Personal Privacy Metric based on Public Social Network Data

Wei Qing HUANG¹, Jian Feng XIA¹, ², Min YU¹, ², *, Chao LIU¹

¹Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
²School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

*corresponding author
yumin@iie.ac.cn

Abstract. Attackers can exploit the published data of the social network by the technology of big data analysis to find the user’s privacy without any permission or friendship. In this paper, to solve the problem of privacy metric under the undefined background knowledge, we derive a new metric to quantify the privacy in the complex network circumstances which is inspired by a theory named set pair analysis, meanwhile we put forward a privacy measurement model based on that. At last, we carry out an experiment with the open data in social network. The result shows that the proposed method can achieve the goal of privacy measures under the uncertain background knowledge.

1. Introduction

With the development of social network and big data analysis technology, more and more people suffer property and even loss of life, due to personal privacy was compromised. In August 2016, ‘XU incident’ attracted a great deal of attention in China, the girl committed suicide by leaking personal information. In 2017, 360 Technology Co Ltd released "2016 Websites Disclose Personal Information Analysis Report ", which stated that about 4.01 billion users behaviors could be leaked by 62.4 percent of web sites vulnerabilities, what is more serious is that a website vulnerability can cause two or three different types of personal information to be compromised. How to solve the problem of information use and privacy protection has become a difficult problem of current research.

Attacks on data and personal information are very serious, as benefits drive and the development of technology, such attacks are likely to increase more. With the extensive use of smart phones and laptops, social applications such as QQ, WeChat, MicroBlog, and so on, which have become more and more popular and used to contact friends or make people relaxation. Therefore, people have post their personal private information on the social networks for some purposes, for example, education and work experience, phone numbers, emails, photos, comments and online activities. What’s more, some social applications have authority to publish users’ location and other sensitive information on public. But the excessive personal information will bring us potential threat, so it is meaningful to study the problem of public data on privacy and to measure the degree of privacy exposure. Personal privacy metric of personal privacy on social network or applications can not only reflect the possibility or severity of privacy thus to warn the excessive exposure of one’s privacy, at the same time it can also be used to measure the effectiveness of the privacy protection method. There are many researches and techniques proposed to solve this problem, however, most existing approaches assume specific and restrict network structure as background knowledge and ignore semantic level prior belief of attackers,
which are not always realistic in practice and do not apply to arbitrary privacy scenarios. Our work in this paper addresses this problem on both attributes and semantic level analyses.

The rest of this article is organized as follows. Section 2 summarizes the related research in social network privacy metric and issues that need to resolve. Section 3 presents the privacy metric model and the method of network personal privacy measurement. Section 4 combines the experiments, analyzes the privacy problems of the social network in detail and uses the privacy metric method of this article to measure it. Finally, we conclude in section 5.

2. Related Work

In a social network, users can often protect their privacy by setting privacy policies or by some default properties. But in order to obtain the ideal social effects, users and social platform service providers will retain a certain amount of information in an open form to let others access. For a single social network site, the default part of the privacy properties can play a role in privacy protection. However, different social network service providers have different default public properties and different social focus, which leads to the possibility of the user-sensitive information cross disclosure. For example, the attribute value in a social networking platform cannot be obtained, but in the B social platform can be. So in the social data environment, there is a threat of privacy disclosure to the users.

The research on privacy is mainly about privacy preservation in the meanwhile research of privacy metric is relatively less, which mainly focuses on the privacy metric of position or track[1-6], by the information entropy, the probability of deducing true location and sustainable track time measure, etc. It is widely accepted that data opened to the public is not private, the information that do not want anyone else know or only known by the particular group is recognized as the privacy. So people often ignore the secure of their own information on the net-work public information. Academic research shows that in large data era using data mining and other analytical tools, the privacy information can also be leaked out[7-12]. This kind of privacy problem mainly reflects in social network privacy which belongs to the network privacy preservation of particular application environment. As a result of the mobile Internet, LBS, social network privacy preserving needs both position and data privacy, more complexity and more challenges in terms of privacy metric.

At present, there are few researches on the social network privacy metric. The paper [13] using the random sampling questionnaire survey data method, the formulation correlation factor scale, has studied the influence Internet user network privacy preserving behavior. Through the analysis of relevant theories, the paper [14] put forward a concept model of the influence factors of network privacy concern and behavior intention, but had no data collection and correlation analysis. And paper [15] quantitates the models’ privacy from the perspective of the actual user's privacy information preservation, considering of user and system center's impact on the privacy of net-work users. According to the privacy settings of Renren and Sina, the paper [16] proposes a privacy quantitative model, and uses the theory of the right of business and the data statistics to analyze the privacy concerns of users. But its research model and analysis object are only for three attributes: the permission of strangers to visit your home page; the reception of stranger's information; the application of the location information service. However, it is independent on the personal privacy information which is involved. Abroad research takes the use of noise based data perturbation techniques, the original data C data interval [a, b], and evaluates the real data and the closeness of the disturbance data to measure the degree of privacy [17]. By the privacy field involved in the published micro-blog, the personal privacy data can be quantified into a certain degree of uncertainty. For example, social network uses the mentioned name, address and other privacy information to quantify the privacy of the article [18]. The article [19] proposes a privacy index function PIDX (i, j) to value a user’s privacy exposure, but it only apply in one social network what is the limitation of this theory.

All of the research above has some limitations, mainly to quantify some of the attributes or limit the research done under certain laboratory conditions which are not always realistic in practice and do not apply to arbitrary privacy scenarios. However, the actual social network environment is very complex, especially on the analysis of large data privacy mining. Attackers can use all the information
collected in the social network (background knowledge) to analyze the user’s information, because the attacker cannot know master background knowledge, resulting in a lot of uncertainty, difficult to measure privacy. Since the user can not know the amount of the attacker’s back-ground knowledge, which brings a lot of uncertainty, it is difficult to measure the privacy. Based on the improvement of K anonymity, article [20] is the first to mention the background knowledge attack, but it only considers the privacy metric in the case of missing a sensitive property. On the basis of that, David Martin defines the range of background knowledge by a set of K implicit language related to user's sensitive attributes [21]. Paper [22] proposed using a tuple (l, k, m) to describe the range of background knowledge about the target user in which 'l' indicates the number of missing value sensitive attributes, 'k' indicates the number of sensitive attributes of other related users and 'm' indicates the number of the same attributes as other related users including the target user. The article [23] proposes three different levels of assumptions about the attacker's background knowledge, using the conditional probability to represent the correlation strength between the background knowledge and the user. Although all the above studies have considered the problem of background knowledge, but it does not reflect the uncertainty of the background knowledge, the approach of the research only extends the single attribute to multiple attributes. The background knowledge is delineated into a range, and then converted to the certain probability to measure. There is still a problem in nature, and it does not have universal. Thesis [24] tries to measure the semantic level privacy of social network based on background knowledge.

In view of the complexity of privacy metrics in the context of background knowledge, the traditional methods fail to meet the demands. Our work addresses this problem on both attributes and semantic level analyses and we divide the results of personal privacy metric into two parts: the certain value and uncertain parameter. The determined value can reflect the degree of privacy exposure, and the uncertain parameter can also indicate the impact of the changes of the background knowledge on the degree of privacy exposure. Privacy analysis can be carried out under the limited conditions of data obtaining and also the dynamic adjustment in the changing background knowledge. It is more applicable to the complex situation.

3. Privacy Metrics and Model
The public data in this paper refers to the data which can be obtained without any friend relationship. For the public data, the user cannot control what the obtainer applied this data to do. If the data is used to analyze the user's privacy information, we consider that the public data is privacy data as well. In this case, we treat the public data as the privacy measure, which can reflect the overall privacy situation of the user.

3.1 Set Pair Analysis Theory
Set pair analysis theory is a kind of research method to solve the theory of uncertainty and certainty, which is presented by domestic scholar Keqin Zhao [25]. This theory can handle with random, indistinct and uncertain problems. Because the theory has certain advantages in dealing with multi factors and levels, it has been widely recognized and applied in practical applications.

This theory takes complicated things as a set pair to analyze. The set pair system H is composed of set A and set B, which can be expressed as H=(A,B). Under a certain background (set to W), the set of all S attributes of the S is analyzed. P attributes are a common part of the set A and B. N attributes are opposite in the set A and set B. The membership degree of the remaining U=S-P-N attribute is temporarily uncertain, P, N, U∈R+. The ratio P/S is identical degree of set pair H under the background W. N/S is the degree of opposition, U/S is uncertainty. The whole system is studied with a complete connection degree \( r(W) \):

\[
r(W) = \frac{P}{S} + \frac{U}{S} + \frac{N}{S}
\]  

It is simplified to \( r = a + bi + cj \), \( r \in [0,1] \), a, b, c meet the normalization: \( a+b+c=1 \). i represents
uncertainty, the $j$ value of $-1$ represents the opposition. Based on the theory above and according to the practical problems and needs, we can make it expanded and extended. What’s more, this theory has made a lot of achievements in the fields of artificial intelligence, system theory, program evaluation, information retrieval, medicine, biology, weather prediction, structural mechanics, control theory, disaster assessment and so on at present.

3.2 Privacy Metrics Model and Methods

The current social networking sites and social networking applications are numerous. Typically, users will be registered on a number of sites or applications and fill out personal information. Due to the different privacy policies of each website, the attribute data which can be obtained through the public means is varied. In order to measure the user's privacy exposure in social networks better, obtaining more access to personal information as much as possible is necessary.

One user registers in the $m$ social networking sites and fills out personal information. Privacy property list is $Q \{q_1, q_2, q_3 \ldots q_n\}$, and $q_1, q_2, q_3 \ldots q_n$ are independent. The user's privacy $X$ can be expressed as a two-dimensional matrix:

$$
X = \begin{bmatrix}
    x_{11} & x_{12} & \cdots & x_{1n} \\
    x_{21} & x_{22} & \cdots & x_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{m1} & x_{m2} & \cdots & x_{mn}
\end{bmatrix}
$$

(2)

Matrix $X$ is the user's privacy property matrix. $x_{kh}$ is value of attribute $q_h$ in the $k$ social networking site or application, $k=1,2,3 \ldots m$, $h=1,2,3 \ldots n$.

For properties $q_h=\{x_{1j}, x_{2j}, x_{3j}, \ldots x_{mj}\}$, the user's privacy is set $A$, and the public data is set $B$. Use set pair to represent properties $q_h$:

$$
T_{qh} = (A,B)
$$

(3)

For property $q_h$, connection degree of privacy metrics is

$$
r_{qh} (\text{public data}) = \sum_{h=1}^{n} \frac{P_h}{m} + \frac{u_h}{m} I
$$

(4)

$r_{qh} \in [0,1], h=1,2,3 \ldots n, i \in (0,1]$. $P_h$ represents the number of the $q_h$'s different attribute values that have been acquired already. $U_h$ represents the number of null values for certain property.

Using information entropy represent the average information content of the $q_h$ property in the $m$ social networking websites which can obtained by the attacker is that:

$$
H(\langle q_h \rangle) = - \sum_{h} p_h \log p_h - u_h \log u_h i
$$

The $p_h$ represents the probability of each value of the $q_h$ attribute. $u_h$ represents the proportion of the null value of the $q_h$ attribute, which is $\frac{u_h}{m}$. Some attribute values which cannot be obtained directly can be obtained by using the deeper data mining. The total number of attributes extracted from the overall data is $N$, and the total number of valid data is $C$. In the valid data, the number of different attribute values is $\{C_1, C_2, C_3 \ldots C_l | C_1+C_2+C_3+\ldots+C_l = C, l \leq C\}$. When $C_1 < \varepsilon < C$, $C_1$ are the noise data. The information entropy of this attribute is

$$
H(\langle q_h \rangle) = - \sum_{c \in \varepsilon} p_h \log p_h - \sum_{c \in \varepsilon} u_h \log u_h i
$$

$p_h = p(C_i|N) = p(C) \cdot p(C_i|C)$, $u_h$ is similar. The whole user privacy metrics model can be expressed as
\[ H(Q) = - \sum \sum_C p_h \log p_h - \sum \sum_C u_h \log u_h = a + bi \]  \hspace{1cm} (5) 

### 3.3 Privacy Disclosure Rating Criteria

Obtaining user’s \( n \) attribute values through \( m \) social data source, we can find that if the property does not exist the real value, the reveal of privacy is 0, else, according to the general K anonymous criterion and maximum entropy principle, the attribute of privacy minimum probability is \( 1/m \), corresponding to that the maximum entropy is \( \log \frac{1}{m} \). Therefore, we can transform the qualitative evaluation of the leakage of the attribute privacy into 5 grades, as the following table:

| Attribute Value Probability | Grade Evaluation |
|-----------------------------|------------------|
| \([0,1/m]\)                | Very safe        |
| \((1/m,1/4]\)              | Basic safe       |
| \((1/4,1/2]\)              | Unsafe           |
| \((1/2,3/4]\)              | Very unsafe      |
| \((3/4,1]\)                | Extremely unsafe |

When there is no real value, the corresponding probability is 0, the entropy is not significant, the privacy of personal information can be divided into 5 levels as follows:

| Privacy metrics \( H(Q) \) | Grade Evaluation |
|----------------------------|------------------|
| \(n \log m\)              | Very safe        |
| \([n \log 4, n \log m]\)  | Basic safe       |
| \([n \log 2, n \log 4]\)  | Unsafe           |
| \([n \log \frac{1}{2}, n \log 2]\) | Very unsafe   |
| \([0, n \log \frac{1}{2}]\) | Extremely unsafe |

For the formula: \( H(Q) = a + bi \), \( a \) is set pair potential of \( H(Q) \), according to the theory of set pair analysis, when \( a \) in one level of the rating charting, if \( \frac{a}{b} < 1 \), turn the \( H(Q) \) a lower level.

### 4. Experiment

#### 4.1 Data Acquisition

The data sources are mainly from Sina microblog, renren.com, Tencent microblog, LinkedIn, dajie.com, QQ, WeChat, maimai.cn, which \( m=8 \). Due to the large number of users and high openness, we can use the web crawler to get a large amount of relevant data for data mining analysis to get more user privacy information.

To select users around as the experimental object, so that we can obtain all the real personal information. In order to measure the degree of privacy exposure in social networks and collect information on the research objects, we determine the list of the user's personal information privacy \{name, gender, age, ID number, mobile phone, email, birthplace, residence, work unit, industry, graduate school, academic degree, profession, interest and hobby\} primarily, where \( n=14 \). The
improvement of privacy information is divided into two steps. The first step is to improve the privacy list by direct access to the information, as Table 3 shows. Next step is the data analysis, through clustering, classifying and other algorithms to speculate other privacy information.

Table 3. User Privacy Acquisition Based on Public Data of Social Network

| Name       | Gender | Age | Mobile | Email                  | Location | Birthplace |
|------------|--------|-----|--------|------------------------|----------|------------|
| renren.com | T      | 23  | x      | x                      | Beijing  | Beijing    |
| Sina microblog | F    | Female | 23 | x | jiang***@163.com | Beijing | x          |
| Tencent microblog | F    | Female | x | x | 7980***@qq.com | Beijing | x          |
| LinkedIn   |        |     | x      | x                      | x        | x          |
| dajie.com  | T      | Female | x | x | x | Beijing | x          |
| QQ         | x      | Female | 100 | x | 7980***@qq.com | Shenzhen | Beijing   |
| maimai.cn  | T      | Female | x | x | x | Beijing | x          |
| WeChat     | x      | Female | x | x | x | Beijing | x          |

| Residency  | Company | Occupation | Graduate School | Academic Degree | Profession | Interest and Hobby |
|------------|---------|------------|-----------------|-----------------|------------|--------------------|
| renren.com |         |            | Beijing College of Politics and Law | Bachelor Degree | Law        | x                  |
| Sina microblog |     |            |                |                 |            |                    |
| Tencent microblog |     |            |                |                 |            |                    |
| LinkedIn   |         |            | Beijing College of Politics and Law | Master Degree | Law        | x                  |
| dajie.com  |         | Headline Daily Personnel Management |                |                 |            |                    |
| QQ         |         |            | Secondar y     |                 | x          | x                  |
| maimai.cn  |         | Headline Daily Personnel Management |                |                 |            |                    |
| WeChat     |         |            |                |                 |            |                    |

For the purpose of preserving the user's privacy, the real name in Table 1 is T, and false name is F. The email information is omitted. As seen from the above table, in addition to mobile phones, work and hobbies cannot be directly access, other privacy attributes can be obtained through different social networking sites or applications.

4.2 Data Analysis
The data Web crawlers obtain the public content of the social network, such as ‘sharing’, ‘status’, ‘log’
in Renren; ‘Microblog’, ‘concern list’, ‘fan list’ in Sina, ‘broadcast’, ‘listen list’, ‘audience list’ in Tencent. And through the segmentation engine which is provided by SAE (Sina App Engine), we can extract keywords from the content and compare keywords with the established thesaurus to get the interest hobby and the mobile phone information. In the ‘concern’ list, ‘fans’ list, ‘listening’ list, ‘audience’ list, classifying the users which have a strong association with the target (mutual concern) and obtaining the position information of the social circle, in order to acquire the resident area. Table 4 is the result of the statistical classification of the 2268 items of the interests and hobbies which the user open to the public. Table 5 and Table 6 are the statistical classification of user’s social relation situation and the distribution area of strong relationships.

Table 4. The Proportion of Interest and Hobby in Publishing Information

| Type  | Travel | Food | Cosmetic | Sports | Sum |
|-------|--------|------|----------|--------|-----|
| Number| 284    | 272  | 124      | 42     | 722 |
| Ratio | 12.52% | 11.99%| 5.46%    | 1.85%  | 31.83% |

Table 5. Distribution of User Social Relations

| Type         | Strong Relation | weak Relation |
|--------------|-----------------|---------------|
| Number       | 94              | 114           |
| Ratio        | 45.19%          | 54.81%        |

Table 6. Distribution of Users in Strong Relationships

| Type      | Beijing | Shanghai | others |
|-----------|---------|----------|--------|
| Number    | 69      | 6        | 19     |
| Ratio     | 73.4%   | 6.38%    | 20.22% |

4.3 Personal privacy metrics
The data We According to the privacy measurement model, the metrics of whole individual privacy are the sum of the individual privacy attributes. Therefore, it is needed to analyze the user's attributes. For the sake of the space, this paper takes the name and the residence measurement as the example.

The total number of name is 8, real name is 3, false name is 2, and unknown is 3. The connection degree of the name can be expressed as:

\[ R_{\text{name}} = p_{\text{true}} + p_{\text{false}} + p_{\text{NA}} \]

\[ = \frac{3}{8} + \frac{2}{8} + \frac{3}{8} \]

\[ H(\text{name}) = - \sum_{i} p_{\text{name}} \log p_{\text{name}} \]

\[ = - \frac{3}{8} \log_{2} \frac{3}{8} - \frac{2}{8} \log_{2} \frac{2}{8} - \frac{3}{8} \log_{2} \frac{3}{8} \]

\[ = 1.03 + 0.53i \]

The result means that by the interference of two in-accurate names and null value, he attacker wants to confirm the user’s real name, requiring at least 1.03bit data. In some cases, there is no access to the name information, so the measurement results are still variable. If attackers get to the real name
by further analysis and the result is still a false name, the privacy will be increased.

The true residence of the user is Beijing. In its 208 concerns, there are 94 for mutual concerns, which can be seen as a group with the user's social more closely, and the analysis is also more meaningful. As analysis of the location of 94 users, it includes 69 in Beijing, 6 in Shanghai. The other is dispersed, belonging to the outlier noise point. Among that, Beijing and Shanghai are the focuses of attacker’s analysis. According to the theory of set pair analysis, the total number is 208, the support of the count is 69, the opposition interference number is 6, the privacy of the residence can be expressed as:

\[
R_{\text{residence}} = p_{\text{beijing}} + p_{\text{shanghai}} + p_{\text{other}}
\]

\[
= \frac{69}{208} + \frac{6}{208} + \frac{133}{208}
\]

\[
(9)
\]

\[
H(\text{residence}) = - \sum p_h \log p_h - u_h \log u_h i
\]

\[
= \frac{69}{208} \log_2 \frac{69}{208} - \frac{6}{208} \log_2 \frac{6}{208} - \frac{133}{208} \log_2 \frac{133}{208}
\]

\[
= 0.68+0.41i
\]

Similarly, other attribute privacy metrics are shown in table 7.

| Attribute      | Privacy Uncertainty | Attribute      | Privacy Uncertainty |
|----------------|---------------------|----------------|---------------------|
| Name           | 1.03+0.53i          | Residence      | 0.68+0.41i          |
| Gender         | 0.17+0.375i         | Unit           | 0.5+0.31i           |
| Age            | 0.875+0.42i         | Industry       | 0.5+0.31i           |
| Mobile         | NA                  | Graduate School| 0.5+0.31i           |
| Email          | 0.875+0.42i         | Academic Degree| 1.125+0.42i         |
| Location       | 0.685+0.375i        | Profession     | 0.5+0.31i           |
| Birthplace     | 0.5+0.31i           | Interest and Hobby | 0.526+0.377i    |

The user's entirely social network public data privacy uncertainty \( H(Q) \) is 8.466+4.877i. If the attackers want to eliminate all the uncertainty to get the correct data in the list of user attributes, they will have to obtain at least 8.466bit. Due to the fact that, 8.466 \( \in [n\log_4 2, n\log_2 2] \), \( n=14 \) and 8.466/4.877 > 1, based on the definition of the privacy evaluation model of this paper, it can be regarded as a bad privacy preservation. The reason is that the records is mainly about the real information makes the data easily to be used and altered by the attackers. In the same way, we can compute another social network’s privacy evaluation \( H(Q) \) is 8.11+4.865i, even if 8.11/4.865>1, the degree of privacy disclosure is more serious than 8.466. In addition to using vertical calculation to
obtain the privacy of a single attribute, horizontal calculation can get the user's privacy in a particular social networking sites or applications, which makes this method flexible.

5. Conclusion
Based on the theory of set pair analysis, this paper combines the information of public data and data mining analysis to measure the privacy of personal social network public data. It mainly solves the problem of privacy metrics in the context of the network big data and uncertain background knowledge environment. This paper lists some common social networking sites or applications while the method described in this paper can be extended, the principle and nature does not change. For the entire Internet, the data obtained and the results of the analysis are more sensitive and credible. Privacy issues are more serious, but it also can carry out privacy metrics using the above methods.

In all, there is a need for further research on the weight of privacy attributes in privacy metrics. Because, the importance among the attributes is different, when privacy value is small, the actual exposure situation is more serious. Furthermore, next step we will study and improve the problem to make it perfect and complete.

Acknowledgment
This work is supported by National Natural Science Foundation of China (No. 61173008, 61402124, 61303244).

References
[1] Eurocode 3: Design of Steel Structures: Part 1-1: General Rules and Rules for Buildings[S]. European Committee for Standardization, 2005.
[2] Zhou F, Chen Y, Young B. Cold-formed high strength stainless steel cross-sections in compression considering interaction effects of constituent plate elements[J]. Journal of Constructional Steel Research, 2013, 80: 32-41
[3] Schafer B W. Review: the direct strength method of cold-formed steel member design[J]. Journal of constructional steel research, 2008, 64(7): 766-778.
[4] Li Y. Extension of the Direct Strength Method to hot-rolled and welded H profile cross-sections[D]. Université de Liège, Liège, Belgique, 2014.
[5] Hibbett, Karlsson, Sorensen. ABAQUS/standard: User's Manual[M]. Hibbitt, Karlsson & Sorensen, 1998.
[6] Joanna Nseir. Development of A New Design Method for the Cross-Section Capacity of Steel Hollow Sections[D]: Université de Liège, Liège, Belgique, 2015.
[7] ECCS, Ultimate Limit State Calculation of Sway Frames with Rigid Joints [J]. ECCS publications, 1984.
[8] Eurocode 3: Design of Steel Structures: Part 1-5: Plated Structural Elements[S]. European Committee for Standardization, 2006.
[9] ANSI B. AISC 360-10-Specification for Structural Steel Buildings[J]. Chicago AISC, 2010.
[10] GB. 50010-2003, Code for design of steel structures[S]. Beijing: China Architecture & Building Press, 2003.
[11] Shokri R, Theodorakopoulos G, Le Boudec J Y, et al. Quan-tifying location privacy[C]//2011 IEEE Symposium on Se-curity and Privacy. IEEE, 2011: 247-262.
[12] Ma Z, Kargl F, Weber M. A location privacy metric for v2x communication systems[C]/Sarnoff Symposium, 2009. SARNOFF'09. IEEE. IEEE, 2009: 1-6.
[13] Gao F, He J, Peng S, et al. A Quantifying Metric for Privacy Protection Based on Information Theory[C]/Intelligent In-formation Technology and Security Informatics (IITSI), 2010 Third International Symposium on. IEEE, 2010: 216-220.
[14] Wang C M, Guo Y J, Guo Y H. Privacy metric for user's trajectory in location-based services[J]. Ruanjian Xuebao/Journal of Software, 2012, 23(2): 352-360.
[15] Zhang X J, Gui X L, Feng Z C, et al. A quantifying frame-work of query privacy in location-based service [J]. Journal of Xi’an Jiaotong University, 2014, 48(2): 8-13.

[16] Xinxin Zhu. Research on Location Privacy Metric Based on a Common Framework LPPM [D]. Central China Normal University, 2014.

[17] Yang C C. Information sharing and privacy protection of terrorist or criminal social networks[C]//Intelligence and Security Informatics, 2008. ISI 2008. IEEE International Conference on. IEEE, 2008: 40-45.

[18] Hsieh C L. Privacy Disclosure: Personal Information and Images on Social Networking Sites in Taiwan[C]//Applications and the Internet, 2009. SAINT’09. Ninth Annual International Symposium on. IEEE, 2009: 294-295.

[19] Pimenta P C, de Freitas C M. Security and privacy analysis in social network services[C]//5th Iberian Conference on Information Systems and Technologies. IEEE, 2010: 1-6.

[20] Cutillo L A, Molva R, Onen M. Analysis of privacy in online social networks from the graph theory perspective[C]//Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE, IEEE, 2011: 1-5.

[21] Chester S, Srivastava G. Social network privacy for attribute disclosure attacks[C]//Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on. IEEE, 2011: 445-449.

[22] Shaoqing LU, Yuqing ZHANG, Ping NI. Privacy leakage in online social networks based on public information [J]. Journal on Communications, 2013, 1.

[23] Kang GAO X Y. Factors Affecting Internet Users' Information Privacy Protection [J]. Journal of Intelligence, 2011, 4.

[24] Xiao J, Shao-bo J I. Conceptual model of the factors influencing consumer online privacy concern and behavior intention[J]. Science-Technology and Management, 2009, 5: 021.

[25] Wang B, Duan Y. Research on information privacy quantization method facing ubiquitous computing environment [J]. Jisuanji Gongcheng yu Yingyong(Computer Engineering and Applications), 2011(27).

[26] Hanyu Z H U, Lianren W U, Tingjie L U. Research on quantifying user privacy on social networking sites [J]. Journal of Tsinghua University (Science and Technology), 2015, 54(3): 402-406.

[27] Agrawal R, Srikant R. Privacy-preserving data mining[C]//ACM Sigmod Record. ACM, 2000, 29(2): 439-450.

[28] Ngoc T H, Echizen I, Komei K, et al. New approach to quantification of privacy on social network sites[C]//2010 24th IEEE International Conference on Advanced Information Networking and Applications. IEEE, 2010: 556-564.

[29] Wang Y, Nepali R K, Nikolai J. Social network privacy measurement and simulation[C]//Computing, Networking and Communications (ICNC), 2014 International Conference on. IEEE, 2014: 802-806.

[30] Machanavajjhala A, Kifer D, Gehrke J, et al. 1-diversity: Privacy beyond k-anonymity [J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2007, 1(1): 3.

[31] Martin D J, Kifer D, Machanavajjhala A, et al. Worst-case background knowledge for privacy-preserving data publishing[C]//2007 IEEE 23rd International Conference on Data Engineering. IEEE, 2007: 126-135.

[32] Chen B C, LeFevre K, Ramakrishnan R. Privacy skyline: Privacy with multidimensional adversarial knowledge[C]//Proceedings of the 33rd international conference on Very large data bases. VLDB Endowment, 2007: 770-781.

[33] Wang C M. Research on Trajectory privacy metric in Location-based services [D]. Central China Normal University, 2011.

[34] Qian J, Li X Y, Zhang C, et al. Social Network De-Anonymization and Privacy Inference with Knowledge Graph Model[J]. IEEE Transactions on Dependable and Secure Computing, 2017.
[35] Zhao K Q. Set pair analysis and its preliminary application [J]. Zhejiang Science and Technology Press, Hangzhou, 2000.