Stalking Online: on User Privacy in Social Networks

Yuhao Yang†, Jonathan Lutes†, Fengjun Li†, Bo Luo†, Peng Liu‡
† Department of EECS, The University of Kansas, Lawrence, KS, USA
‡ College of IST, The Pennsylvania State University, University Park, PA, USA
{yhyang, jonlutes, fli, bluo}@ku.edu; pliu@ist.psu.edu

ABSTRACT

With the extreme popularity of Web and online social networks, a large amount of personal information has been made available over the Internet. On the other hand, advances in information retrieval, data mining and knowledge discovery technologies have enabled users to efficiently satisfy their information needs over the Internet or from large-scale data sets. However, such technologies also help the adversaries such as web stalkers to discover private information about their victims from mass data.

In this paper, we study privacy-sensitive information that are accessible from the Web, and how these information could be utilized to discover personal identities. In the proposed scenario, an adversary is assumed to possess a small piece of "seed" information about a targeted user, and conduct extensive and intelligent search to identify the target over both the Web and an information repository collected from the Web. In particular, two types of attackers are modeled, namely *tireless attackers* and *resourceful attackers*. We then analyze detailed attacking mechanisms that could be performed by these attackers, and quantify the threats of both types of attacks to general Web users. With extensive experiments and sophisticated analysis, we show that a large portion of users with online presence are highly identifiable, even when only a small piece of (possibly inaccurate) seed information is known to the attackers.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information systems**]: Security and Protection; K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*

General Terms

Security, Experimentation, Measurement

Keywords

Web, social networks, privacy, attacks

CODASPY'12, February 7–9, 2012, San Antonio, Texas, USA.

1. INTRODUCTION

The Internet has changed the ways we publish, search and consume information. Even with the static web, a huge amount of personal-related content has been made available online. More recently, various types of online social network (OSN) products have been introduced to the Internet, which further promotes the sharing of personal information. In addition to the great commercial success and social impacts of the OSNs, they also brought new challenges to the research community (e.g., [12, 27, 24]). With enormous number of users and tremendous amount of personal information available over various online social networks, it is critical to ensure that user privacy is well preserved. However, although many researchers have been working on extracting information or learning knowledge from online social networks, very little research effort has been put so far into the study of security and privacy issues until very recently [2, 42, 28, 15, 44, 40, 17, 16, 43, 32].

In online social networks, users voluntarily share personal information within the community under some implicit assumptions that: (1)these information is only accessible to the targeted readers; (2) one's true identity cannot be discovered if he/she only provides limited/incomplete profile information (e.g. an email address and a phone number); (3) a small amount of information is not significant and the disclosure will not hurt one's privacy; and (4) it is very difficult, if not impossible, to collect and link pieces of information scattered over various online social networks or data sets, and associate them to one's real identity. Unfortunately, these assumptions are proven to be either false or at least questionable, in both research literatures and news reports. Several types of privacy attacks in social networks have been proposed, such as the structural re-identification attacks [2, 42, 28, 15, 44, 40], the inference attacks [17, 16, 43], the information aggregation attacks [32, 25], and the traditional attribute re-identification attacks [19, 14]. Although different types of attacks and countermeasures have been proposed in recent literature, only a few of them have been well tested on real data. Moreover, most of the attacks and corresponding protection mechanisms are based on the graph topologies of social networks. Privacy attacks that focus on the attributes are not well studied.

Personal information is scattered over various sources, including online social networks and the general Web. We believe a thorough understanding of the nature of how these information are distributed and retrievable is the key to an effectively defense. This paper takes a first step towards studying private information online, especially the online

social networks data. In this paper, we intensively examine the vulnerability of private information in online sources as well as the validity of different types of attribute-based privacy attacks. In particular, we define two types of attackers, *resourceful attacker* and *tireless attacker*, based on their different attack capabilities and strategies. Both types of attackers obtain small amounts of information about their targets, known as seed attributes, from external sources and launch advanced re-identification attacks. The seed information could be non-identifiable attributes, such as names of schools where the target gets degrees. A resourceful attacker is capable of retrieving a large amount of personal information about potential targets from online social network sites and creating his/her own resource database, and re-identifies the target by checking the seed attributes against his/her resource database. On the other hand, a tireless attacker only submits such attributes to search engines, and tirelessly browses and studies the results for clues. We have simulated both types of attacks on our database, with 3 million records collected from an online social network and a phonebook data set, to check their reality and severeness. From the results, we can see that large portions of users with online presence are identifiable even with a small piece of seed information, where the seed information could be inaccurate. Our simulation also shows that it does not require extensive resources or efforts to successfully conduct attributed-based attacks to hurt user's privacy online.

2. RELATED WORKS

In online social networks, users are sometimes either oblivious about their privacy, or concerned but underestimate the privacy risks. Surveys and general discussions on social network privacy and security could be found at [21, 5, 41]. In this paper, we are interested in the nature of user identity and personal information that are voluntarily released to social networks, and how such information could be valuable to attackers. Along this thrust, four types of privacy threats have been discovered in the literature: (1) an individual information item (e.g. an identifiable profile image) may be accessed by adversaries; (2) information items of the same user may be collected from different sources, and aggregated to reveal user privacy; (3) values of hidden information items may be inferred from public information items; and (4) user identities could be recovered from anonymized data sets. We briefly cover them below.

2.1 Private information disclosure

Users give out information to trusted social network community. They also implicitly assume that their information would stay within the community. Unfortunately, this assumption is not always valid. For instance, messages sent to an email-based social network may be archived at a repository and accessible to the open public [11], stalkers may follow people through social networks [9], gadgets and add-ons may access users' profiles [20], code errors reveal user profiles [3], etc. To further understand how people value their secrets and the patterns of information revelation, researches on user behavior study, user education, or policy/legal issues have been proposed [19, 14, 4]. For instance, [19] shows that people value their privacy based on context – i.e., the desirability of the traits in a target group. In this sense, people may be willing to publicize private information if they feel they are "somewhat typical or positively atypical compared to the target group" [19]. Meanwhile, a study of the Facebook users within the CMU student community shows that, about 80% of the users adopt identifiable or semi-identifiable images in their profile, and less than 2% of the users made use of the privacy settings [14]. In [29, 30], authors proposed a framework that assesses potential privacy risks to a privacy score, which is computed from the sensitivity of the disclosed information and the visibility of such information. Finally, [23] shows that online social networks and applications leak users' personally identifiable information to third parties. In a position paper [37], the author identifies an attack that uses Sybil nodes and search functions to discover hidden social relationships in LinkedIn.

2.2 Information aggregation

When people participate in online social networks, they voluntarily release different types of personal information: name, screen name, telephone numbers, email addresses, locations, etc. Moreover, when users post messages in forums, blogs, and bulletin boards, they also disclose small pieces of private information. However, with the development of information retrieval techniques, adversaries could collect pieces of such personal information of the targeted user [32]. Though a single piece of such information may be harmless, it discloses a significant amount of private information when associated with other pieces of information. Moreover, adversaries could use evidences such as identical email addresses, screen names, similar posts, and attribute/structural re-identification attacks to bridge profiles across different social networks [32, 25]. Particularly, [38, 13, 25] have shown that people are highly identifiable with very little information, which make cross-network aggregations quite feasible. In all cases of information aggregation attacks, private information of the same user from multiple resources is aggregated and severely hurts user privacy.

2.3 Inference attacks

Aside from voluntary disclosure of explicit personal information, [17, 16, 43] study a type of indirect private information inference through social relations. [17, 16] notice that hidden attributes could be inferred from friends' attributes using a Bayesian network. They study the factors that impact inference accuracy, and suggest that selectively hiding social connections or friends' attributes could help preserve privacy. More recently, [43] also focuses on social networks with mixed public and private user profiles. They found that both friendship links and group membership information could be used to infer sensitive hidden attributes. For instance, membership of a local engineer society discloses location information of the user.

2.4 Privacy Threats in Published Social Network Data

When social network data sets are published for various legitimate reasons, user identity and some profile information are often removed to protect the user privacy. Some of the well-known techniques for this purpose includes *k-anonymity* [39, 1], *l-diversity* [33] and *t-closeness* [26]. For instance, in a k-anonymized data set, an individual cannot be distinguished by attributes from other k-1 records. Possibilities of attribute re-identification attacks on publicly available data sets have been studied in [38, 13, 25]. More recently, in [7]

authors introduce an attribute-based anonymization method for social network data.

On the other hand, due to the nature of social network data, just anonymizing node attributes is not enough. Graph structure contains significant amount of information which could be utilized to hurt user privacy, i.e. structural re-identification attacks. A good survey on structural anonymization and re-identification attacks could be found at [45]. Notably, [2] first identified the problem that although quasi-identifiers are removed before publishing, node identities could be inferred through graph structure. They show that node identities are vulnerable to both passive and active attacks. In [35], authors introduce a new metric, namely *topological anonymity*, to quantify the level of anonymity using the topology properties of network graph. [44] introduces neighborhood attacks, in which an adversary knows the neighborhood subgraph of the target, and tries to re-identify the user from an anonymized network graph. They propose an approach to further anonymize vertexes by modifying edges to construct isomorphic neighborhoods. In [28], authors define *k-degree anonymity*: in a k-degree anonymized graph, each node has the same degree with at least k other nodes. They also efficiently propose k-degree anonymize graphs with minimal edge additions and deletions. Moreover, [15] models three types of adversary knowledge that could be used to re-identify vertexes from an anonymized social network graph. They tackle the problem through graph generalization – dividing the graph into partitions and publishing summarized partition-level data. K-Automorphism is introduced in [46] to defend against multiple attacks. In [18], authors propose a graph anonymization approach that maximally preserves original graph structure and statistical features. Finally, [31] considers social network as a weighted graph, in which edge labels are also considered to be sensitive. They propose to protect sensitive edge labels while keep certain global features of the graph.

3. INFORMATION, VULNERABILITIES AND ATTACKS

3.1 Information and Vulnerabilities

With the Internet explosion, huge amounts of information have been made online. Moreover, advances in information retrieval techniques and Web search engines have enabled easy access to such information. However, large amount of personal information is also exposed to public, not always with the consent of the information owner. In particular, we believe there are three primary channels for personal information disclosure:

Personal information on the general web. In the Web 1.0 era, especially in the early days, personal homepages sometimes contain large amount of personal information. Such information is usually published by owners who are somewhat familiar with the Web. They usually understand the risks better than the novices, hence, the contents may be carefully tailored to protect privacy. On the other hand, some personal information maybe published in sources such as news, employee directories, etc. Overall, this channel is better administered although sensitive information could be disclosed by careless users.

Digitalized public records. With governmental and industrial efforts, a large amount of public records (e.g. phone books) have been digitalized and made available online. Many of them are indexed by commercial search engines, while others require a minimum subscription fee for full access – the barrier is usually low for an adversary to query or even collect the entire databases. Some public information could be highly personal (e.g. salaries of faculty members in public universities).

Online social networks. As online social networks get extremely popular, they become gold mine for adversaries. Large volume of personal information have been collected at social network sites for socialization, career development, and other purposes. As shown in [14], most social network users are poorly protected and their personal information is highly accessible. In this way, social network users may be very vulnerable.

All types of information summarized above are accessible to adversaries, who strive to collect personal information about the targeted users. From the adversaries' perspective, user information could be categorized as (i) private information, (ii) identifiable information, and (iii) non-identifiable information. In the literature, a lot of work has been done on the risks associated with (i) and (ii), and on preventing (i) private information from been disclosed to the Internet. However, seed information obtained by the adversary (from offline) is not always identifiable, hence, the attacker's first objective is to discover the true identity of the target (i.e. from category iii to ii).

3.2 Attacker Models

In this work, we define and simulate two types of attackers, *resourceful attackers* and *tireless attackers*, with different attacking capabilities and strategies.

Resourceful attacker: a *resourceful attacker* is assumed to have enough resource (bandwidth, storage, technique, etc) to construct his/her own database by collecting information from the Web. The database could be constructed in three ways: (1) crawling the general web, extracting personal information from web pages, and storing the data in a local database; (2) implementing a focused crawler to collect data from online public record datasets; (3) crawling online social networks, or downloading research data sets published by social network sites.

In the information retrieval community, many work has been done for entity extraction from the "surface Web", e.g. [6]. However, to populate a local database requires intensive crawling of a significant portion of the surface web, which is very time-consuming. Comparably, collecting information from public records and online social network user profiles is more feasible since the information has been concentrated on such websites. Moreover, considering the user data are usually published in well-structured templates, resourceful attackers can easily implement niche parsers to extract structured personal information. One practical obstacle could be the restriction for massive crawling, which usually violates the terms of use for most online social networks. However, with technical assists, e.g. anonymous routing [8], such crawling is very doable and hard to detect. As such, it is reasonable for us to assume a resourceful attacker has certain technical capability to crawl from typical online data sources.

With the collected databases, resourceful attackers compare his/her external knowledge about the target with information in the database, and search for candidate records for further examination. Meanwhile, if the target is identified from one database, it becomes trivial to use the discovered identity to retrieve more information from other databases. A real-world example for cross-database attacks is given in [25]. The example in [25] is a manually executed attack, but the risk is valid when a resourceful attacker possesses multiple overlapping databases.

Tireless attacker: a *tireless attacker* does not have the resources or techniques to create and maintain a local database. As a compensation, a tireless attacker devotes more time and labor in the attacking process to maximize the chance of success. In particular, a tireless attacker knows some of the attributes of his/her target (seed attributes), and submits such attributes to search engines, and tirelessly browsing and examining the results for clues. Due to the size of the Web, the results returned from search engines are mostly noise, and the attacker needs to be very patient to discover any useful information. The chance of success highly depends on the amount of information provided by the seed attributes. For instance, if the attacker knows that the target get a Bachelor's degree from a large public university and nothing else, it is very unlikely to identify the target in tireless attack. However, if the attacker knows the first name of the target, and the fact that he/she gets a Ph.D. from a small university, the attacker is more likely to discover the true identity (full name) and more personal information of the target.

Meanwhile, a tireless attacker also tries to search or browse in social networks, public records, etc. Furthermore, besides the "brute-force" attack, a tireless attacker can get "smarter" by constructing advanced queries with his/her knowledge about the attacker. For instance, if the attacker knows that the target is currently employed at a university, it is more likely that the target's information will be discovered from webpages within the domain of the university. This type of "advanced search" functions are provided by all major search engines.

4. RESOURCEFUL ATTACKERS

In this section, we focus on two types of privacy attacks (i.e. the re-identification attack and the cross-database aggregation attack) conducted by a resourceful attacker, who is capable of maintaining a private database of a large volume of publicly available online user profiles. In our study, we simulate the power of the resourceful attackers by crawling user data from two publicly available resources, a social networking site and an online phone book data repository, and study the feasibility, difficulty, and the success rate of resourceful attacks with different types of seed attributes.

4.1 Data Collection

To implement a proof-of-concept attacking mechanism, we design niche crawlers to collect data from two resources to simulate the proposed resourceful attacker.

4.1.1 Collecting data from LinkedIn

LinkedIn[1] is a professional online social networking site that provides open access to detailed identifying user profiles. We implemented a specialized crawler to retrieve data based on the public index of LinkedIn.com, using methods and technologies that are available to any potential resourceful attacker. We collected approximately 9 million (8,943,014) user profiles in total in 10 months. The crawled html profiles are indexed alphabetically by the last name of profile owners and stored in a MySQL database for further offline processing, which includes two major procedures, *data extraction* and *data cleaning*.

[1]http://www.linkedin.com

Table 1: Seed attributes in the resource database created by a resourceful attacker.

Name		Work		Education	
FN	first name	*TI*	title	*S*	school name
LN	last name	*AF*	affiliation	*D*	degree
		IND	industry	*ST*	start time
		LO	location	*ET*	end time
				T	time period

Table 2: Approximate information on attributes.

Attribute	Approximation	Notation
name	initials	*N.in*
school	state	*S.st*
	region	*S.re*
	country	*S.ct*
	continent	*S.cn*

Data extraction: The LinkedIn profile contains rich information about one's educational history that is useful in identifying a target. However, the raw data are in html profiles, which need to be extracted to reconstruct corresponding records in the resource database. We implemented a specialized parser to do that. Currently, our parser only extracts data from three fields *name*, *work* and *education* fields, which contain the most useful information for re-identification. In the future, we consider to extend our parser to include more fields such as working experiences. Data from the three fields are further processed and categorized into 11 **seed attributes**, as shown in Table 1. For instance, data in *name* field are segmented as first, middle and last names. Data in *work* field are decomposed to *current title*, *affiliation* (e.g. Software Engineer at XYZ company), *industry* type (e.g. Internet, Higher Education, Research), and current *location* (e.g. San Francisco Bay Area). Similarly, *school name*, *degree* earned, *major*, degree *starting time* and *ending time*, and the entire degree *time period* are extracted from *education* field. Please note that one profile may have multiple education records. Also, not all the profiles contain education information. In some profiles, the education field is either left blank or hidden from non-registered LinkedIn users.

To simulate the attacks where the attacker only had approximate information about the target, we consider two common scenarios. In the first case, the attacker only knows the initial of the target, and in the second case, the attacker only know the approximate location of the school that the target has attended. Therefore, we added four new attributes to our seed attribute table, as shown in Table 2. After cross-checking with the school reference lists, we have successfully added country and continent information to all schools and state information to all the US schools, for 80% of the records.

Data cleaning: The collected data may have redundant or ambiguous contents, which makes data cleaning operations important in data collection. Some of the ambiguity is caused by inaccurate or wrong inputs of careless users, for instance, a user mistakenly includes department name or year of graduation as part of school name. A more common problem resulting in redundant content is that many universities are referred by different names. For example, we noticed University of Cambridge is referred as Cambridge University instead of its formal name in some profiles. We corrected this problem by cross-checking with the school reference lists that contain formal forms for most of the schools. After a quick browsing of the data, we created manually coded heuristic rules to map most of the school names to their formal forms, and removed all the redundant elements and special characters [2].

Another important processing we took in data cleaning is to set aside the schools with less than 3 attendees, which we think are highly likely to be invalid or mistaken entries (proved by later manual check). After all the operations, we successfully obtained about 2,466,721 clean profiles, with 3,417,550 clean education records.

4.1.2 Collecting data from online phone books

Many data sets with private personal information are now publicly available for commercial or administration purposes. Such information is open to public, unless data owners explicitly opt out. Residential phone book data is such a resource, which has been made online through various sources. All the online phone book sites list phone numbers and residential locations for free access. For the registered users, more detailed residential information are also available. Moreover, a few of online phone books even show the names and addresses of the holders as unlisted phone book entries, for instance, while the phone numbers are hidden, the owners' info is displayed on http://www.phonesbook.com.

We assume the resourceful attackers are capable of retrieving all types of online data to enrich their resource databases. Therefore, we crawled residential phone book entries for three regions, two college towns and one state capital city[3], from an online phone book data repository to simulate attackers' knowledge in this category.

After creating his own resource database, the resourceful attacker is capable of launching two types of attacks, the *re-identification attack* and the *cross-network attack*.

4.2 Re-identification attacks

The re-identification attack is to explore the identity (and/or other information-of-interest) of the target by linking or matching the known information about the target to the data in the resource database. In this section, we first simulate a number of re-identification attacks over the crawled LinkedIn data to assess the risk of re-identification attack against profile data that users voluntarily submitted to online social networks. Then, we employ an information theory based approach to theoretically estimate the re-identification risk.

[2]Due to lack of referencing lists for high schools and lower level schools, we have to remove all education records at high school level or lower.

[3]City names are anonymized as required by double-blind review policy.

4.2.1 Re-identification attack model

To launch a re-identification attack, the attacker needs to know some information about the target. It is assumed that the attacker obtains such knowledge from external resources. When the attacker obtains offline information about the target, he expresses this knowledge in the form of *seed attributes* that he collects for the resource database.

The attacker's knowledge about each target varies. In some cases, the attacker knows only one seed attribute about the target, e.g. "John has a bachelor's degree". In other cases, the attacker may know more about the target, which can be interpreted as multiple seed attributes, e.g. "John graduated from college in 2004". Sometimes, the knowledge about the target is not accurate. For example, the attacker may only know that "John graduated from a school in Midwest". Since the inaccuracy in the name and school location fields are addressed by the new approximate attributes in Table 2, we can simulate certain inaccurate inputs in attacker's knowledge. For instance, the attacker may know that: "John graduated from a school in Midwest", which indicates Attribute $SchoolRegion = "Midwest"$.

Therefore, we model the attacker's knowledge about a target as an identity-attribute tuple $<I, v_1,..., v_t>$, where I is the identity of the target, and $\{v_1,...,v_t\}$ are the values of the known seed attributes $\{A_1,...,A_t\}$. For instance, the attacker's knowledge "John graduated from college in 2004" can be expressed as:

$$\begin{aligned} \text{Identity}: &\quad I = John \\ \text{Attribute } FirstName: &\quad v_1 = "John" \\ \text{Attribute } EndTime: &\quad v_2 = "2004" \end{aligned}$$

In the defined resourceful attack model, to re-identify the target, the attacker needs to send the known identity-attribute tuple into the resource database that is built upon the data retrieved from online sources. The severeness of such re-identification attack highly depends on the completeness and identifiability of the records in the resource database. Therefore, the first-step approach towards assessing the risk of such re-identification attack is to study the resource database. In particular, we explore the *identifiability* of the crawled LinkedIn user profiles in our simulated resource database to assess the re-identification risk.

4.2.2 Assessing risk with profile identifiability

The resource database and the seed information are two key components for a successful re-identification. Consider a resource database $\mathbb{D}$ with n records, where each record is associated with one identity. To an attacker, the ideal case for the resource database is that it is large enough to contain records of all the targets and each record contains all information about the target. The ideal case for the seed information is that it is accurate and adequate to distinguish the target from records of others in the database. However, it is very difficult, if not impossible, to meet both conditions in real-world cases. Therefore, for a resourceful attacker, it is important to measure the identifiability of the records in the resource database $\mathbb{D}$.

Definition 1. For a database $\mathbb{D}$ whose scheme is $\mathbb{D}(A_1, ..., A_t)$, we define the **identifiability of a target T in $\mathbb{D}$** as $I_T^{\{v_1,...,v_r\}} = k$, if T cannot be distinguished from other $k-1$ profiles with known seed attributes $\{attr_1, ..., attr_r\} = \{v_1, ..., v_r\}$, where $r \leq t$.

This definition is similar to the *k-anonymity* concept of privacy in data publishing, but interpreted from the attacker's perspective. For each target whose record is in $\mathbb{D}$, given any adequate and accurate seed information $\{v_1, v_2, ..., v_t\}$, his/her identifiability should be 1, which means he/she is uniquely identified. Typically, since the attacker's seed information is limited, the identifiability of a target, k, is much larger than 1. However, for the attacker, the size of potential profile set (that may contain the target) under this definition is successfully decreased from n to k.

To assess the **identifiability of** $\mathbb{D}$, we further count n_{k-}, which is the number of profiles that cannot be identified from at most k other profiles given seed information $\{attr_1, ..., attr_r\}$. In other words, for every possible value set $\{v_1, ..., v_r\}$ in the seed attribute tuple space $\mathbb{R}^r$,

$$n_{k-} = sum(I_T^{\{v_1,...,v_r\}}), \quad for \quad I_T^{\{v_1,...,v_r\}} < k.$$

Then, we calculate k-or-less proportion $p(k)$ as an indicator of the identifiability of $\mathbb{D}$, where

$$p(k) = \frac{n_{k-}}{n}, \; for \quad k \in [1, max(k)],$$

and $max(k)$ is the largest k for all possible values of seed attribute tuple $\{attr_1, ..., attr_r\}$.

Next, we select several seed attribute tuples, and assess the identifiability of the resource database with crawled LinkedIn data. First, we simulate the scenario where the attacker only knows a single seed attribute value about the target. Then we measured the k-or-less proportion $p(k)$ for each seed attribute in Table 1. The results of three seed attributes, first name FN, work location LO, and school name S, are shown in Figure 1(a). In the figure, a slowly growing curve indicates better anonymity, since less people are identifiable among smaller sets. As we can see from the figure, users' identifiability shows different patterns for different attribute. Overall, when the adversary only knows one attribute, most people cannot be identified among a relatively large set.

Then, we consider the scenario in which a weaker attacker only knows approximate values of the attributes, as summarized in Table 2. Some of the results are shown in Figure 1(b), when the attacker knows (i) the first and last initials ($N.in$) of the target, but not the name, e.g. the attacker knows "JD", not "John Doe", or (ii) the region where the target goes to school ($S.re$), e.g., "the person went to school in West coast". As we have expected, knowing approximate values on an attribute usually gives the adversary very limited information.

The third type of scenarios that we examine is that the resourceful attacker knows multiple attributes about the target. Figure 1(c) shows the population vs. k-anonymity curves when the resourceful attacker knows (i) first name and affiliation: <*FN, AF*>, e.g. "John works at XYZ company"; (ii) school name and starting time: <*S, ST*>, e.g. "the person went to Stanford in 2001"; and (iii) first name, work location and school name <*FN, LO, S*>, e.g. "John went to Berkeley, and now works at New York". Note that the k axis is scaled to [1, 100]. As we can see from the figure, users become very vulnerable when the adversary knows multiple seed attributes.

We also consider the case where the adversary knows approximate information on multiple attributes. The rightmost figure in Figure 2(b) shows the population vs. k-anonymity curve when the attacker knows the states in which the target goes to school (given that the target goes to at least two schools), e.g. "the person went to school in California and Massachusetts". Obviously, the database is less identifiable under approximate seed information.

Table 3: Information gain (IG) by knowing a single seed attribute with precise values.

Category	Attribute	IG (bit)
Name	*FN*	13.348
	LN	16.461
Work	*TI*	14.433
	AF	12.979
	IND	6.405
	LO	8.011
Education	*S*	11.8231
	D	1.8336
	ST	5.149
	ET	5.026
	T	7.537

4.2.3 Assessing risk using information gain

To quantify the *amount of information* provided by an attribute, we further analyze the problem from a information theory perspective. In our scenario, the goal of the attacker is to identify the particular record which corresponds to the target. Without any prior knowledge, all the records are equally likely to be the target. Hence, to achieve the goal, the average amount of information that the attacker needs to collect (i.e. adversary's expected information gain) is denoted as:

$$E(I(X)) = H(X) = -\log_2 \frac{1}{N}$$

where N is the number of records in the database. In our simulation, $E(I(X)) = 21.23$(bits), i.e. on average, the attacker needs to obtain *21.23 bits* of information in order to identify a target from our database.

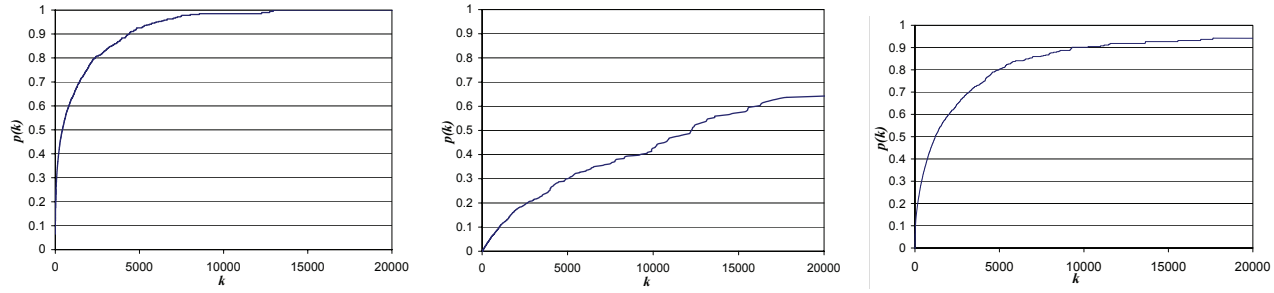
When the attacker knows the value v of attribute $attr$, the conditional entropy is denoted as:

$$H(X|attr = v) = -\log_2 \frac{1}{N_{attr=v}}$$

where $N_{attr=v}$ is the number of records that satisfy the condition `attr=v`. On average, the information gain of knowing attribute A is denoted as:

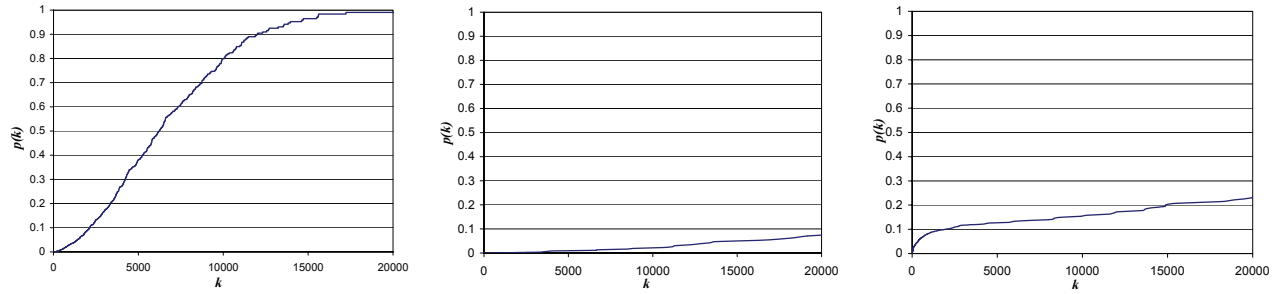
$$\begin{aligned} I(X;A) &= H(X) - H(X|A) \\ &= H(X) - \sum_{v \in V_A} p(A=v) H(X|A=v) \end{aligned}$$

where $H(X|A)$ is the conditional entropy of knowing attribute A. In our settings, an information gain of m bits indicates that the attacker has successfully discovered that the target is among $\frac{N}{2^m} = \frac{2,466,721}{2^m}$ records, on average. In additions, the attacker will need to further obtain $21.23 - m$ bits of information in order to exactly identify the target. Most importantly, if we assume that our data set is a random sample of the general population, attackers' information gain will be the same if he obtains the same attribute in the general population. In that case, $H(X)$ and $H(X|A)$ increases proportionally, while $I(X;A)$ will remain the same (statistically).



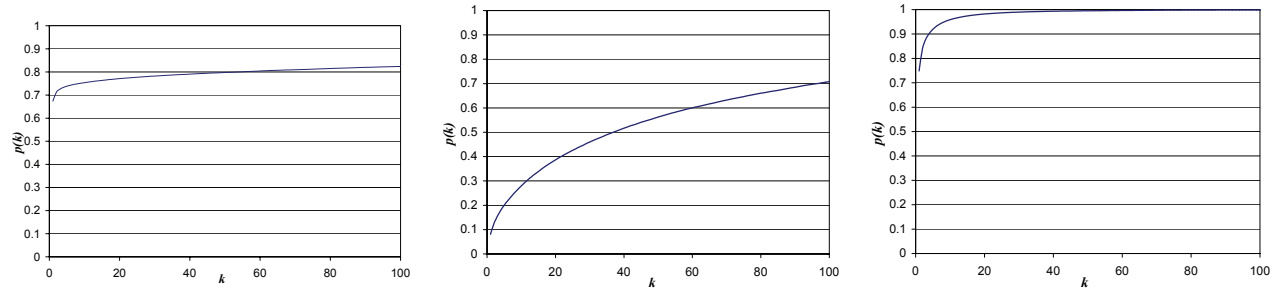
seed attribute: first name FN seed attribute: work location LO seed attribute: school S

(a) Measured k-or-less proportion $p(k)$ using seed attribute tuple with precise values.



seed attribute: initial $N.in$ seed attribute: state $S.st$ seed attribute tuple: $<S.st, S.st>$

(b) Measured k-or-less proportion $p(k)$ using seed attribute with approximate values.



seed attribute tuple: $<FN, AF>$ seed attribute tuple: $<S, ST>$ seed attribute tuple: $<FN, LO, S>$

(c) Measured k-or-less proportion $p(k)$ using seed attribute tuple with precise values.

Figure 1: Estimate the risk of *resourceful attacks*.

In Table 3, we show the information gain of the attacker when he/she knows one seed attribute. As we expected, the last name carries the largest amount of information, while first name and school name also carries significant amount of information. However, knowing one attribute alone is not enough for the attacker to identify the target, or to narrow down to a very manageable range. Attribute ln is somehow an exception, which on average narrows the search to less than 30 candidates (i.e. $H(X|A) = 4.77bit$). When the attacker only know approximate information on an attribute, the information he/she learns from the knowledge is even less, as shown in Table 4.

In the scenario that the attacker knows multiple attributes, the information gain is denoted as:

$$I(X; A_1A_2) = H(X) - H(X|A_1A_2)$$

When two attributes A_1 and $A_2$are independent, we should have:

$$\begin{aligned} I(X; A_1A_2) &= H(X) - H(X|A_1A_2) \\ &= H(X) - H(X|A_1) - H(X|A_2) \end{aligned}$$

Table 4: Information gain (IG) by knowing seed attribute with approximate values.

Attribute	IG (bit)	Attribute	IG (bit)
$N.in$	8.807	$S.st$	4.795
$S.re$	2.360	$S.ct$	1.853
$S.cn$	1.328		

Table 5 shows the information gain when the attacker knows multiple attributes.

4.3 Cross-database aggregation

As we have introduced, a resourceful attacker is capable of collecting multiple databases from different sources. When

Table 5: Information gain (IG) by knowing multiple seed attributes.

Attributes	IG (bit)	Attributes	IG (bit)
<*FN, S*>	20.316	<*S, ST, ET*>	16.549
<*FN, S.st*>	15.068	<*FN, S.ct*>	12.848
<*FN, ST*>	16.092	<*FN, ST, ET*>	17.362
<*FN, ET*>	15.679	<*D, ST, ET*>	5.685

the attacker identifies the target (i.e. discovers the full name of the target) from one of the databases, it becomes trivial to retrieve relevant records from other databases to learn more about the target.

In our experiments, we simulate cross-database aggregation attack by matching LinkedIn data with online phone book data. We have crawled phone book data for three cities: two college towns and a state capital city. We try to link records from both databases by matching full names. The results are shown in Figure 2. As we can see, approximately 20% of the LinkedIn users from town A could be identified in phone book, while 14% and 14% of the LinkedIn users from town B and town C are re-identified, respectively. According to the literature [25, 13], with known full name and location information, people are very identifiable. We are confident that most of the linked records are true positives (i.e., the two linked records reflect one unique offline identity). For linked records, the attacker will further learn the home address and phone number of the user. In many cases, the attacker also learns the names of the family members of the user.

In cross-database aggregation attacks, when a resourceful attacker identifies a target using attribute-reidentification attacks on one of his databases, it is likely that he can learn more information about the target. In our experiments, we only collected information about users whose phone numbers are listed. As we have mentioned, there are websites (e.g. http://www.phonesbook.com/) that publish addresses of users who opt to exclude their information from the phone book. From our observation, this websites contains 20% more user records than the phone book data set we crawled. Meanwhile, with a small fee, the attacker could subscribe to various databases that collect personal information from public and commercial records. Therefore, a resourceful attacker has great potential to become more powerful than we have demonstrated in this work.

We can also see that the phonebook size is much larger in state capital C, which shows that a relatively larger population who do not have LinkedIn accounts (or configured their accounts as private), but are still visible in the phone book. In this case, although these users are not actively releasing their information online, or are successfully protecting their online identities, unfortunately, their personal information is still accessible from online sources.

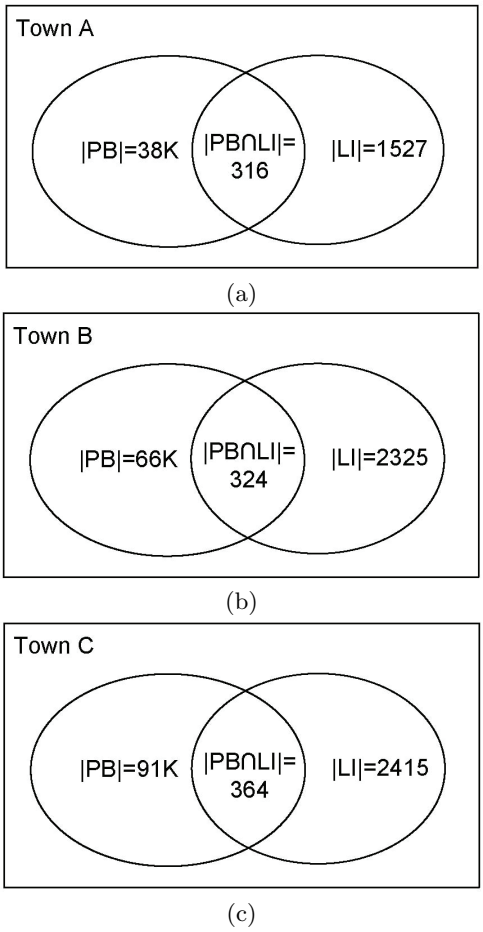


Figure 2: Cross-database aggregation for three cities.

5. TIRELESS ATTACKERS

5.1 Tireless Attackers

Tireless attackers do not possess a local database of personal information, as a compensation, they devote their time and energy. In our simulation, the tireless attacker knows some (non-identifiable) attributes about the target. The attacker queries a Web search engine (we use Google in our experiments) with the known attributes, and examines the results returned by the search engine for any clue.

To simulate tireless attacks, we have randomly sampled 50,000 users from education and healthcare industry, including faculty, students, researchers, doctors, etc. We simulated tireless attacks on different combinations of known attributes. In Figure 3, we show the success rate when the tireless attacker knows the target's: (1) first name and the name of last school that the target attended <*FN, S*>; (2) last name and school <*LN, S*>; (3) first name and current affiliation <*FN, AF*>; (4) last name and current affiliation <*LN, AF*>; (5) names of two schools, knowing that the target has attended two or more schools <*S, S*>; and (6) school name, degree and year of graduation <*S, D, ET*>. When the full name of the target was discovered in a returned web page in the form of "John Doe" or "Doe, John", we treat the result as *positive*. An attack is successful when at least one positive result is found in the top 200 results returned from the search engine. Please note that in tireless attacks, we exclude all the results from LinkedIn, i.e., an attack is successful only if the target is re-identified from non-LinkedIn sources.

Figure 4 and Figure 5 give more insights on tireless attacks. Figure 4(a) shows a histogram of the number of positive results for successful attacks when the attacker knows

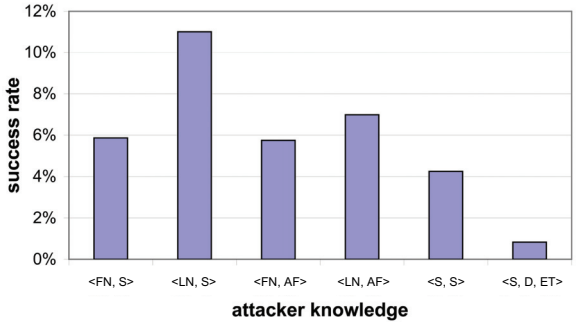


Figure 3: Success rate of tireless attackers.

the first name and affiliation of the target *<FN, AF>*. Figure 5(a) shows the same histogram for *<FN, S>* case. We do observe a significant portion of targets who have been re-identified from multiple websites (excluding LinkedIn). Meanwhile, no victim has been re-identified from more than 20 websites. Figure 4(b) and Figure 5(b) show a histogram of the rank of the first positive result for successful attacks of the *<FN, AF>* and *<FN, S>* cases, respectively. We observe that most of the positive results came from top 10 results, which indicates that a tireless attacker does not need to be very "tireless" to achieve a successful attack. On the other hand, we also observe that positive results do not always come in top 2 search results.

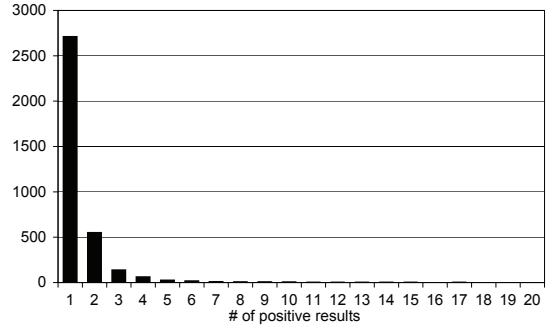
To further validate the successful attacks, we have manually checked 200 randomly-sampled positive results for each type of attacks. We have discovered that around 70% of them were true positives that also contain further personal information about the target. Meanwhile, we do have some false negatives. For instance, in the *<LN, AF>* attack, we have found a few pages of conference program committee members. They contain the name of the school, and a person with exactly the same name as the target, but affiliated with a different school or organization. Another major category of false positive appears when the name "Doe, John" is discovered in the context of "Jay *Doe, John* Smith".

Last but not least, when the targets are identified in tireless attacks, we continue the attack by issuing new queries using their identity (i.e. full name) and known attributes. For most of the cases, we can easily discover more sources (again, excluding LinkedIn) that contains further information about the target. A major reason is that we use the LinkedIn user profiles from education and healthcare domains as seeds, and such users are more active on the Internet.

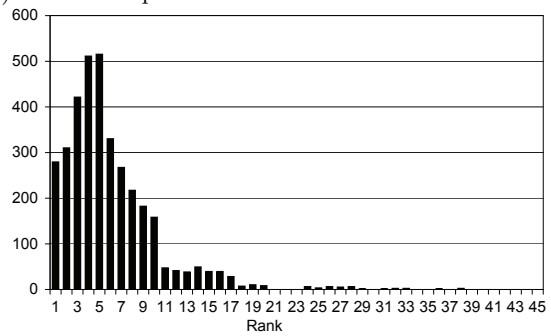
5.2 Smart Tireless Attackers

Regular tireless attackers use a simple textual combination of all the known attributes as the query to be sent to search engines. However, existing web search engines support not only free text queries, but also advanced queries (e.g. Google Advanced Search[4]). Tireless attackers can get smarter by utilizing such functions. In our simulation, when the tireless attacker knows the affiliation of the target (e.g. this person works at XYZ University), it is highly likely that information about the target could be found in the employers' domain (e.g. xyz.edu). A smart attacker first queries

[4]http://www.google.com/advanced_search



(a) Number of positive results in each successful attack.

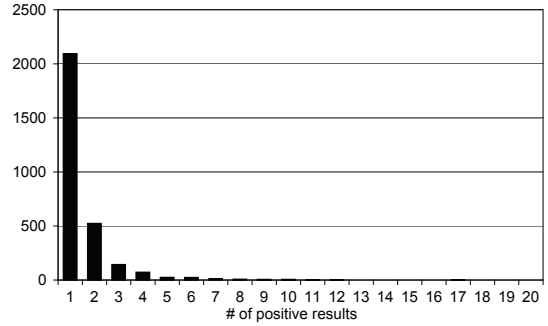


(b) Rank of the first positive result in each successful attack.

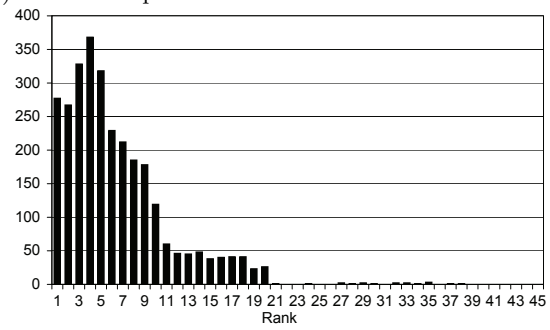
Figure 4: Results of successful *tireless attacks* with seed attribute tuple *<FN, AF>*.

the search engine (e.g. Q="XYZ university") to get the official website of the employer, which is usually included in the top 3 returned results. The attacker then issues an advanced query, which contains textual terms and a domain constrain. The textual terms include the other known attributes about the target (e.g. first name "John"), while the domain constrain forces to search within the employer's domain (e.g. "site:xyz.edu").

We simulate smart tireless attacks for the case with seed attribute tuple *<FN, AF>* (i.e., the attacker knows the first name and current affiliation of the target). We have simulated attacks for 10,000 users, randomly sampled from the 50,000 records that we used for regular tireless attacks. Figure 6 shows the simulation results of such smart tireless attacks. As we can see, the re-identification rate of smart tireless attacks is lower than the re-identification rate of regular tireless attacks. It means that, in at least 50% of the successful regular tireless attacks, the targets are identified from information sources other than websites of their workplaces. When we further look into the successful smart tireless attacks, we observe that most of them are true positives. Moreover, as we can see from Figure 6(b), on average, the rank of the first positive result is higher in smart tireless attacks. Therefore, smart attacks are more effective – less effort is required for the attacker to browse and examine the results. For both regular and smart tireless attacks, we can see that most of the users are either identified in top results, or never identified. It means that when the user is not "highly visible", his/her information is most likely buried in the massive amount of online information and becomes invisible. However, consider the fact that only a small portion of the general population have disclosed their information

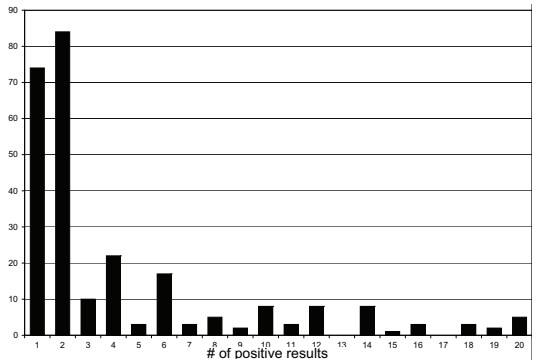


(a) Number of positive results in each successful attack.

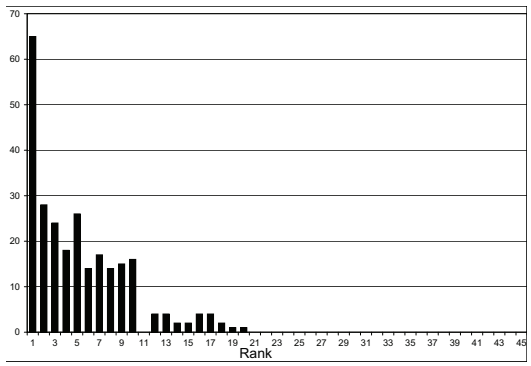


(b) Rank of the first positive result in each successful attack.

Figure 5: Results of successful *tireless attacks* with seed attribute tuple <*FN, S*>.



(a) Number of positive results in each successful attack



(b) Rank of the first positive result in each successful attack

Figure 6: Results of successful *smart tireless attacks* with seed attribute tuple <*FN, AF*>

on the Web, people with an online presence is still highly distinguishable.

6. ANALYSIS AND REFLECTION

Information. We have observed a large amount of personal information available over the Internet. Each information item may include both identifiable and non-identifiable attributes. Not all such information is published by the owner (of the identity), or with the consent of the owner. For instance, we have observed webpages such as news stories published by the employer. Moreover, the user might be completely unaware that his/her information has been accessible and searchable over the Internet. From the simulation results, we can see that it is very difficult, if not impossible, to completely hide one's online identity in the Internet age.

On the other hand, we have introduced an information-theory-based approach to evaluate the values of personal information items to the attackers. We believe that the results will help users determine the types and amounts of information to be published on personal and social networking sites.

Vulnerability. We have simulated the data collection process of resourceful attackers. We can see that personal information could be easily collected by attackers, especially from social networking or public record sites, where information is published in well-structured templates. On the other hand, automatically and accurately extracting *large amounts* of structured information from free (unstructured) text is not an easy task. Named-entity extraction [10, 6] is a very hard problem. Although we have seen successes in controlled datasets or for popular entities that appear on many web contexts, the general problem of arbitrary entity extraction is still far from being solved. In particular, diversity of web documents and limited evidences (e.g. a user's phone number only appears on one webpage) make it very difficult to precisely extract and collect large amounts of entities from the web. However, we have shown that it takes little effort for a human attacker to exploit search engines to locate webpages containing such information.

Next, with the simulation results of resourceful and tireless attackers, we have shown that **people with web presences are highly identifiable**, even with very limited or approximate information. Moreover, information from multiple resources could be linked to provide more information to the attacker. A major reason behind the phenomenon is that many people do not have a web presence, as confirmed by our cross-network aggregation attacks. On the contrary, people with web presence are very likely to appear in multiple sources. In this sense, we have a group of people who are more active on the Web, while the mass majority of the population mostly remain silent online. As a result, the online population becomes very identifiable. There appears to be a dilemma: if we have more people online, the identity of the existing users will be better "shadowed" than they are right now. However, in this way, we may put more people under risk.

Attacks. Recent advances in information retrieval techniques are shown to be a double-bladed sword – they provide great functions to the users, but also reveal their private information to attackers with sufficient capabilities and resources, or strong wills. Intuitively, we can interpret the

goal of the attacker as taking a piece of seed information as input against large data that are available online to successfully find a hit.

Ideally, if the seed is precise and adequate and the data is large enough to guarantee that it contains the target, the attack will always succeed. While the results are constrained in reality, the attacker manages to increase his chance and efficiency by meeting the conditions at his most. The first (and often hidden) assumption is that the focused data should be large enough to contain data of a particular target. In the resourceful attack, the focused data is the resource database created by the attacker, which in turn motivates the long-term and multi-source data collection. The second condition that affects the success rate of the attack is the identifiability of the user with the seed information (in terms of either seed attributes or search terms). In the tireless attack, it is assumed implicitly that the related data should be in the high-rank results returned by search engines. This in turn explains why tireless attack is only effective when the target is highly distinct against proper search terms (or combination of search terms). The study of the identifiability will also shed light on how to tailor one's online presence to shadow his identity within an indistinguishable group.

7. CONCLUSION

A large amount of personal information has been made available over the Internet. The advances in information retrieval technologies provide powerful Web search engines to legitimate users. However, they also provide adversaries with a convenient access to the abundant personal information on the Web.

In this paper, we have studied personal information that is disclosed to the Web through various sources, especially through online social networks. We also analyze the vulnerabilities in information and possible attacks. In particular, we have presented two types of attackers: the resourceful attackers and the tireless attackers. We assume that an attacker possesses a small piece of "seed" information about his target. A resourceful attacker searches local database with data collected from various online sources, including social networks and online public records. On the other hand, a tireless attacker queries web search engines with his seed information, and untiringly examines the results. We have simulated both attacks with real data collected from online social networks and phonebook, and quantitatively analyzed the results. From the results, we can see that large portions of users with online presence are very identifiable, even with a small piece of seed information, and the seed information could be inaccurate. We also show that it does not require extensive resources or efforts to successfully conduct such attacks.

8. ACKNOWLEDGEMENTS

Bo Luo was partially supported by NSF OIA-1028098 and University of Kansas General Research Fund 2301420. Peng Liu was partially supported by AFOSR FA9550-07-1-0527 (MURI), ARO W911NF-09-1-0525 (MURI), NSF CNS-0905131 and NSF CNS-0916469.

9. REFERENCES

[1] G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy, D. Thomas, , and A. Zhu. k-anonymity: Algorithms and hardness. Technical report, Stanford University, 2004.

[2] L. Backstrom, C. Dwork, and J. Kleinberg. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of ACM international conference on World Wide Web*, pages 181–190, 2007.

[3] P. Cashmore. Privacy is dead, and social media hold smoking gun. CNN, October 2009.

[4] J. Caverlee and S. Webb. A large-scale study of myspace: Observations and implications for online social networks. In *Proceedings of the International Conference on Weblogs and Social Media*, 2008.

[5] X. Chen and S. Shi. A literature review of privacy research on social network sites. In *Multimedia Information Networking and Security, 2009. MINES '09. International Conference on*, volume 1, pages 93 –97, nov. 2009.

[6] T. Cheng, X. Yan, and K. C.-C. Chang. Entityrank: searching entities directly and holistically. In *VLDB '07: Proceedings of the 33rd international conference on Very large data bases*, pages 387–398, 2007.

[7] S. Chester and G. Srivastava. Social network privacy for attribute disclosure attacks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, pages 445 –449, july 2011.

[8] R. Dingledine, N. Mathewson, and P. Syverson. Tor: the second-generation onion router. In *USENIX Security Symposium*, 2004.

[9] B. Dubow. Confessions of 'Facebook stalkers'. USA Today, March 2007.

[10] O. Etzioni, M. Cafarella, D. Downey, A.-M. Popescu, T. Shaked, S. Soderland, D. S. Weld, and A. Yates. Unsupervised named-entity extraction from the web: An experimental study. *Artificial Intelligence*, 165(1):91 – 134, 2005.

[11] G. Eysenbach and J. E. Till. Ethical issues in qualitative research on internet communities. *BMJ*, 323:1103–1105, 2001.

[12] L. Garton, C. Haythornthwaite, and B. Wellman. Studying online social networks. *Journal of Computer-Mediated Communication*, 3(1), 1997.

[13] P. Golle. Revisiting the uniqueness of simple demographics in the us population. In *WPES '06: Proceedings of the 5th ACM workshop on Privacy in electronic society*, pages 77–80, New York, NY, USA, 2006. ACM.

[14] R. Gross, A. Acquisti, and I. H. John Heinz. Information revelation and privacy in online social networks (the facebook case). In *Proceedings of ACM workshop on Privacy in the electronic society*, pages 71–80, 2005.

[15] M. Hay, G. Miklau, D. Jensen, D. Towsley, and P. Weis. Resisting structural re-identification in anonymized social networks. *Proc. VLDB Endow.*, 1(1):102–114, 2008.

[16] J. He and W. W. Chu. Protecting private information in online social networks. In *Intelligence and Security Informatics*, pages 249–273, 2008.

[17] J. He, W. W. Chu, and Z. Liu. Inferring privacy information from social networks. In *IEEE*

International Conference on Intelligence and Security Informatics, pages 154–165, 2006.

[18] X. He, J. Vaidya, B. Shafiq, N. Adam, and V. Atluri. Preserving privacy in social networks: A structure-aware approach. In *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT '09. IEEE/WIC/ACM International Joint Conferences on*, volume 1, pages 647 –654, sept. 2009.

[19] B. A. Huberman, E. Adar, and L. R. Fine. Valuating privacy. *IEEE Security and Privacy*, 3(5):22–25, 2005.

[20] M. Irvine. Social network users overlook privacy pitfalls. USA Today, April 2008.

[21] P. Joshi and C.-C. Kuo. Security and privacy in online social networks: A survey. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pages 1 –6, july 2011.

[22] V. Kostakos, J. Venkatanathan, B. Reynolds, N. Sadeh, E. Toch, S. A. Shaikh, and S. Jones. Who's your best friend?: targeted privacy attacks in location-sharing social networks. In *Proceedings of the 13th international conference on Ubiquitous computing*, UbiComp '11, pages 177–186, 2011.

[23] B. Krishnamurthy and C. E. Wills. On the leakage of personally identifiable information via online social networks. In *WOSN '09: Proceedings of the 2nd ACM workshop on Online social networks*, pages 7–12, New York, NY, USA, 2009. ACM.

[24] J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 915–924, 2008.

[25] F. Li, J. Y. Chen, X. Zou, , and P. Liu. New privacy threats in healthcare informatics: When medical records join the web. In *ACM SIGKDD Workshop on Data Mining in Bioinformatics*, 2010.

[26] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *Proceedings of the 23rd International Conference on Data Engineering*, pages 106–115, 2007.

[27] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng. Facetnet: a framework for analyzing communities and their evolutions in dynamic networks. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 685–694, 2008.

[28] K. Liu and E. Terzi. Towards identity anonymization on graphs. In *Proceedings of the 2008 ACM SIGMOD*, pages 93–106, 2008.

[29] K. Liu and E. Terzi. A framework for computing the privacy scores of users in online social networks. In *Data Mining, 2009. ICDM '09. Ninth IEEE International Conference on*, pages 288 –297, dec. 2009.

[30] K. Liu and E. Terzi. A framework for computing the privacy scores of users in online social networks. *ACM Trans. Knowl. Discov. Data*, 5:6:1–6:30, December 2010.

[31] L. Liu, J. Wang, J. Liu, and J. Zhang. Privacy preserving in social networks against sensitive edge disclosure. Technical Report CMIDA-HiPSCCS 006-08, University of Kentucky, 2008.

[32] B. Luo and D. Lee. On protecting private information in social networks: A proposal. In *Workshop on Modeling, Managing, and Mining of Evolving Social Networks - in conjunction with IEEE ICDE*, 2009.

[33] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1):3, 2007.

[34] A. Masoumzadeh and J. Joshi. Preserving structural properties in anonymization of social networks. In *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2010 6th International Conference on*, pages 1 –10, oct. 2010.

[35] L. Singh and J. Zhan. Measuring topological anonymity in social networks. In *GRC '07: Proceedings of the 2007 IEEE International Conference on Granular Computing*, page 770, Washington, DC, USA, 2007. IEEE Computer Society.

[36] A. C. Squicciarini, M. Shehab, and J. Wede. Privacy policies for shared content in social network sites. *The VLDB Journal*, 19:777–796, December 2010.

[37] J. Staddon. Finding "hidden" connections on linkedin an argument for more pragmatic social network privacy. In *Proceedings of the 2nd ACM workshop on Security and artificial intelligence*, AISec '09, pages 11–14, New York, NY, USA, 2009. ACM.

[38] L. Sweeney. Uniqueness of simple demographics in the u.s. population, 2000.

[39] L. Sweeney. k-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570, 2002.

[40] X. Ying and X. Wu. Randomizing social networks: a spectrum preserving approach. In *SIAM International Conference on Data Mining (SDM)*, 2008.

[41] C. Zhang, J. Sun, X. Zhu, and Y. Fang. Privacy and security for online social networks: challenges and opportunities. *Network, IEEE*, 24(4):13 –18, july-august 2010.

[42] E. Zheleva and L. Getoor. Preserving the privacy of sensitive relationships in graph data. In *International Workshop on Privacy, Security, and Trust in KDD (PinKDD)*, pages 153–171, 2008.

[43] E. Zheleva and L. Getoor. To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles. In *18th International World Wide Web conference (WWW)*, April 2009. Earlier version appears as CS-TR-4926.

[44] B. Zhou and J. Pei. Preserving privacy in social networks against neighborhood attacks. In *Proceedings of the 24th International Conference on Data Engineering (ICDE)*, April 2008.

[45] B. Zhou, J. Pei, and W. Luk. A brief survey on anonymization techniques for privacy preserving publishing of social network data. *SIGKDD Explor. Newsl.*, 10(2):12–22, 2008.

[46] L. Zou, L. Chen, and M. T. Özsu. k-automorphism: a general framework for privacy preserving network publication. *Proc. VLDB Endow.*, 2:946–957, August 2009.