Involuntary Information Leakage in Social Network Services *

Ieng-Fat Lam, Kuan-Ta Chen, and Ling-Jyh Chen

Institute of Information Science, Academia Sinica
{iengfat,ktchen,cclljj}@iis.sinica.edu.tw

Abstract. Disclosing personal information in online social network services is a double-edged sword. Information exposure is usually a plus, even a must, if people want to participate in social communities; however, leakage of personal information, especially one's identity, may invite malicious attacks from the real world and cyberspace, such as stalking, reputation slander, personalized spamming and phishing.
Even if people do not reveal their personal information online, others may do so. In this paper, we consider the problem of *involuntary information leakage* in social network services and demonstrate its seriousness with a case study of Wretch, the biggest social network site in Taiwan. Wretch allows users to *annotate* their friends' profiles with a one-line description, from which a friend's private information, such as *real name, age, and school attendance records*, may be inferred without the information owner's knowledge. Our analysis results show that users' efforts to protect their privacy cannot prevent their personal information from being revealed online. In 592,548 effective profiles that we collected, the first name of 72% of the accounts and the full name of 30% of the accounts could be easily inferred by using a number of heuristics. The age of 15% of the account holders and at least one school attended by 42% of the holders could also be inferred. We discuss several potential means of mitigating the identified involuntary information leakage problem.

1 Introduction

Social network services (SNS) represent one of the most important applications of the Internet in recent years, with some SNSs hosting millions of profiles, for example, Myspace, Facebook, Flickr, Orkut, and Yahoo! 360. Such services provide a virtual playground for participants to meet new friends, maintain contact with acquaintances, and share resources with others over the Internet. To let others know about themselves, users normally publish personal information

* This work was supported in part by Taiwan Information Security Center (TWISC), National Science Council under the grants NSC 97-2219-E-001-001 and NSC 97-2219-E-011-006. It was also supported in part by Taiwan E-learning and Digital Archives Programs (TELDAP) sponsored by the National Science Council of Taiwan under NSC Grants: NSC 96-3113-H-001-010, NSC 96-3113-H-001-011 and NSC 96-3113-H-001-012.

online, such as their appearance, nationality, school attendance records, work experience, and hobbies. The information not only lets people know more about a person, but also enables others to find the user through web searches. Thus, users are normally encouraged to disclose personal information in order to receive higher exposure (more "eyeball counts" in Internet jargon) in the community.

Disclosing personal information online is a double-edged sword. Information exposure is usually a plus, even a must, if people want to join certain types of social communities; however, leakage of personal information, especially one's identity, may invite malicious attacks from the real world and cyberspace. Many studies have addressed the problems of privacy invasion and security threats raised by information exposure online, e.g., stalking, reputation slander, personalized spamming and phishing. We discuss some of the threats in Section 6.1 and refer interested readers to [1].

Even if people do not reveal their personal information, *others may do so.* We illustrate how this could happen with the following example. Suppose Alice uses the pseudonym `boulder_987` to protect her identity. Bob, a friend of Alice, may reveal Alice's age, occupation, and even her real name during their interaction in the following ways:

- Bob may recommend[1] Alice as "`the best steak chef in Boston`" on his own page, and thereby inadvertently reveal Alice's occupation and the city she lives in.
- Suppose Bob also uploaded a photo taken with Alice to his online albums. He may annotate the photo with "`Bob and Alice Lewis at George's Wedding`" and link the photo to Alice's profile. This action would reveal Alice's full name and the fact that Bob, Alice, and George know each other.

Thus, Bob may unintentionally reveal a great deal of information about Alice without her knowledge. In other words, Alice's efforts to protect her identity may be easily nullified by others' behavior. Moreover, it is difficult to detect occurrences of such leakages due to their *distributed* nature. This problem, which we call *involuntary information leakage*, is becoming a serious threat to privacy because of the popularity of social network services.

In this paper, we investigate the extent of *involuntary information leakage in social network services.* We analyze data gathered from Wretch, the biggest social network site in Taiwan. The data set contains 592,548 effective profiles, the social connections between the profiles, and annotations describing the social connections. To quantify the degree of such leakages, we attempt to infer the real name, age, and school attendance records of each user based on annotations made by friends. Our results show that the first name of 72% of users and the full name of 30% of users can be easily inferred by a number of heuristics. The age of 15% of users and at least one school of 42% of users can also be inferred. The high ratio of information leakage evidences that users tend to annotate their friends by using real names and by describing their offline relationships. Based on our

[1] Many SNSs provide a recommendation/endorsement system in which a user can "recommend" another user to the public.

analysis results, we consider several possible ways of mitigating the identified involuntary information leakage problem.

The remainder of this paper is organized as follows. In Section 2 we provide an overview of earlier studies related to social networks and online privacy. In Section 3, we describe our data collection procedures and examine the demography and levels of self-disclosure. We investigate the leakage of real names in Section 4 and the leakage of age and school attendance records in Section 5. Section 6 considers threats and risks that may occur due to information leakage and potential solutions to the problem. Then, in Section 7 we present our conclusions.

2 Related Work

Online social network services have attracted the attention of both entrepreneurs and researchers in recent years [2]. From an academic perspective, the services provide the research community with an unprecedented opportunity to analyze the structure and properties of online social networks on a large scale.

Ahn et al. analyzed the structure of online social networks and found that they have many similarities with offline social networks [3]. Mislove et al. conducted a large-scale study of snapshot graph structures of online social networks. They compared the structures of online social networks and the Web, and validated online social networks with the structural properties of offline social networks [4]. Kumar et al. analyzed the evolution of two popular online social networks, and identified the high-connectivity core and the star structure of each network [5]. O'Murchu et al. compared and classified different categories of social and business networking communities [6].

User interaction and relationships in social network services have also been investigated by researchers. Boyd analyzed social network services from the perspective of human factors, and found that knowledge of, or trust between, users is not required to establish online relationships [7]. Moreover, it has been shown that online social networks engender much weaker ties between users than their offline counterparts [2].

Based on $4,000$ profiles gathered from Facebook.com, Gross et al. analyzed the degree of information disclosure and the subsequent risks. They found that *"personal data is generously provided, and limiting privacy preferences are hardly used"* [2], while Acquisti observed that *"technology alone or awareness alone may not address the heart of the privacy problem"* [8]. The privacy issues raised by social network services present a difficult challenge to both information technology and social science researchers.

3 Data Description

In this section, we begin with an introduction to Wretch, the social network service we studied, and then describe our data collection procedures.

3.1 Wretch

Wretch (`http://www.wretch.cc`) was established in 1999 and acquired by Yahoo! Taiwan in 2006. It is currently the most popular social networking site in Taiwan. At the time of writing (Feb 2008), it hosted about 4 million profiles. Like other social network systems, Wretch provides an array of services, including albums, blogs, a bulletin board system (BBS), video sharing, and a discussion forum. Anyone can freely browse all the profiles on Wretch without an account. Joining the service as a registered member is free. Members can upgrade their service levels with a yearly subscription to obtain a larger storage space and more functionalities.

3.2 Data Collection

To collect users' profiles and information about their social relations, we developed a crawler program to fetch profile pages from Wretch. We began the data crawling with an initial set of accounts. In each round, the crawler fetched an account's profile and its friend list; and HTML processing techniques were employed to extract the desired information. If the friend list contained an account the crawler had not seen before, it was added to the job queue. At the end of each round, the crawler randomly selected an account from the queue, and targeted that account in the next round. The crawler continued fetching users' data until the job queue was empty.

We collected the data in September 2007. In sum, we fetched 766,972 profiles, which constituted 20% of Wretch's population at the time. As we focus on name leakages caused by friends, profiles with an empty friend list are irrelevant to our study. Therefore, we removed profiles that did not have any outgoing friend connections. This yielded a reduced set of 592,548 profiles, corresponding to 15% of the population. Our data set is summarized in Table 1.

Table 1. Overview of crawled data

Wretch Data	
Number of users	766,972 (20%)
Number of Effective users	592,548 (15%)
Number of Connections	7,619,212
Avg Connections per user	11.5

3.3 Self-Information Disclosure

Self-disclosure is defined as "*telling others previously unknown knowledge so that it becomes shared knowledge*" [9]. It is normally intended to increase understanding between people, build trust, strengthen the ties between people, and bind romantic relationships or friendships [10].

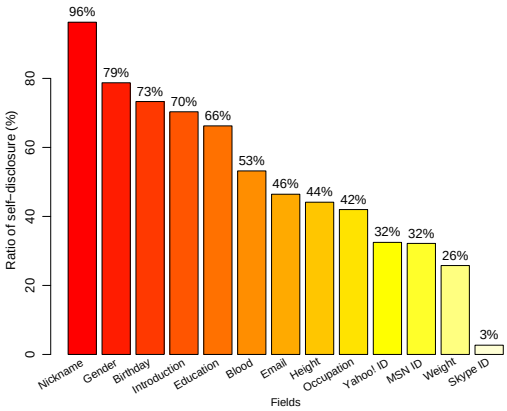

Fig. 1. Ratio of self-disclosure. The gender and birthday fields have a high disclosure ratio.

To determine the degree that Wretch users reveal information about themselves, we summarize the disclosure ratio of personal statistics in Fig. 1. We observe that most users provide gender and birthday information, and many list their email addresses. In addition, details of instant messaging accounts (e.g., MSN Messenger and Yahoo Messenger) are often given, with a disclosure ratio higher than 30%. This is not surprising as real-time messaging applications are now one of the main communication methods used by young people [11].

We define the *degree of self-disclosure* (DSD) in order to quantify a user's tendency to disclose his/her own personal information. A user's DSD is defined as follows:

$$DSD = \sum_{i=1}^{n} F_i \times W_i, \tag{1}$$

where n is the total number of fields, F_i is a binary value indicating whether field i has been completed, and W_i is the weight of field i. We compute W_i by the following equation:

$$W_i = 1 - \frac{R_i}{\sum_{j=1}^{n} R_j}, \tag{2}$$

where R_i is the *disclosure ratio* of field i. The ratio is computed by dividing the number of users who complete field i by the total number of users. For example, if there are $1,000$ users, and 100 of them provided the weight information, the disclosure ratio of the field "weight" would be $100/1000 = 0.1$. We compute the DSD for all fields, except nicknames, introductions, and instant messaging accounts, as the account fields may not be applicable to every user.

4 Involuntary Name Leakage

On Wretch, a user can provide a free-form text (limited to one line) to annotate a friend's profile, which we call *friend annotations* (or *descriptions*). However, people invent their own ways of using this field. After viewing thousands of annotations, we identified some typical patterns and found that a typical annotation is comprised of three parts: 1) a tag, which is used for *classification*; 2) the *real name or nickname* of the friend to be annotated; and 3) a description of the friend's features or the *relationship* between the two friends. Some examples are "∗Beauty∗ Cathy Brown -- The hottest girl of Nightingale High School" and "[[School Mate]] Tony MY BUDDY." In these ways, a great deal of personal information could be revealed through friend annotations without the information owner's knowledge. (Hereafter, we refer to the information owner as the *annotatee*.)

For example, if we find 5 incoming annotations for a user contain the substring "Jane Garcia" and 10 descriptions contain "Jane," then we may safely guess that the annotatee's real name is Jane Garcia.

4.1 Inference Methodology

To infer the real name of a profile, we first collect all of its incoming annotations, i.e., those that the user's friends compile for him/her. Then, from each description, we extract name candidate tokens from the text as follows:

1. We break the text into disconnected tokens by using several types of delimiters, for example:
   (a) symbols: <SPACE>, <TAB>, #
   (b) punctuation marks: ' " , . ( ) [ ]
2. As most Wretch clients use Chinese, we employ Chinese-specific naming rules to determine whether a delimited token is potentially a Chinese name. A Chinese name is usually composed of two or three characters[2] — a one-character family name, e.g., Chen, Wang, Lin, and a one- or two-character first (given) name, e.g., Xin, Kuan-Ta, Ieng-Fat. Thus, if a token contains two or three characters, it could be a Chinese first name or full name, and we consider it as a real name candidate token.
3. We associate each name candidate token with a *duplication count*. For example, if a name candidate token C_1 appears in annotations from three friends, the duplication count of C_1 will be 2.

We summarize the statistics of friend annotations and extracted name candidate tokens in Table 2. In our data set, we find that, on average, a user has 7 online friends and 6.8 incoming annotations from them.

[2] Some rare Chinese family names are comprised of two characters, e.g., Ouhyoung; thus, a four-character full name is possible. However, to avoid false identification of real names, we only consider two- or three-character names.

Table 2. Summary statistics of friend annotations and extracted name candidate tokens.

Friend Annotations and Name Candidates	
Avg In-Degree	7.10
Avg In-Degree with Annot.	6.81 (96%)
Avg In-Degree with Dup. Tokens	3.46 (49%)
Avg # Unique Name Candidates	3.81

Although the friend description is not a required field, 96% of the connections are annotated. In addition, 49% of the incoming annotations for a user contain at least one name candidate token that appears in other annotations for the same user. For each user, we extract an average of 3.8 unique name candidate tokens, which will serve as our input for real name inference.

Inference of Full Names Given the extracted name candidate tokens for a user, we apply the following heuristic rules to infer the user's full real name.

1. **Common Family Name**: We consider a token as a full real name if 1) its first character is a common family name (according to the 100 family names listed in [12]); and 2) its duplication count is greater than 1.
2. **First Name as a Substring of the Full Name**: We consider a token C_i as a full real name if 1) there is another token C_j equal to C_i without its first character; and 2) the duplication count of C_i is greater than 1. For example, if C_i is "Wang Ta-Ming," C_j is "Ta-Ming," and C_i appears more than once, then we consider that C_i is probably the full real name of the user.
3. **Common Full Name**: We consider a token as a full real name if it was one of the $574,010$ names on the enrollment list for the national college exams for the years 1994 to 2007 [13].
4. **Nickname Decomposition**: In Chinese, it is common for a person to have a nickname that is derived from his/her family or given names. For example, a man called Wang Ta-Ming may have a nickname like "Old Wang," "Bro Wang," "Bro Ta," "Little Ming," or "Bro Ta-Ming." Generally, for a person with a name in the format "FN GN1-GN2," some possible nicknames could be:
   (a) $prefix + X$,
   (b) $prefix + X + X$,
   (c) $X + postfix$,
   where X could be FN, GN1, GN2, or GN1-GN2. We examined the nicknames that appeared in our data set and manually picked 38 common prefixes and postfixes for nickname composition. We consider that a token C_i is a full real name if 1) it contains the predefined nickname prefixes or postfixes as specified; and 2) after removing the corresponding prefix or postfix, the remaining part is the same as other name candidate tokens.

5. **Common Word Removal**: If we do not find any matched name candidate based on the above rules, a name candidate token is considered as a full real name if 1) its duplication count is greater than 1; 2) it does not contain any nickname-composition prefix or postfix; and 3) it does not contain any general word that people use in daily life, e.g., he, she, friend, lover, classmate, good, or bad. The removal of common words is based on a dictionary containing 100,511 words [14]. If more than one name candidate matches these criteria, then the one with the highest duplication count is deemed the real name of the user.

Inference of First Names The method used to infer first names is the same as that used for full names, except for the following three points. 1) We use first name candidates (two characters) instead of real name candidates (three characters). 2) Because heuristics 1 and 2 are used specifically for full name inference, we only use heuristics 3, 4, and 5 for the first name. 3) A common first name table (comprising 208,581 names) is used in heuristic 3 instead of the common full name table [12]. Also, because users may include the names of other people, such as "Dolly's sister," we only apply rule 3 to name candidate tokens that have a duplication count greater than 1.

4.2 Inference Results

Here, we summarize the results of the name inference procedures. Table 3 lists the ratio of users whose real names we were able to infer. We successfully inferred first names for 72% of users, and full names for 30% of users. If we count both first names and full names, totally 78% of users are subject to the risk of name leakage. In addition, we detail the inference success ratio of each heuristic rule for real names and first names in Table 4.

Table 3. Ratios of Correctly Inferred Names.

Type of name	Ratio of Name Inference
Nickname	60%
Real name	30%
First name	72%
Real name or first name	78%

4.3 Validation

A complete validation of our name inference results was not possible because we did not know the true real names of the users. Therefore, we employed manual validation. By randomly selecting 1,000 profiles, and examining the real names

Table 4. Ratio of Users Whose Names Can be Inferred, by Different Heuristic Rules.

Method	Real Name	First Name
Common family name	3%	N/A
Relation of first name	11%	N/A
Common full/first name	9%	57%
Relation of nickname	2%	14%
Removal of common words	27%	69%

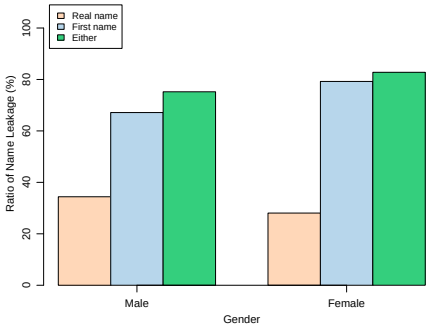

Fig. 2. Ratio of name leakage based on gender.

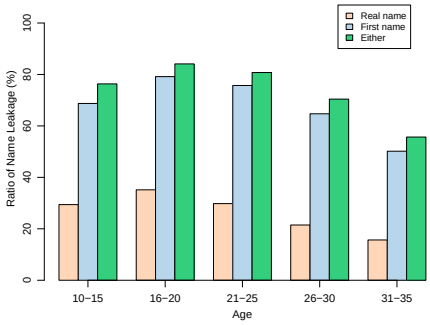

Fig. 3. Ratio of name leakage based on age.

we inferred for them. We believe that at least 738 of the selected names are correct. The majority of cases of incorrectly inferred names are caused by mistaking a user's nickname for the real name, as we cannot enumerate all possible prefixes and postfixes for nicknames derived from real names. Moreover, our methods cannot distinguish a nickname from a real name if the former is not a literal derivative of the latter.

We acknowledge that our proposed heuristics cannot always derive correct real names. However, exact real name inference is not our primary goal. Instead, we seek to verify the qualitative fact that "*involuntary real name leakage occurs in real-life social network systems, and the degree of leakage is significant.*" In this way, our inference results, though not very accurate, are sufficient to support our conjecture.

4.4 Demographic Analysis

Fig. 2 shows that males are susceptible to higher full name leakage risks than females. However, the first names of males are less likely to be released involuntarily than those of females. This interesting finding implies that people are more likely to use full names to describe a male friend and first names to describe a female friend. Fig. 3 shows that the risk of name leakage is lower for older users. This phenomenon can be simply explained by the number of friends

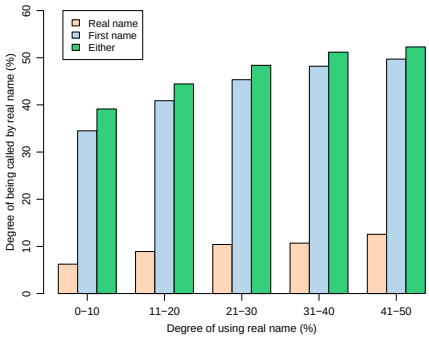

Fig. 4. DUR vs DCR. DCR has a positive relation to DUR.

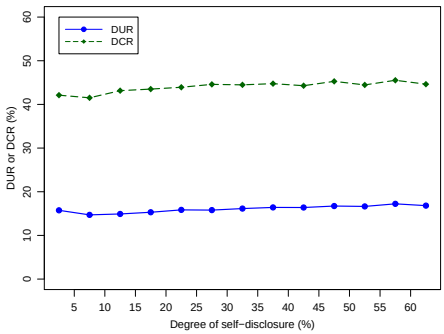

Fig. 5. DCR and DUR distribution over DSD.

that people have in the social network. Because the 16–25 age group constitutes the core population of Wretch and social connections are mostly between users of similar age, younger people tend to have more online friends than other age groups. Therefore, they have more incoming friend annotations and a higher risk of name leakage than users in other age groups.

4.5 Risk Analysis

To confirm that identity leakage is definitely *involuntary*, we checked whether users whose names were inferred correctly had disclosed their real names in their profiles. We found that less than 0.1% of those users had voluntarily revealed their real names (either their full name or first name) in their profiles. This supports our contention that name leakage is caused by annotations made by online friends, rather than self-disclosure.

To determine the source of name leakage, i.e., who reveals the identities of other users, we investigate the usage of real names in friend annotations. We define two metrics to quantify the tendency to use real names when describing online friends:

- **Degree of Using Real Name (DUR)**: DUR quantifies a person's tendency to use real names when annotating his/her friends' profiles. It is computed as the ratio of that person's outgoing annotations that contain the annotatee's real name.
- **Degree of Being Called by Real Name (DCR)**: DCR quantifies the extent that a user is annotated with his/her real name by online friends. It is computed as the ratio of incoming annotations containing the user's real name.

We first investigate whether real name use behavior is *symmetric*, i.e., are DCR and DUR positively correlated? Fig. 4 suggests that there is a consistent positive correlation between the two metrics. The result indicates that users

who receive annotations containing their real names also tend to use real names when sending friend annotations. We also consider the impact of the degree of self-disclosure (DSD) on DCR and DUR, as shown in Fig. 5. Even though both DCR and DUR have a slight positive relationship with DSD, the correlation is insignificant, which indicates that name leakage is not strongly related to self-disclosure. That is, friends may still reveal a user's real name unintentionally no matter how much the user tries to protect his/her identity online.

5 Involuntary Leakage of Age and Education Records

In online social networks, the connections between people are usually a duplicate of their *offline relationships*, which suggests that people have *common attributes*. For example, if the relationship between two people is described as "classmates," they should study in the same school, live in nearby locations, and be a similar age; likewise, the relationship "colleague" implies that two people work in the same organization and probably have similar professions.

If we know two users are classmates in a primary school and one of them disclosed his/her age in the profile, we can infer that the other one is a similar age. To assess the risk of personal information being revealed involuntarily, in the following, we infer the age and education records of users who did not provide this information. In our data set, 56% (331, 827) of users disclosed their ages, but only 10% (59, 255) disclosed their current schools in their profiles. By applying a heuristic inference method, we successfully inferred the ages for 18% of those who did not disclose their ages (15% of the total number of accounts), and inferred at least one school for each of 42% of users in our data set.

5.1 Inference Methodology

Inferring Age Our inference procedures are executed in a round-based manner. Initially, a job queue is filled with all the accounts that provide age information. In each round, we fetch an account X from the queue, and check all of its incoming and outgoing friend annotations. If we determine that a user Y on X's friend list should be the same age, we set the age field of Y equal to that of X and add Y to the job queue. Our heuristic rule for "same age" is based on the offline relationships of "classmate" or "schoolmate in the same year." We identify such relationships by searching with keywords like "classmate", "class leader", "same class," and "same grade" in friend annotations. The inference procedure continues until the job queue is empty.

Inferring Education Records The inference methodology for education records is similar to that for age inference except for three points. First, while a person's age is unique, his/her education records will include information about attending several schools. For simplicity, we assume that each user attended at most one school in each of the four education levels, namely, elementary school, junior high school, senior high school, and college. Second, while users are unlikely to

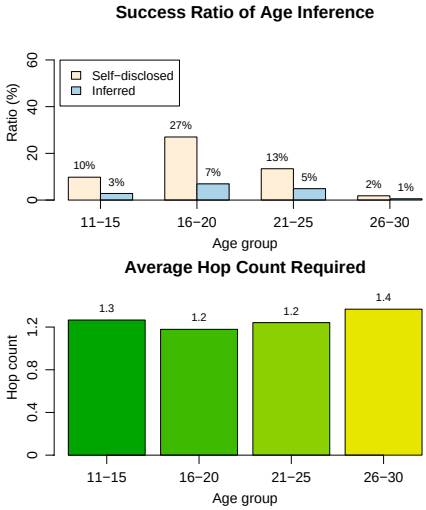

Fig. 6. Inference results for users' ages.

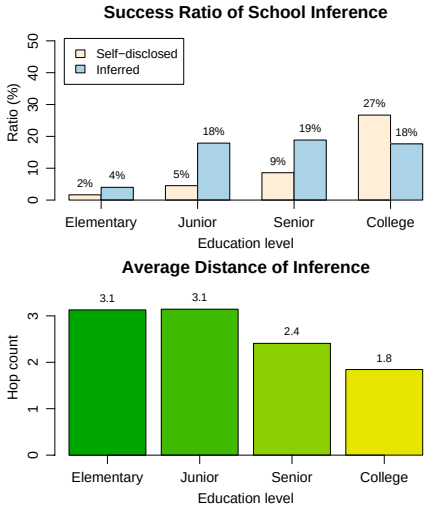

Fig. 7. Inference results for users' education records.

specify a friend's age in annotations, they may use a school's name as a category name. Thus, we can infer not only if two users attended the same school, but also the name of a school the annotatee ever enrolled in. Third, in addition to school names and offline relationships like "classmate" and "schoolmate," we can determine if two users ever studied in the same school by keywords like "same school," "same college," or "*some* department."

5.2 Inference Results

Here we summarize the inference results of age and education records. The upper graph of Fig. 6 shows that the success ratio for age inference is higher for users in the 16–25 year age group. In sum, we successfully inferred the ages of 15% of users, which corresponds to 18% of the users who did not provide age information. The lower graph suggests that the average hop count required for inference is about 1.2, which indicates that, in most cases, the inferred age of a user is directly propagated the profile of a user who discloses his/her own age.

Fig. 7 shows the inference results for users' education records. The success ratios for school inference were approximately 20% for all education levels, except elementary school. Totally, we inferred at least one school for 42% of users in the data set. This success ratio is reasonably high, as only 5% and 9% of users self-disclosed their respective junior high and senior high schools. The results suggest that many Wretch users are linked by relationships established in high school. The lower graph in Fig. 7 shows that the average hop count required for inference decreases by education level, which implies that users at higher education levels have stronger connections with their schoolmates online.

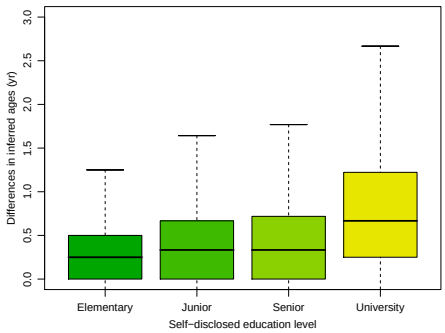

Fig. 8. The inferred age differences between pairs of self-disclosed schoolmates in the four education levels.

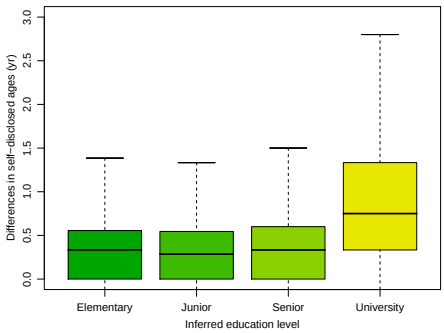

Fig. 9. The self-disclosed age differences between pairs of inferred schoolmates in the four education levels.

5.3 Validation

We apply a cross-validation approach to verify the inferred ages and education records. That is, we verify the inferred ages based on the self-disclosed school information, and verify the inferred school names based on the self-disclosed ages.

To verify if the inferred age information is correct, we find all the users who satisfy the following three criteria: 1) they disclose their current schools, 2) they do not disclose their ages, and 3) we inferred their ages by the above methodology. Intuitively, if two people currently attend the same school, their ages should be similar. Thus, we computed the age differences between pairs of schoolmates, and identified a total of 208,086 valid pairs. We summarize the age differences between pairs of users at different education levels in Fig. 8.

We verify the correctness of the inferred education records based on the same intuition. However, the users being compared should satisfy the following criteria: 1) they disclose their ages, 2) they do not disclose their current schools, and 3) we inferred their education records. In this case, 42,896 pairs of users who described their relationship as "schoolmates" were identified. The distributions of the age difference between each pair of users are shown in Fig. 9. Both plots show that the average age differences between pairs of users in the four education levels are generally less than 2 years. This result suggests that our inference results for users' ages and education records are accurate. It also confirms our conjecture that significant involuntary information leakage occurs in real-world social network services.

6 Discussion

In this section, we discuss the problems that may be caused by identity leakage, such as personalized email spams and spear phishing, and suggest several ways to mitigate the involuntary identity leakage problem.

6.1 Threats Caused by Name Leakage

Spamming. Spam mail has become a major problem in recent years. It is actually a business activity whereby spammers send emails with product information based on massive email address lists. In effect, anyone who has an email address is regarded as a potential customer [15].

To protect users from spamming, academia and industry have developed a number of anti-spam mechanisms. One method of protection against spam involves using a white list [16–18], so that emails from trusted parties will not be mistaken for spam. However, as spamming is profitable, spammers are always devising new ways to penetrate spam filters. To combat the white list approach, spammers collect email addresses and information about social relationships from social network services [15]. In our data set, 46% of users disclose a well-formed email address that spammers may use to send spam mail as if it has been sent by one of the target's friends. In this way, spam mail can bypass the filter and be delivered to the target's mailbox.

Spammers may make use of inferred real names in two ways. 1) They may use the recipient's real name in the mail's content. 2) They may make spam mails look like they have been sent by a friend of the target by using the friend's real name. Consider an email containing the recipient's full real name, where the sender is specified as a friend with his/her correct email address and full real name. The user may have difficulty verifying the email's authenticity.

Phishing. Similar difficulties also exist with respect to phishing detection. Phishers normally send emails, which contain a link to a forged web page, to obtain people's sensitive information, such as an account ID, social security number or credit card number [19]. Some phishing-targeted companies, including eBay and PayPal, and Internet security vendors provide guidelines for recognizing phishing emails. Common rules include "*checking if the email includes your real name because phishers do not have personal information*" [20, 21]. However, the assumption that phishers do not have personal information might be incorrect as self-disclosure is becoming more frequent in social network services. The involuntary name leakage problem will exacerbate the problem further, as it will be more difficult for users and phishing detection mechanisms [19, 22, 23] to verify the authenticity of a web page.

6.2 Potential Solutions

We consider three possible ways to mitigate the problem of involuntary name leakage in social networking services.

A. Personal Privacy Settings. Social network service providers should provide the following preference settings for every account, no matter whether it is free or not:

1. options to hide personal information;
2. options to hide social connections (the level of connections to be hidden should be configurable, e.g., direct friends or friends of friends);

3. options to prevent a user's friends annotating the user's profile with certain words, or deny any incoming annotation completely.
4. options to deny specific people access to a user's incoming and/or outgoing annotations.

B. Browsing Scope Settings. It is recommended that social network service providers limit the profile browsing scope of users. For example, a user could be restricted to browsing friend annotations made by users who are at most two degrees away. One common way to limit browsing scope is through group partitioning, i.e., only users belonging to a certain group, where the action of joining the group requires the approval of a moderator, can access more sensitive information about users in the same group. Such mechanisms could prevent malicious parties from downloading users' personal information, social connections, and annotations in an automated way. We believe that this is the key to solving the problem of large-scale information leakage in social network services.

C. Owner's Confirmation. Every operation involving user information should be confirmed by the information owner. For example, friend annotations should only be shown if the annotatee agrees. This would at least prevent unintentional personal information leakage by the user's friends.

7 Conclusion

In this paper, we consider the involuntary information leakage problem in social network services. To assess the seriousness of the problem, we conduct a case study of Wretch, the most popular social network service in Taiwan. Because Wretch users are allowed to annotate their friends' profiles with a one-line, free-form description, sensitive information, such as a person's real name, can be disclosed without the annotatee's knowledge. We show that 78% of users in our data set were subject to involuntary name leakage. In addition, the ages of 15% of users and the school attendance records (partial or complete) of 42% of users could be easily inferred using a simple heuristic. We also show that the risk of information leakage is not related to the degree of self-disclosure; thus, telling a user not to disclose his/her personal information is not an effective way to reduce the risk that his/her identity could be revealed by online friends.

We discuss three possible ways to mitigate the identified leakage problem, namely, providing personal privacy settings, regulating the browsing scope, and requiring an owner's authorization to release personal information. However, as most users do not change the default settings, we believe that the default mechanisms provided by the services are the most effective methods available. We suggest operators should at least mandate that annotations do not contain any sensitive information (as judged by the annotatee) unless the annotatee agrees.

References

1. ENISA: Enisa position paper no.1, security issues and recommendations for online social networks (October 2007) http://www.enisa.europa.eu/doc/pdf/deliverables/enisa_pp_social_networks.pdf.
2. Gross, R., Acquisti, A., Heinz III, H.: Information revelation and privacy in online social networks. In: Proceedings of the 2005 ACM workshop on Privacy in the electronic society, ACM Press New York, NY, USA (2005) 71–80
3. Ahn, Y., Han, S., Kwak, H., Moon, S., Jeong, H.: Analysis of topological characteristics of huge online social networking services. In: Proceedings of the 16th international conference on World Wide Web, ACM Press New York, NY, USA (2007) 835–844
4. Mislove, A., Marcon, M., Gummadi, K., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, ACM New York, NY, USA (2007) 29–42
5. Kumar, R., Novak, J., Tomkins, A.: Structure and evolution of online social networks. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM Press New York, NY, USA (2006) 611–617
6. OMurchu, I., Breslin, J., Decker, S.: Online social and business networking communities. In: Proceedings of ECAI 2004 Workshop on Application of Semantic Web Technologies to Web Communities. (2004)
7. Boyd, D.: Friendster and publicly articulated social networks. Conference on Human Factors and Computing Systems (CHI 2004), Vienna, Austria, April (2004) 24–29
8. Acquisti, A.: Privacy in electronic commerce and the economics of immediate gratification. In: Proceedings of the 5th ACM conference on Electronic commerce, ACM Press New York, NY, USA (2004) 21–29
9. Jourard, S., Lasakow, P.: Some factors in self-disclosure. Journal of Abnormal and Social Psychology **56**(1) (1958) 91–98
10. Joinson, A.N., Paine (Schofield), C. Oxford Handbook of Internet Psychology. In: Self-Disclosure, Privacy and the Internet. Oxford University Press (2007) 237–252
11. Farmer, R.: Instant messaging–collaborative tool or educator's nightmare. In: The North American Web-based Learning Conference (NAWeb 2003). (2003)
12. Tsai, C.H.: Common chinese names http://technology.chtsai.org/namefreq/.
13. Tsai, C.H.: A list of chinese names http://technology.chtsai.org/namelist/.
14. Tsai, C.H.: A review of chinese word lists accessible on the internet http://technology.chtsai.org/wordlist/.
15. Judge, P., Alperovitch, D., Yang, W.: Understanding and reversing the profit model of spam. In: Workshop on Economics of Information Security 2005. (WEIS 2005). (June 2005)
16. Oscar, P., VWANI, R.: Personal Email Networks: An Effective Anti-Spam Tool. IEEE Computer **38**(4) (2005) 61–68
17. Seigneur, J., Dimmock, N., Bryce, C., Jensen, C.: Combating spam with TEA (trustworthy email addresses). In: Proceedings of the Second Annual Conference on Privacy, Security and Trust (PST04). 47–58
18. Garcia, F., Hoepman, J., van Nieuwenhuizen, J.: Spam Filter Analysis. In: Proceedings of 19th IFIP International Information Security Conference, WCC2004-SEC, Kluwer Academic Publishers (2004)

19. Zhang, Y., Egelman, S., Cranor, L., Hong, J.: Phinding phish: Evaluating anti-phishing tools. In: Proceedings of the 14th Annual Network and Distributed System Security Symposium (NDSS 2007). (2007)
20. Microsoft.com: Recognize phishing scams and fraudulent e-mails http://www.microsoft.com/athome/security/email/phishing.mspx.
21. PayPal: Phishing guide part 2 https://www.paypal.com/us/cgi-bin/webscr?cmd=xpt/cps/securitycenter/general/RecognizePhishing-outside.
22. Wu, M., Miller, R., Garfinkel, S.: Do security toolbars actually prevent phishing attacks? In: Proceedings of the SIGCHI conference on Human Factors in computing systems, ACM Press New York, NY, USA (2006) 601–610
23. Florêncio, D.A.F., Herley, C.: Analysis and improvement of anti-phishing schemes. In: SEC 2006. (2006) 148–157