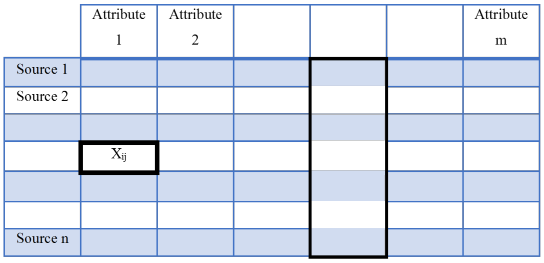


Figure 2: Accessibility score matrix for privacy measurement for each user

online social network a user participates. Based on the nature of each online social network, accessibility value may vary. As a case in point, the interest of a user can be accessible more easily than his social network like academia. Let $i$ be source, $n$ be the number of sources, $j$ be an attribute, and $m$ be the total number of attributes. Figure 2 shows the accessibility score matrix for calculations. After assigning the accessibility values to each of the attributes for each social network, we provide an algorithm ($F_{acc}$) to calculate the accessibility given by:

**Algorithm 1:** Algorithm for Computing Accessibility for an attribute

```
Data: Input: User response matrix with m columns and
      n rows
Result: Accessibility score of each attribute;
// Initialize the temp matrix;
for $k = 1 : m$ do
    // Extract the $k^{th}$ column and put it in col
       variable
    Based on the input, delete the entries that does not
      meet the condition i.e. if the difference between the
      maximum number and minimum number in each
      column is equal to three, delete ones;
    // calculate the mean after checking the
       defined condition
end
for $j = 1 : m$ do
    Initiate counter and sum variables with value equal
      to zero;
    for $i = 1 : n$ do
        if $temp(i, j)! = 0$ then
            sum = sum + input(i,j);
            counter = counter + 1;;
        end
        Display the calculated accessibility values;
    end
    means(1,j) = sum/counter;
end
```

It should be noted that the reason for deleting '$x_{ij}$' when the range is equal to 3 is that we have an attribute publicly available in one source while its accessibility is completely private in another source(s). While user data in a source is publicly available and anyone can have access to it, users only accessibility does not make sense in other sources. Therefore, it can be concluding that we should just calculate the data that have an impact on user privacy. In another scenario, the data accessibility might not be publicly accessible or can only be accessed by the user and the privacy measurement can be calculated by the defined permission to the information by users. In this case, we simply calculate the mean of accessibility value of each attribute for each user.

#### 4.2.2 Data Extraction Difficulty

One factor that is important to compute the privacy is the difficulty of extracting private information from different formats of data. Extracting attributes from structured data is much easier than unstructured data. For example, it is hard to understand one's religious view from his/her picture than the place he/she clearly stated his religious view. For calculation of difficulty three level has been defined (3 → low difficulty, 2 → medium difficulty, 1→ high difficulty). Naturally, the more a profile item is accessible; the difficulty of data gathering is less. To compute how much it is difficult to extract an attribute, we calculate the mean of extraction difficulty of each attribute for each of the social networks.

$$F_{dif} = \sum_{n} (dif_j)$$

where $dif_j$ indicates the difficulty of extracting an attribute from a social network data.

#### 4.2.3 Data Reliability Calculation

Reliability is a criterion that can provide with what confidence a particular attribute has been disclosed in one or multiple sources. In this context, for each attribute of a user, we consider the overall reliability of data disclosure for each attribute of the users to consider it in total visibility calculation. As reliability of a sensitive information will increase with more number of resources validating it, we are using a sigmoid function to measure the reliability of data. The reason for using the sigmoid function was that this function supports us in differentiating the reliability. The equation for calculating the reliability given by:

$$F_{rel} = \frac{2}{1 + e^{-s}} - 1$$

Where 's' indicates the number of sources the attribute has been revealed. The output boundary for this function is [0,1], where the number of sources of disclosure increases, the reliability increase.

#### 4.2.4 Total Visibility Calculation

Our proposed method for calculating the overall visibility score for the users is based on a set of fuzzy rules that is occurring for users in different situations. The reason for choosing the fuzzy inference system (FIS) [30] [31] was based on the nature of system and process complexity, which involves various interacting parameters. Hence, FIS considered as a suitable method for applying in this type of decision systems. After defining the rules based on the inputs (Calculated numbers for the accessibility, difficulty of data extraction and frequency of occurrence), the Mamdani fuzzy inference [32] is used. Assume that a user wants to know what is his/her visibility level if he/she discloses his/her information in multiple datasets. The designed fuzzy system can explain him/her that in which level of privacy (in context of visibility). The process of FIS (Figure 3) based on Madani's method [32] would be as:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2720187, IEEE Access

1) **Fuzzification (of inputs):** antecedent evaluation for each rule – obtain membership values from crisp values
2) **Implication:** obtain the consequences of each rule
3) **Aggregation:** combining step 2 output for each rule into a single fuzzy set by using a fuzzy aggregation operator
4) **Defuzzification:** obtain a crisp number as the output

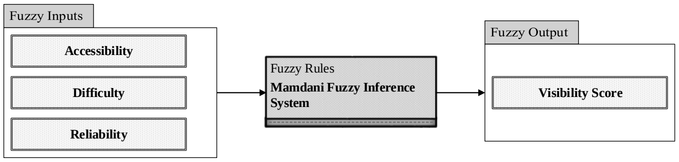


Figure 3: Fuzzy Inference System Overview

In the fuzzification step, a generalized bell function was selected as the membership function to define the fuzzy sets. The generalized bell function is given by:

$$f(x,[a,b,c]) = \frac{1}{1+|\frac{(x-c)}{/}a|^{2b})}$$

As can be seen, this function depends on the three parameters $a$, $b$ and $c$. Each of these parameters has a physical meaning. Parameter $c$ determines the center of the corresponding membership function. Parameter $a$ is the half width; and $b$ controls the slope at the crossover points. Figure 4 shows an example of a plotted generalized bell function. Table 3 presents details of each membership function used in fuzzy inference model.

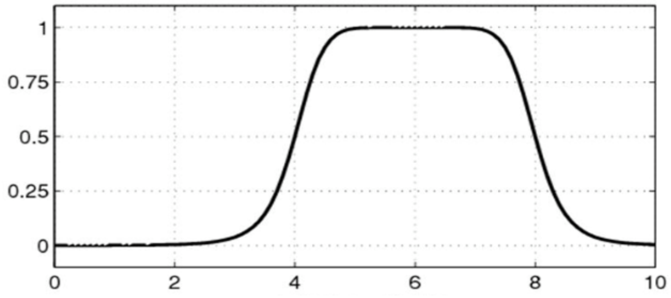


Figure 4: An Example of Generalized bell function

Apart from the membership function details, a set of rules were defined to make a logical calculation for the visibility of a user' attributes based on the inputs (Table 4). According to fuzzy inference system model and fuzzy logic, the logical 'AND' operator were treated as 'min' while the $OR$ operator treated as a 'max' operation on the corresponding membership function. Thereafter, we applied $max$ operation for the aggregation of the database rules on the resulting of the resultant (corresponding) rules.

The fuzzy rules in the model have been obtained after several consultations with experts in the domain knowledge. The membership functions are defined such that they most precisely match the values of a particular attribute. For example, if the accessibility of data is high (i.e the data is publicly available), and the frequency of occurrence is high (data published in more than 3 sources) as well, then the visibility (regardless of the value of data extraction difficulty) is high. The rest of the rules have been defined by this method (applying the experts knowledge).

The last step in a fuzzy inference system is defuzzification. A Defuzzification method permits to obtain a crisp number from a fuzzy value. The two most practical methods are: mean of maxima and centroid (center of mass) [33]. In this paper, we exploit the centroid function (which provide us the better result compared with other fuzzy functions), which indicates the center of the area under the curve to obtain a crisp value for the output (visibility). This method computes the output (a crisp number) from defined rules (as input) as:

$$F_{vis}(x) = \frac{\int_{x1}^{x2} xf(x)dx}{\int_{x1}^{x2} f(x)dx}$$

Where the centroid function of the area bounded by $B = [x1, \ x2]$ and the x-axis, and the function $F_{vis}(x)$ converts points of B to a crisp value. The obtained value can be considered as the visibility score for the users' attribute. The FIS model for a sample is illustrated in Figure 5.

### 4.3 Calculation of Privacy Score

By considering the $\beta_i$ as the sensitivity of each attribute and $F_{vis}(x)$ as the visibility of each attribute, the overall privacy disclosure score of each user can be calculated by privacy disclosure score function given by:

$$Privacy = \frac{\sum_{i=1}^{m} \beta_i * F_{vis}(xi)}{m}$$

where $i$ indicates the i-th attribute of a user and $m$ is the number of attributes. As the calculated value increases, it indicates that a user is more likely in a risk of privacy and information disclosure, where the less is better.

## 5 Experimental Evaluation

In this section we present the evaluation of our proposed privacy model. First we present evaluation for validation of the model. Then, we evaluate the accuracy of the model using real case study by comparing with the privacy scoring model by Liu & Terzi [13] which is the most recent polytomous approach for structured data.

### 5.1 Model Validation

To validate the privacy model, we did a simple test where we have choosen three different cases: a user who has all the data public, a user who has all data private and a user who has some data public and other private. In a valid model, the obtainied privacy scores should match the expected theoritical scores for the three choosen cases, i.e. highest score, lowest score and in between score, respectively. Table 5 illustrates the visibility scores obtained from our proposed model for three different cases. Clearly, the scores are fairly close to the theoretical expectation. For example, for user a, the visibility score is close to the highest expected score '8'. As in our proposed privacy function, for a given user sensitivity values are constant, therefore the privacy function validity is verified by the validity of our proposed visibility function.

Table 3: MF Database; Linguistic Variable (LV), Membership Function (MF)

| LV | Type | MF | Range | a | b | c |
|---|---|---|---|---|---|---|
| Accessibility | Input | Very Low | [0,4] | 1.171 | 11.8 | 0 |
| Accessibility | Input | Low | [0,4] | 0.449 | 7.73 | 1.85 |
| Accessibility | Input | Medium | [0,4] | 0.367 | 5.439 | 2.67 |
| Accessibility | Input | High | [0,4] | 0.667 | 14.28 | 4 |
| Difficulty | Input | Low | [0,3] | 1.16 | 14.48 | 0 |
| Difficulty | Input | Medium | [0,3] | 0.492 | 6.67 | 1.81 |
| Difficulty | Input | High | [0,3] | 0.5 | 8.56 | 3 |
| Reliability | Input | Very Low | [0,1] | 0.172 | 18.1 | 0.25 |
| Reliability | Input | Low | [0,1] | 0.136 | 19.2 | 0.567 |
| Reliability | Input | Medium | [0,1] | 0.086 | 13.2 | 0.798 |
| Reliability | Input | High | [0,1] | 0.02 | 12.3 | 0.919 |
| Reliability | Input | Very High | [0,1] | 0.132 | 32 | 1.08 |
| Visibility | Output | Very Low | [0,8] | 1 | 11.2 | 0.799 |
| Visibility | Output | Low | [0,8] | 0.628 | 7.59 | 2.64 |
| Visibility | Output | Medium | [0,8] | 0.792 | 23.6 | 11.4 |
| Visibility | Output | High | [0,8] | 0.688 | 12.3 | 5.69 |
| Visibility | Output | Very High | [0,8] | 1 | 11.1 | 7.608 |

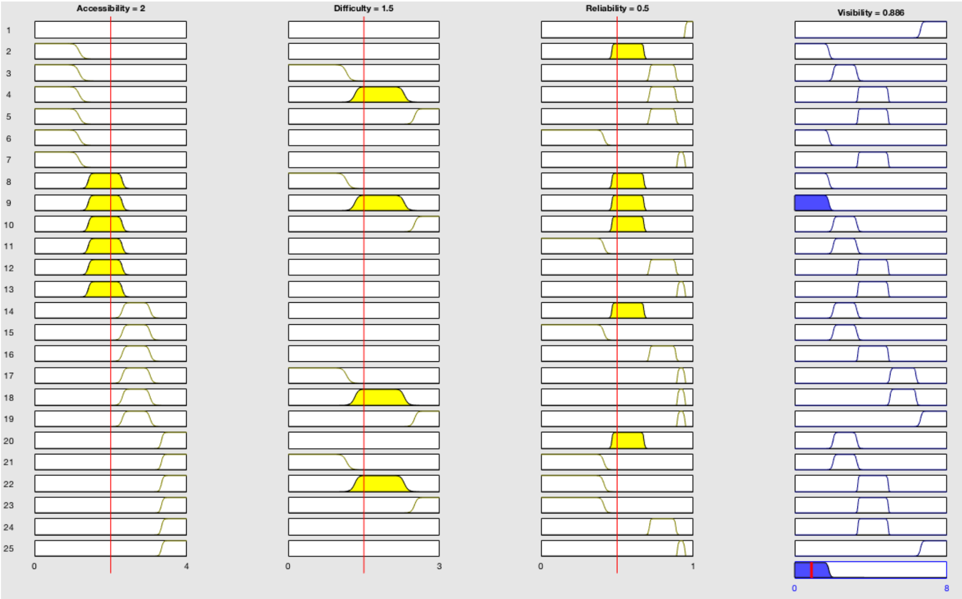


Figure 5: FIS model for visibility score calculation

## 5.2 Case Study

We measured to what extent the users' personal information is revealed in multiple sources of online social networks. In other words, our intent is to calculate the privacy risk based on how much information a user has disclosed overall in all the social networks. We gathered the data of 15 users who were involved in four different online social networks (Facebook, ResearchGate, LinkedIn, and Google+) containing 11 attributes (as in Table 1) for each user to measure the information disclosure and privacy risk of those users. The chosen number of users cover diverse range of values from user profiles that is needed to show effectiveness of the proposed privacy score method. Then, to calculate the privacy disclosure score of users we considered two factors (sensitivity and visibility) that have direct influence on users' privacy. We obtained the sensitivity value from the literature review and historical work. In order to compute the visibility score, we deployed a Mamdani fuzzy inference system to obtain the visibility score after calculating the related functions (Accessibility, Difficulty of data extraction and Reliability). Then, we compute the overall privacy disclosure score for each user. At the final step, we compared the privacy disclosure score of all the users with the the privacy scoring model by Liu & Terzi [13] to evaluate the accuracy of our proposed model.

Table 4: Fuzzy rules database (VL= Very low, L= Low,M=Medium,H=High, VH=Very High, X=Can be in any state )

| | Inputs | | | Output |
|---|---|---|---|---|
| | Accessibility | Difficulty of data extraction | Frequency of occurrence | Visibility |
| 1 | X | X | VH | VH |
| 2 | L | L | VL | VL |
| 3 | L | M | VL | VL |
| 4 | L | H | VL | L |
| 5 | VL | X | VL | VL |
| 6 | VL | L | M | L |
| 7 | VL | M | M | M |
| 8 | VL | H | M | M |
| 9 | VL | X | L | VL |
| 10 | VL | X | H | M |
| 11 | L | X | L | L |
| 12 | L | X | M | M |
| 13 | L | X | H | M |
| 14 | M | X | VL | L |
| 15 | M | X | L | M |
| 16 | M | L | M | M |
| 17 | M | L | H | H |
| 18 | M | M | H | H |
| 19 | M | H | H | VH |
| 20 | H | X | VL | L |
| 21 | H | L | L | L |
| 22 | H | M | L | L |
| 23 | H | H | L | M |
| 24 | H | X | M | M |
| 25 | H | X | H | H |

Table 5: Users visibility comparison

| User | Inputs | | | Output | Expected Theoritical Score |
|---|---|---|---|---|---|
| | Facc | Fdif | Frel | Fvis | Fvis |
| User a (All public data) | 4 | 0 | 1 | 7.32 | 8 |
| User b (All private data) | 0 | 3 | 0 | 0.884 | 0 |
| User c (Partically public data) | 2 | 1.5 | 0.5 | 4.12 | (3-5) |

### 5.2.1 *Analysis of Results*

For our experiments, we calculate the value of $F_{acc}$ and $F_{dif}$ functions by considering the accessibility and difficulty values of each attribute for each user as the input. These values may be varying in each social network.

By considering the user accessibility (Table 6) and difficulty of data extraction (Table 7), we provide the final calculated values for the accessibility and difficulty score for two users (user o and user b - FB=Facebook, RG=ResearcGate, LD=LinkedIn, G+=Google+) as a sample.

Table 8 illustrates the gained values from the fuzzy inference system. It should be noted that if the accessibility of all sources has the value equal 1 (only accessible to the user), the reliability of the data would be zero and we exempt that parameter from the final calculation for visibility (in other words, the visibility for that parameter is zero).

For the moment, the fuzzy inference system was used to calculate the visibility of each attribute for the users. Table 9 shows the obtained value for the visibility.

By comparing the results of the table, it can be seen that users do not tend to provide the information, which is more sensitive than the other attributes. Based on our experiments, we found out that users may likely disclose their information such as their Email address, current town, and

Table 6: Two Users' Accessibility

| User (Accessibility) | User o (FB, RG, LD, G+) | User b (FB, RG, LD, G+) |
|---|---|---|
| Contact number | (2,2,4,3)→ Facc = 2.75 | (1,1,1,2)→ Facc = 1.25 |
| Email | (2,2,3,4)→ Facc = 2.75 | (1,1,1,2)→ Facc = 1.25 |
| Address | (2,2,2,4)→ Facc = 2.5 | (2,1,2,1)→ Facc = 1.5 |
| Birthdate | (3,2,3,4)→ Facc = 3 | (1,1,1,1)→ Facc = 1 |
| Home town | (3,2,3,4)→ Facc = 3 | (2,1,1,1)→ Facc = 1.25 |
| Current town | (3,3,2,4)→ Facc = 3 | (3,1,2,1)→ Facc = 1.75 |
| Job details | (2,4,4,4)→ Facc = 3.5 | (2,1,4,1)→ Facc = 3 |
| Relationship Status | (3,2,2,4)→ Facc = 2.75 | (2,1,1,1)→ Facc = 1.25 |
| Interests | (3,3,2,3)→ Facc = 2.75 | (2,1,3,1)→ Facc = 1.75 |
| Religious views | (3,2,2,4)→ Facc =2.75 | (1,1,1,1)→ Facc = 1 |
| Political views | (2,2,1,1)→ Facc = 1.5 | (1,1,1,1)→ Facc = 1 |

Table 7: Two Users' Data Extraction Difficulty

| User (Difficulty) | User o (FB, RG, LD, G+) | User b (FB, RG, LD, G+) |
|---|---|---|
| Contact number | (2,2,3,3)→ Fdif = 2.5 | (1,1,1,2)→ Fdif = 1.25 |
| Email | (2,2,3,4)→ Fdif = 2.5 | (1,1,1,2)→ Fdif = 1.25 |
| Address | (2,2,2,3)→ Fdif = 2.25 | (2,1,2,1)→ Fdif = 1. 5 |
| Birthdate | (3,2,3,3)→ Fdif = 2.75 | (1,1,1,1)→ Fdif = 1 |
| Home town | (3,2,3,3)→ Fdif = 2.75 | (2,1,1,1)→ Fdif = 1.25 |
| Current town | (3,2,3,3)→ Fdif = 2.75 | (3,1,2,1)→ Fdif = 1.75 |
| Job details | (2,3,3,3)→ Fdif = 2.75 | (2,1,3,1)→ Fdif = 1.75 |
| Relationship Status | (3,2,2,3)→ Fdif = 2.5 | (2,1,1,1)→ Fdif = 1.25 |
| Interests | (3,3,2,3)→ Fdif = 2.75 | (2,1,3,1)→ Fdif = 1.75 |
| Religious views | (3,2,2,3)→ Fdif = 2.5 | (1,1,1,1)→ Fdif = 1 |
| Political views | (2,2,1,1)→ Fdif = 1.5 | (1,1,1,1)→ Fdif = 1 |

Table 8: FIS-based visibility calculation

| User (Difficulty) | User o | | | | User b | | | |
|---|---|---|---|---|---|---|---|---|
| | Facc | Fdif | Frel | Fvis | Facc | Fdif | Frel | Fvis |
| Contact number | 2.75 | 2.5 | 0.96 | 7.32 | 1.25 | 1.25 | 0.46 | 1.5 |
| Email | 2.75 | 2.5 | 0.96 | 7.32 | 1.25 | 1.25 | 0.48 | 1.52 |
| Address | 2.5 | 2.75 | 0.97 | 7.3 | 1.5 | 1.5 | 0.76 | 4.11 |
| Birthdate | 3.2 | 2.75 | 0.96 | 7.32 | 1 | 1 | 0 | 0 |
| Home town | 3.2 | 2.75 | 0.96 | 7.32 | 1.25 | 1.25 | 0.46 | 1.5 |
| Current town | 3.2 | 2.75 | 0.96 | 7.32 | 1.75 | 1.75 | 0.76 | 4.11 |
| Job details | 3.5 | 2.75 | 0.96 | 7.32 | 3 | 1.75 | 0.76 | 7.32 |
| Relationship Status | 2.75 | 2.5 | 0.96 | 7.32 | 1.25 | 1.25 | 0.46 | 1.5 |
| Interests | 2.75 | 2.5 | 0.96 | 7.32 | 1.75 | 1.75 | 0.76 | 4.11 |
| Religious views | 2.75 | 2.5 | 0.96 | 7.32 | 1 | 1 | 0 | 0 |
| Political views | 1.5 | 1.5 | 0.76 | 4.11 | 1 | 1 | 0 | 0 |

Table 9: FIS model for visibility score calculation

| Users | Contact Number | Email | Address | Date of Birth | Home Town | Current Town | Job Details | Relationship Status | Interests | Religious Views | Political Views |
|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 3.59 | 3.6 | 3.6 | 3.5 | 1.5 | 2.65 | 3.59 | 0 | 0 | 0 | 0 |
| b | 1.51 | 1.52 | 4.11 | 0 | 1.5 | 4.1 | 7.32 | 1.5 | 4.1 | 0 | 0 |
| c | 4.1 | 3.6 | 1.01 | 1.64 | 1.01 | 5.66 | 5.26 | 0 | 4.12 | 3.54 | 0 |
| d | 1.52 | 1.5 | 1.5 | 1.5 | 0 | 4.1 | 4.12 | 0 | 4.12 | 2.64 | 4.14 |
| e | 1.5 | 7.33 | 4.42 | 4.1 | 4.12 | 7.32 | 7.32 | 0 | 4.12 | 0 | 0 |
| f | 1.5 | 1.51 | 4.11 | 4.1 | 0 | 7.32 | 7.32 | 4.1 | 4.12 | 4.12 | 0 |
| g | 4.12 | 4.12 | 5.66 | 4.11 | 1.5 | 5.66 | 5.66 | 1.5 | 7.32 | 4.11 | 0 |
| h | 2.65 | 2.65 | 7.32 | 4.12 | 0.88 | 7.16 | 2.65 | 2.59 | 7.31 | 2.64 | 1.01 |
| i | 4.1 | 7.3 | 7.32 | 4.11 | 4.12 | 4.12 | 7.31 | 1.51 | 7.3 | 1.01 | 1.51 |
| j | 7.32 | 4.12 | 4.41 | 4.12 | 4.11 | 7.31 | 7.32 | 1.5 | 4.1 | 4.11 | 1.5 |
| k | 4.1 | 4.12 | 4.1 | 4.1 | 4.11 | 7.3 | 7.32 | 4.11 | 7.32 | 4.12 | 4.12 |
| l | 3.59 | 2.64 | 7.32 | 0.88 | 4.12 | 7.32 | 7.32 | 1.2 | 7.3 | 7.3 | 1.01 |
| m | 4.11 | 4.11 | 4.12 | 3.54 | 7.32 | 7.3 | 7.3 | 0 | 7.32 | 7.3 | 4.11 |
| n | 4.1 | 1.52 | 3.54 | 4.11 | 4.12 | 7.32 | 7.3 | 4.11 | 7.18 | 4.12 | 7.3 |
| o | 7.32 | 7.3 | 7.32 | 7.32 | 7.3 | 7.31 | 7.32 | 7.32 | 7.32 | 7.31 | 4.11 |

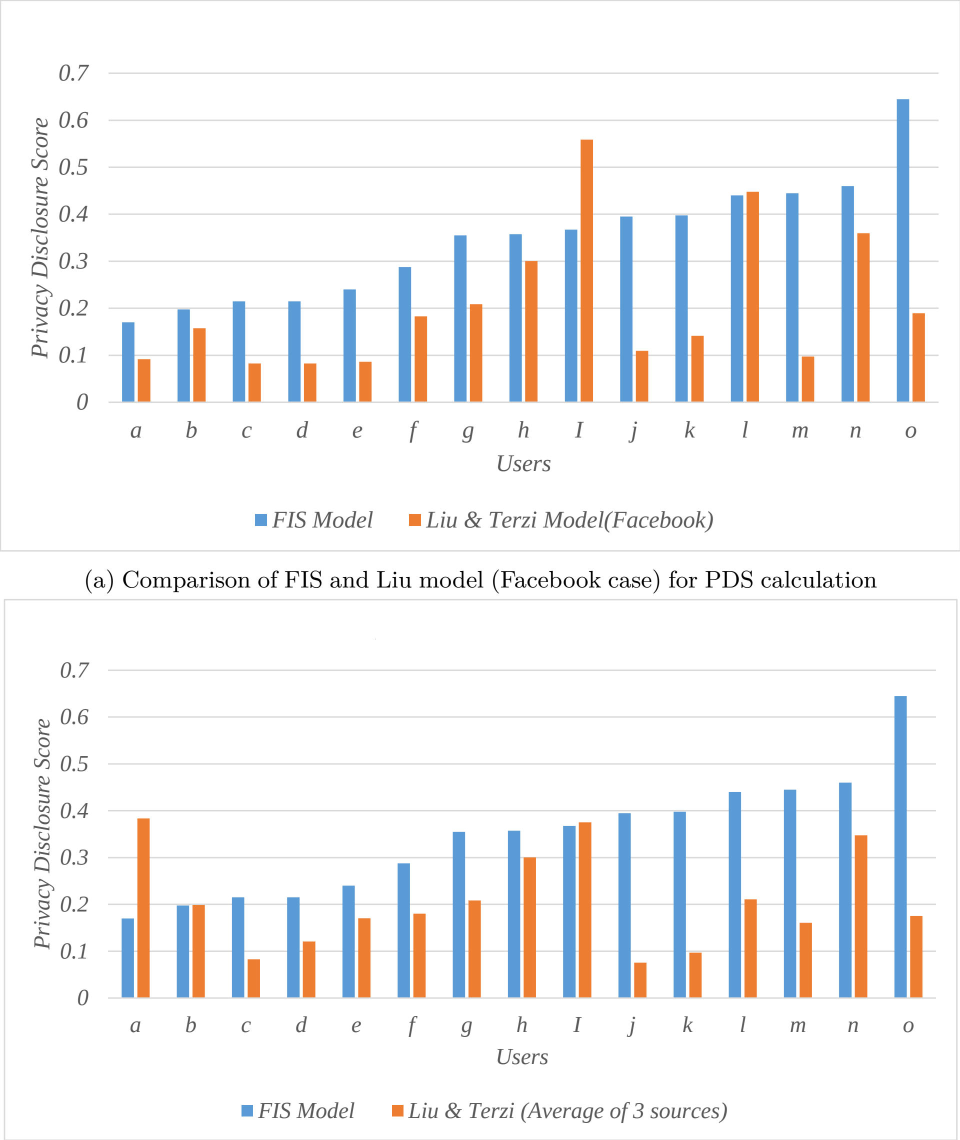
Table 10: User's final privacy disclosure score calculation

| Users | Contact Number | Email | Address | Date of Birth | Home Town | Current Town | Job Details | Relationship Status | Interests | Religious Views | Political Views | Privacy Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 2.154 | 0.659 | 1.060 | 0.408 | 0.225 | 0.308 | 0.718 | 0 | 0 | 0 | 0 | 0.684 |
| b | 0.904 | 0.278 | 3.493 | 0 | 0.225 | 0.478 | 1.464 | 0.624 | 1.230 | 0 | 0 | 0.790 |
| c | 2.460 | 0.659 | 0.858 | 0.191 | 0.151 | 0.599 | 1.252 | 0 | 1.215 | 2.009 | 0 | 0.861 |
| d | 0.912 | 0.273 | 1.275 | 0.174 | 0 | 0.478 | 0.824 | 0 | 1.218 | 1.493 | 2.828 | 0.863 |
| e | 0.900 | 1.341 | 3.757 | 0.478 | 0.618 | 0.851 | 1.464 | 0 | 1.236 | 0 | 0 | 0.967 |
| f | 0.900 | 0.276 | 3.493 | 0.478 | 0 | 0.853 | 1.464 | 1.708 | 1.215 | 2.334 | 0 | 1.158 |
| g | 2.472 | 0.754 | 4.811 | 0.479 | 0.225 | 0.659 | 1.132 | 0.624 | 2.195 | 2.328 | 0 | 1.429 |
| h | 1.590 | 0.485 | 5.222 | 0.480 | 0.132 | 0.834 | 0.530 | 1.078 | 2.193 | 1.495 | 0.690 | 1.430 |
| i | 2.460 | 1.336 | 5.222 | 0.479 | 0.618 | 0.480 | 1.462 | 0.629 | 2.190 | 0.372 | 1.031 | 1.589 |
| j | 4.392 | 0.754 | 3.748 | 0.480 | 0.616 | 0.852 | 1.464 | 0.624 | 1.230 | 2.328 | 1.031 | 1.593 |
| k | 2.460 | 0.754 | 3.485 | 0.478 | 0.615 | 0.851 | 1.464 | 1.712 | 2.196 | 2.334 | 2.813 | 1.742 |
| l | 2.154 | 0.483 | 5.222 | 0.102 | 0.618 | 0.851 | 1.464 | 0.499 | 2.190 | 4.136 | 0.690 | 1.764 |
| m | 2.466 | 0.752 | 3.502 | 0.412 | 1.095 | 0.851 | 1.460 | 0 | 2.196 | 4.136 | 2.808 | 1.798 |
| n | 2.460 | 0.278 | 3.009 | 0.479 | 0.618 | 0.863 | 1.460 | 1.712 | 2.154 | 2.334 | 4.988 | 1.849 |
| o | 4.392 | 1.336 | 5.222 | 0.853 | 1.095 | 0.852 | 1.464 | 3.049 | 2.195 | 4.141 | 2.808 | 2.582 |

interests while information related to their political and religious view has less likelihood of disclosure. After computing the visibility score of each attribute for the users, the next step is to calculate overall privacy disclosure score for the users. Regarding our case study, which involves 15 users, we have deployed privacy disclosure score calculation derived in previous section.

Table 10 shows the computed result for the users and illustrates the final privacy disclosure score of the users in our case study. Regarding the obtained value for the privacy disclosure score of each user, it can be observed that the users who have greater willingness to disclose their information (can be both sensitive and non-sensitive), have the higher risk for their privacy.

Figures 6 illustrates the overall privacy disclosure score of each user measured by two different methods after data normalization (between 0 and 1). In figure 6a, comparison of results indicates that majority of calculations applying FIS model exhibit higher disclosure scores, excluding two exceptions of 'i' and 'l'. This generally higher level of privacy disclosure score from FIS model is because of using higher number of input data from multiple sources which result in revealing more information and also the accuracy and reliability of each attribute itself. For a better clarification, three users with three different patterns has been compared. For the first case, the user 'i' exceptionally exhibits higher privacy disclosure score using Liu & Terzi model (0.56 for Liu & Terzi [13] model vs 0.37 for FIS model) as a reason of



(a) Comparison of FIS and Liu model (Facebook case) for PDS calculation

(b) Comparison of FIS and Liu model (Average of 3 sources) for PDS calculation

Figure 6: Comparison of FIS and Liu model for PDS calculation

sharing high level of sensitive information within one source of input data. This user has shared most of information on Facebook and did not provide sufficient amount of sensitive information on other sources resulting in a lower level of reliability of data (which is calculated in fuzzy phase) and consequently a decrease in the privacy disclosure score calculated by FIS model. Hence, the obtained value by Liu method is higher which is reasonably practical for a single source of data. For the second case, user ‘l’ similarly indicates comparable disclosure scores for both models ( 0.45) due to the high proportion of released data in Facebook as one of the sources of information. Comparing of results obtained from users ’l’ and ’o’ indicates that both users have revealed the same amount of information on their Facebook profile (six out of 11 attributes). For the last case, the user ’l’ has disclosed more sensitive information than the user ’o’, resulting in higher exposure score for the user ’l’ by Liu & Terzi model. While majority of information shared by the user ‘o’ on Facebook are not sensitive, other sources provide publicly available sensitive information. Therefore, unlike obtained low level of disclosure score using the Liu method, the privacy disclosure score for the user ‘o’ computed by fuzzy method is very high. Figure 6b illustrates the results from the FIS model with the average score of three social network sources (i.e. Facebook, LinkedIn and ResearchGate) calculated using Liu & Terzi model. The graph clearly show that the obtained average score is not sufficient to capture the risk of disclosure across multiple sites, excluding the user ‘a’. User ‘a’ significantly exhibits lower score of privacy disclosure by FIS model because of the scattered distribution of the attributes within each individual source. In this case, the reliability of data in the fuzzy function may not provide a high level value as the Liu & Terzi model do.

## 6 Conclusion and Future Work

As social network usage is increasing day by day, privacy concerns are becoming more and more important. Social network users generally have multiple social network accounts for different purposes and in each network they will be sharing their personal information. One of the challenges in addressing privacy concerns is how to measure the privacy of a user participating in multiple social networks. In this context, we considered three aspects to compute the overall privacy disclosure score of a user who is participating in multiple social networks. We have proposed a system to compute information visibility as a factor that has a direct impact on a user's privacy disclosure score. For doing so, we selected the Mamdani fuzzy inference system to compute the visibility score for the attributes of the users. Finally, we compute the overall privacy disclosure score of users who are sharing their information in multiple social networks. Regarding the obtained privacy disclosure scores, we can conclude that users' privacy disclosure scores directly depend on the amount of information a user discloses, such as religious views, political views, and relationship status. Finally, the results obtained in this study allow us to conclude that the proposed framework to measure the privacy of the users can offer a positive perception for the users to have a more detailed examination of the information they want to share in the future.

In future, we would like to explore further generalisation of the privacy scoring framework considering users' perspectives about the sensitivity of their data. Further, there is no explicit criterion for measuring the difficulty of data extraction, which is hence an open problem for further investigation.

## References

[1] L. Gao, M. Li, W. Zhou, and W. Shi, "Privacy protected data forwarding in human associated delay tolerant networks," in *2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE, 2013, pp. 586–593.

[2] Q. Ma, H. H. Song, S. Muthukrishnan, and A. Nucci, "Joining user profiles across online social networks: From the perspective of an adversary," in *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on.* IEEE, 2016, pp. 178–185.

[3] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in *Proceedings of the 2005 ACM workshop on Privacy in the electronic society.* ACM, 2005, pp. 71–80.

[4] E. Zheleva, E. Terzi, and L. Getoor, "Privacy in social networks," *Synthesis Lectures on Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 1–85, 2012.

[5] G. Brown, T. Howe, M. Ihbe, A. Prakash, and K. Borders, "Social networks and context-aware spam," in *Proceedings of the 2008 ACM conference on Computer supported cooperative work.* ACM, 2008, pp. 403–412.

[6] J. Hechinger, "College applicants, beware: Your facebook page is showing," *The Wall Street Journal*, 2008.

[7] H. Jia and H. Xu, "Measuring individuals' concerns over collective privacy on social networking sites," *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 10, no. 1, 2016.

[8] X. Han and L. Wang, "Are you really hidden? estimating current city exposure risk in online social networks," 2016.

[9] S. Livingstone, "Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression," *New media & society*, vol. 10, no. 3, pp. 393–411, 2008.

[10] A. Quattrone, L. Kulik, E. Tanin, K. Ramamohanarao, and T. Gu, "Privacypalisade: Evaluating app permissions and building privacy into smartphones," in *2015 10th International Conference on Information, Communications and Signal Processing (ICICS).* IEEE, 2015, pp. 1–5.

[11] B. Wittes and E. Kohse, "The privacy paradox ii: Measuring the privacy benefits of privacy threats," *Center for Technology Innovation at Brookings, Washington DC*, 2017.

[12] A. Srivastava and G. Geethakumari, "Measuring privacy leaks in online social networks," in *Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on.* IEEE, 2013, pp. 2095–2100.

[13] K. Liu and E. Terzi, "A framework for computing the privacy scores of users in online social networks," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 5, no. 1, p. 6, 2010.

[14] M. H. Veiga and C. Eickhoff, "Privacy leakage through innocent content sharing in online social networks," *arXiv preprint arXiv:1607.02714*, 2016.

[15] M. Beye, A. J. Jeckmans, Z. Erkin, P. Hartel, R. L. Lagendijk, and Q. Tang, "Privacy in online social networks," in *Computational Social Networks.* Springer, 2012, pp. 87–113.

[16] C. Bunnig and C. H. Cap, "Ad hoc privacy management in ubiquitous computing environments," in *Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services, 2009. CENTRIC'09. Second International Conference on.* IEEE, 2009, pp. 85–90.

[17] Q. Ni, E. Bertino, J. Lobo, C. Brodie, C.-M. Karat, J. Karat, and A. Trombeta, "Privacy-aware role-based access control," *ACM Transactions on Information and System Security (TISSEC)*, vol. 13, no. 3, p. 24, 2010.

[18] S. Taheri, S. Hartung, and D. Hogrefe, "Achieving receiver location privacy in mobile ad hoc networks," in *Social Computing (SocialCom), 2010 IEEE Second International Conference on.* IEEE, 2010, pp. 800–807.

[19] J. Kang, "Information privacy in cyberspace transactions," *Stanford Law Review*, pp. 1193–1294, 1998.

[20] D. J. Solove, "A taxonomy of privacy," *University of Pennsylvania law review*, pp. 477–564, 2006.

[21] E. F. Stone, H. G. Gueutal, D. G. Gardner, and S. McClure, "A field experiment comparing information-privacy values, beliefs, and attitudes across several types of organizations." *Journal of applied psychology*, vol. 68, no. 3, p. 459, 1983.

[22] J. Becker and H. Chen, "Measuring privacy risk in online social networks," in *Proceedings of the 2009 Workshop on Web*, vol. 2, 2009.

[23] C. Renner, "Privacy in online social networks," Ph.D. dissertation, master's thesis, Swiss Federal Institute of Tech., Zurich, 2010.

[24] A. Bonti, M. Li, L. Gao, and W. Shi, "Effects of social characters in viral propagation seeding strategies in online social networks," in *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications.* IEEE, 2012, pp. 632–639.

[25] E. M. Maximilien, T. Grandison, K. Liu, T. Sun, D. Richardson, and S. Guo, "Enabling privacy as a fundamental construct for social networks," in *Computational Science and Engineering, 2009. CSE'09. International Conference on*, vol. 4. IEEE, 2009, pp. 1015–1020.

[26] J. Anderson, "Privacy engineering for social networks," Ph.D. dissertation, University of Cambridge, 2013.

[27] J. Domingo-Ferrer, "Rational privacy disclosure in social networks," in *International Conference on Modeling Decisions for Artificial Intelligence.* Springer, 2010, pp. 255–265.

[28] R. K. Nepali and Y. Wang, "Sonet: A social network model for privacy monitoring and ranking," in *2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops.* IEEE, 2013, pp. 162–166.

[29] N. Talukder, M. Ouzzani, A. K. Elmagarmid, H. Elmeleegy, and M. Yakout, "Privometer: Privacy protection in social networks," in *Data Engineering Workshops (ICDEW), 2010 IEEE 26th International Conference on.* IEEE, 2010, pp. 266–269.

[30] S. Ghanei and K. Faez, "Localizing scene texts by fuzzy inference systems and low rank matrix recovery model," *Computer Vision and Image Understanding*, vol. 142, pp. 94–110, 2016.

[31] M. Grabisch, H. T. Nguyen, and E. A. Walker, *Fundamentals of uncertainty calculi with applications to fuzzy inference*. Springer Science & Business Media, 2013, vol. 30.

[32] D. Wang, X.-J. Zeng, and J. A. Keane, "A simplified structure evolving method for mamdani fuzzy system identification and its application to high-dimensional problems," *Information Sciences*, vol. 220, pp. 110–123, 2013.

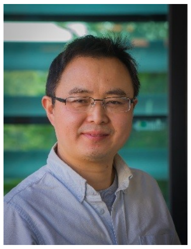
[33] A. A. Torres-García, C. A. Reyes-García, L. Villaseñor-Pineda, and G. García-Aguilar, "Implementing a fuzzy inference system in a multi-objective eeg channel selection model for imagined speech classification," *Expert Systems with Applications*, vol. 59, pp. 1–12, 2016.

**Erfan Aghasian** is currently a second year PhD student in Information Technology (IT) at the University of Tasmania. He received B.Eng degree in Information Technology at Qazvin Azad University (Barajin) and MSc. degree in Information Technology Management at University Technology of Malaysia (UTM). His research interests include Computer Systems and Network Security, Data Security and Data Anonymization.

**Saurabh Garg** is a lecturer at the University of Tasmania, Australia. He is one of the few Ph.D. students who completed in less than three years from the University of Melbourne. He has published more than 40 papers in highly cited journals and conferences. During his Ph.D., he has been awarded various special scholarships for his Ph.D. candidature. His research interests include resource management, scheduling, utility and grid computing, Cloud computing, green computing, wireless networks, and ad hoc networks.

**Longxiang Gao** Longxiang Gao received his PhD in Computer Science from Deakin University, Australia. He is currently a Lecturer at School of Information Technology, Deakin University to teach database, web development and networking units for both undergraduate and postgraduate students. Before joined Deakin University, he was a post-doctoral research fellow at IBM Research & Development, Australia. In IBM R&D, he had been the core team member to develop a crisis event analysis, computing and reporting system to Australia Red Cross, and this project has been selected as the feature project of IBM. His research interests include data processing, mobile social networks and Fog computing.

Dr. Gao has over 30 publications, including patent, monograph, book chapter, journal and conference papers. Some of his publications have been published in the top venue, such as IEEE TMC, IEEE IoT, IEEE TDSC and IEEE TVT. He has being Chief Investigator (CI) for more than 10 research projects. He received 2012 Chinese Government Award for Outstanding Students Abroad (Ranked No.1 in Victoria and Tasmania consular districts). Dr. Gao is active in IEEE Communication Society. He has served as the TPC co-chair, publicity co-chair, organization chair and TPC member for many international conferences.



**Shui Yu** is currently a Senior Lecturer of School of Information Technology, Deakin University. He is a member of Deakin University Academic Board (2015-2016), a Senior Member of IEEE, and a member of AAAS and ACM, the Vice Chair of Technical Committee on Big Data Processing, Analytics, and Networking of IEEE Communication Society. Dr Yu's research interest includes Security and Privacy in Networking, Big Data, and Cyberspace, and mathematical modelling. He has published two monographs and edited two books, more than 150 technical papers, including top journals and top conferences, such as IEEE TPDS, IEEE TC, IEEE TIFS, IEEE TMC, IEEE TKDE, IEEE TETC, and IEEE INFOCOM. Dr Yu initiated the research field of networking for big data in 2013. His h-index is 25. Dr Yu actively serves his research communities in various roles. He is currently serving the editorial boards of IEEE Communications Surveys and Tutorials, IEEE Transactions on Computational Social Systems, IEEE Access, IEEE Journal of Internet of Things, IEEE Communications Magazine, and a number of other international journals. He has served more than 70 international conferences as a member of organizing committee, such as publication chair for IEEE Globecom 2015 and 2017, IEEE INFOCOM 2016 and 2017, TPC co-chair for IEEE BigDataService 2015, IEEE ATNAC 2014, IEEE ITNAC 2015; Executive general chair for ACSW2017.

**James Montgomery** (M'11) received a BInfTech(Hons 1) degree from Bond University, Gold Coast, Australia, in 2000, and a PhD in computer science, also from Bond University, in 2005. He is currently a lecturer in ICT at the University of Tasmania, Hobart, Australia. He has previously held postdoctoral positions at Swinburne University of Technology and the Australian National University. His research interests span evolutionary computation, machine learning and web services.