INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 2258-2267

INFERENCE ATTACK PREVENTION OF PRIVATE INFORMATION ON SOCIAL NETWORKS

Ms. Vina Varade[1], Dr. Monika Deshmukh[2], Dr. Mohammad Muqeem[3], Dr. Pawan R. Bhaladhare[4]

[1]M. Tech , Sandip University, Nashik, Maharashtra, India.

[2,3,4]Professor, Sandip University, Nashik, Maharashtra, India.

ABSTRACT

With increasing user involvement, social networks nowadays serve as a repository of all kinds of information. While there have been various studies demonstrating that private information can be inferred from social networks, few have taken a holistic view on designing mechanisms to detect and alleviate the inference attacks. In this study, we present a framework that leverages the social network data and data mining techniques to proactively detect and prevent possible inference attacks against users. A novel method is proposed to minimize the modifications to user profiles in order to prevent inference attacks while preserving the utility.

1. INTRODUCTION

The Importance of Inference Attack Prevention of Private Information on Social Networks

Along with the increasing popularity of OSNs, OSNs users encounter growing risks of privacy attacks. Online Social Networking site provides a rich digital forum for social interactions. While some OSNs users are open to sharing and posting personal information, others may not (Krishnamurthy & Wills, 2008; Post & Walchli, 2014). Often, the privacy risks posed to OSNs users are unexpected and unaware of. The most immediate danger of posting on OSNs is that it may leave a permanent fingerprint of whatever being posted (Rosenblum, 2007). With the click of a button, what a user just typed could be instantly disseminated and stored in countless independent permanent places. Next-minute damage recovering is almost impossible. Worse, the power of a search engine makes it searchable within a few seconds. Some concrete examples of the privacy risks posed to OSNs users include stalking, spamming, and the possible damage to their future educational and career opportunities. With the information from a user profile, a potential adversary can determine the likely physical location of the user and thus incur real-world stalking (Gross & Acquisti, 2005).

Individual users of an online social network can create profiles containing various personal attributes such as age, group affiliation, lists of personal interests, contact information, and so on. Some OSNs offer the ability to create and join groups that share common interests or affiliations, upload live videos, and hold discussions in forums. To alleviate privacy concerns, OSNs usually have controls that allow users to choose their own privacy settings (Barnes, 2006), for instance, whether to make your profile public or private, who is allowed to view the profile, contact you, add you to their list of contacts, and so on. Despite the availability of the privacy setting controls to the users, such controls may not be sufficient in enforcing privacy of users (e.g., Li, Li, Yan, & Deng, 2015; Singh, Bhola, & Lee, 2009). This has been evidenced by both the incidents reported in media outlets and the studies from the academia (e.g., Dwyer, Hiltz, & Passerini, 2007; Gross & Acquisti, 2005; Hazari & Brown, 2013; Zeller, 2006). In addition, OSNs users may be vulnerable to various privacy attacks, such as automated user profiling (Balduzzi et al., 2010; Dougnon, Fournier-Viger, & Nkambou, 2015), identity attacks via user de-anonymization techniques (e.g., Narayanan & Shmatikov, 2009; Wondracek, Holz, Kirda, & Kruegel, 2010), and inference attacks (e.g., He, Chu, & Liu, 2006; Lindamood, Heatherly, Kantarcioglu, & Thuraisingham, 2009; Zheleva & Getoor, 2009), and so on.

In order to prevent some of the above privacy attacks, technical solutions have been proposed in the literature. For instance, several anonymization solutions have been proposed to protect against identity attack in OSNs (e.g., Backstrom, Dwork, & Kleinberg, 2007; Liu & Terzi, 2008; Zou, Chen, & Özsu, 2009). Yet, preventing inference attacks has been largely neglected until recently that a study has proposed sanitization techniques to prevent inference attacks (Heatherly, Kantarcioglu, & Thuraisingham, 2013). However, there lacks a holistic approach to detect and alleviate inference attacks posed to the users in the general context of online social networks.

In this study, we present a framework that leverages the social network data and existing data mining techniques to proactively detect and prevent possible inference attacks against users. We also propose an approach to minimize the modifications to user profiles in order to prevent inference attacks while preserving the utility. In addition, the modularized design of our framework enables the flexibility in implementing the proposed plug-in components.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 2258-2267

The rest of the article is organized as follows. We present the background research in second section. The third section introduces inference attacks in the context of online social networks. The proposed framework is then provided in the fourth section. Finally, conclusion and future research are presented.

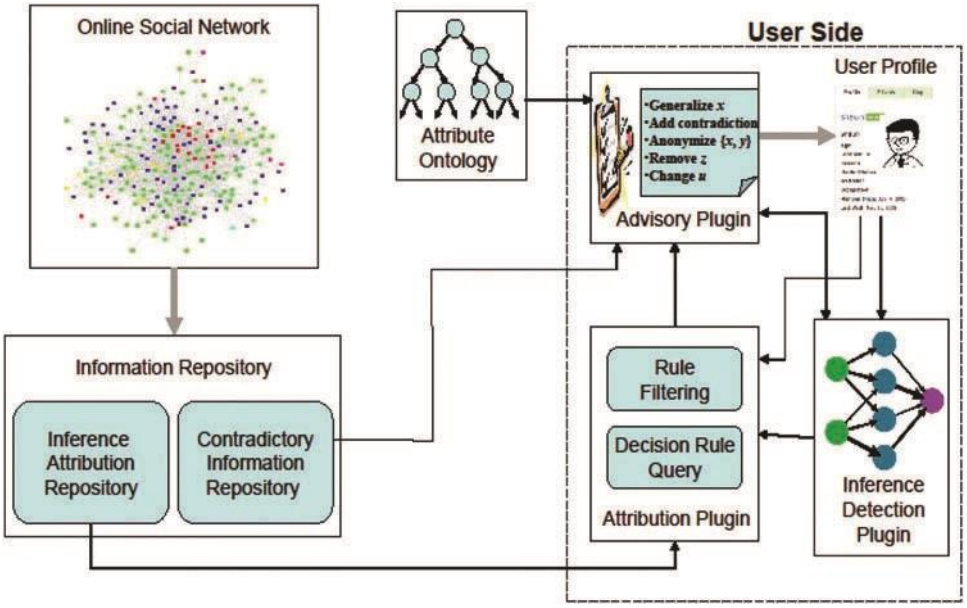

Fig. 1. Figure of The framework for inference detection and prevention.

1.1 Problem Statement

Privacy risks in online social networks

Sweeney (2002a) shows that a large portion of the U.S. population can be re-identified using a combination of 5-digit ZIP code, gender, and date of birth. The majority of OSNs users disclose their location, gender, and age on their profiles. Therefore, an adversary can link a OSNs user to other deidentified data sources such as voter registration record, which leads to re-identification. An additional re-identification risk is that profile information may be used to estimate a user's social security number and exposes him /her to identity theft (Gross & Acquisti, 2005).

To demonstrate the real risk that OSNs users can encounter, several recent studies have instrumented identity attacks on real users in real online social network data (e.g., Bilge, Strufe, Balzarotti, & Kirda, 2009; Narayanan & Shmatikov, 2009; Wondracek et al., 2010). For example, Bilge et al. (2009) demonstrated how easy it was for a potential attacker to launch automated crawling and identity theft attacks against a number of popular social networking sites in order to gain access to a large volume of personal user information.

The premise of these attacks is that network owners often share their online social network and user information with advertising partners and other third parties. The sharing indeed forms a part of their business operations. In some cases, social networks are even published for research purposes. To allay privacy concerns, the networks are typically "sanitized" through anonymization, a simple procedure in which each user's personal identifiable information, i.e., name, age, or location, associated with individual nodes is suppressed or generalized. Can anonymization really protect privacy?

Unfortunately, the answer is "No." To demonstrate the re-identification risk, Narayanan and Shmatikov (2009) developed a de-anonymization algorithm targeting anonymized social network graphs. Despite based purely on the social network topology, the algorithm was used to show that a third of the users who can be verified to have account on both Twitter, a social networking and microblogging service, and Flickr, an online image/video sharing site, can be re-identified in the anonymized Twitter network with only a 12% error rate. That is, even anonymized social network data are still subject to privacy risks of re-identification.

1.2 Applications

Inference attacks in online social networks

Recent studies show that even OSNs users take advantage of the available privacy setting control to make their profiles or sensitive attributes private (e.g., no one but their friends can see their profile detail), it is still possible to infer or predict private information that a user is not willing to disclose (e.g., Davis, Pappa, De Oliveira, & De L Arcanjo, 2011; Dougnon et al., 2015; He et al., 2006; Heatherly et al., 2013; Lindamood et al., 2009; Tang et al., 2011; Zheleva & Getoor, 2009). In fact, these studies demonstrate that a surprisingly large amount of private

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 2258-2267

information can be inferred by just exploiting friendship links and/or group affiliations and/or partial social network data.

Specifically, the problem of inference attacks in an online social network is to infer the hidden values of sensitive attributes in user profiles that are conditioned on the observed attribute values, friendship links and group memberships, and/or other information available. It is commonly assumed that an adversary can apply a probabilistic model for predicting the hidden sensitive values (e.g., Heatherly et al., 2013; Zheleva & Getoor, 2009). By combining the given information of a social network in different ways, the adversary can launch various inference attacks (e.g., attacks without links and groups, attacks using links, attacks using groups, or a mix of them, etc.).

As a simplified example, suppose users have attributes of gender and self-declared political views in a collected Facebook data (Zheleva & Getoor, 2009). Assume that user Alice has set her profile as private. From the group information available in Facebook, Alice belongs to four groups: Crochet Mastery, the National Coalition on Health Care, Health Care for America Now, and Hand Embroidery. Based on these public information, the group-based classification model can accurately predict that Alice is female and a liberal. With assumption of 50% private profiles, the attack accuracy for gender attribute reaches at 73.4% by using group-based classification model; and the accuracy is 72.5% by using both links and groups.

The above inference problem does have significant privacy implications. Since fewer users in OSNs hide their friendship links and even if they do, their friendship links can still be constructed through the backlinks from their public profile friends. Group participation information availability is similar—even if a user keeps his/her profile private, his/her participation in a group is displayed on the group's membership list. Currently, most OSNs (e.g., Facebook and MySpace) do not allow users to hide their group membership from public groups.

In addition, inference attacks to OSNs users may be motivated by the various self-interests of the attackers, such as targeted marketing, insurance screening, email phishing, or political monitoring (Bonneau & Preibusch, 2010). These attacks would have negative effects on the users and even the credibility and attractiveness of OSNs. Further, research has shown that simply removing some attribute information or friendship links may not be enough to prevent inference attacks (e.g., Heatherly et al., 2013; Lindamood et al., 2009). Therefore, it is critical for both OSNs providers and OSNs users to take proactive approaches in protecting them against undesired private information digging and inform them of possible privacy breaches. With large amount of social network data collected by various parties and the advances of data mining techniques, a holistic view on the inference problem and being able to leverage data-driven approaches may hold great promise to detect and prevent the problem in practice.

1.3 Challenges

The Inference detection plugin is simply a classifier model made available to users by the service provider or third party to support inference detection. This could be based on any data mining techniques. For example, the service provider could train a neural network based classifier using the online social network data. When the trained model is provided to the users, they can easily run their profile through the model to identify possible inferences. One requirement of such an approach would be that the underlying models have to be scalable as well as incrementally updatable to incorporate the ever-expanding data accrued in the social network.

2. LITERATURE SURVEY

Functionality description

We now discuss the detailed functionality of each component of the framework and the methods employed to support such functionality.

Inference attribution repository

The inference attribution repository stores a set of rules for all the attributes that are declared private/sensitive in at least one of the users' profile. These rules are derived from the online social network data by applying classification algorithms such as decision tree learners C4.5 for the given set of sensitive attributes. Specifically, for each sensitive attribute, a decision tree can be learned from the data that provides the combination of attribution values leading to the inference of the private data. For example, given a sensitive attribute value y, a standard decision tree generation algorithm C4.5 could potentially come up with an attribution rule that if attribute values x and z are present in the profile or if w and q are present in the profile, y could be inferred. For simplicity, we assume that the decision rules are expressed as the following conjunctive normal form: $Attr_i \wedge AttrVal_j == > TargetValue$

where $Attr_i$ in the conditional part of the rule corresponds to checking the presence of such attribute in the user profile, for example whether user's location is included in his/her profile. $AttrVal_j$ in the conditional part corresponds to checking the value of given attributes for satisfaction of the rule, for example "Location" = Florida. In essence, the

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 2258-2267

rules in the inference attribution repository specify, for a sensitive target value, what the possible lists of affecting conditions are. For convenience, we only use attribute values for illustrations thereafter. For example, given the following simple rule: Location("Sunshine State") = = > "Florida," where the attribute value being affected is "Sunshine State"; this rule states that if a user specifies the location in her profile as "Sunshine State," it simply implies that she is a Florida resident.

Contradictory information repository

Contradictory Information Repository is a component that is used to record inference rules that directly contradict the detected sensitive information. Such contradictory information can also be used to break the detected inferences, since its presence may directly contraindicate the presence of the inferred information. This can be a very appealing option if the user would like to leave their current profile unchanged but still create the illusion of possessing the contradictory attribute.

Then, the question is how we can find such contraindicative information. Such information can actually be derived from the social network data itself, by using a clever trick. The key idea is that we are looking for anticorrelations. To detect these, we thus can reverse or modify the target attribute value in the original data based on which an initial new inference is detected. Then, the same module used for the inference detection (say, decision tree learner) can be run against the newly reversed data. This will give us a new set of inference rules that indicate the given target value. Since we have purposely reversed the original target value, comparing the affecting attribute values in these new rules with those in the initial inference would tell us the contradictory information existing between them. The following steps in pseudo-code depict the basic data of how the contradictory information rules are generated.

1: for each sensitive attribute do

2: Mark it as the class attribute

3: for each data instance do

4: Modify the class label for that instance to be its complement (e.g., change 0 to 1)

5: end for

6: Run the decision tree learner to learn attribution rules

7: end for

In our running example, we can infer that Alice's birthdate is "January 1, 1992." To find out the contradictory information for breaking this inference, we can change her birthdate to, say, "January 1, 1990." After running the detection module on this changed data, the user may get a new inference, say, Location("Yellowhammer State") ∧ GroupAffiliation("NewYearBornParty") ∧ RecentActivity ("Just qualified for full driver license") = = > BirthDate("January 1, 1990"). Comparing the affecting attribute values in this rule with those in the initial inference rule, we can see that the difference is in the Location attribute value. Thus, we may state that Location("Yellowhammer State") is a contradictory information to Location("Sunshine State"), which can be added to the Contradictory Information repository.

Inference detection plugin

The Inference detection plugin is simply a classifier model made available to users by the service provider or third party to support inference detection. This could be based on any data mining techniques. For example, the service provider could train a neural network based classifier using the online social network data. When the trained model is provided to the users, they can easily run their profile through the model to identify possible inferences. One requirement of such an approach would be that the underlying models have to be scalable as well as incrementally updatable to incorporate the ever-expanding data accrued in the social network.

Attribution plugin

The attribution plugin interacts with the information repository to determine the possible attributes or attribute values in the user profile that leads to disclosure of sensitive attributes as determined by the inference detection plugin. Specifically, the attribution plugin queries the inference attribution repository for all the rules that includes the given sensitive attribute in the target value or the target value of such rules directly or indirectly influences the conditional part of the rule with the given sensitive attribute in the target value. For example, for a given sensitive attribute value y, the rules i) a ∧ b ∧ c = = > y; ii) a ∧ b ∧ d = = > z; and iii) z ∧ x ∧ p = = > y will be included in the query result, where i) includes the sensitive attribute y in the target value and ii) includes a target value that directly influences the conditional part of iii) that includes y as its target value.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 2258-2267

Since the attribution repository includes rules created from the entire social network, some of the rules in the query result may not be relevant, as the given user profile may not include the attributes specified in those rules. Therefore, a filtering process is employed to purge out the irrelevant rules. In reference to our running example, there are three attributes in Alice's profile. All the rules related to birth date inference that do not have any direct or indirect influence on any of the attributes in Alice profile are filtered out.

3. METHODOLOGY

Advisory plug-in

Advisory plug-in is run locally at the user site but it interacts with the contradictory information repository, attribute ontology that may be hosted by a third party, and attribution plugin running at user site to suggest possible modifications in the user profile for breaking inference. The suggested modifications such as generalization, perturbation, suppression, etc. have been used for data anonymization or sanitization for privacy protection in the literature (e.g., Heatherly et al., 2013; Sweeney, 2002b). Specifically, some possible modifications can be the following:

Generalize attribute values. This option can utilize an attribute ontology for generalization of the relevant attribute values. For example, for the attribute Location, instead of directly using a state's nickname, it can be generalized into region (i.e., Southeast region).

Perturb attribute values. To break the inference, an attribute value can be distorted. For instance, "Sunshine State" can be perturbed as "Yellow hammer State" changing the user location from "Florida" to "Alabama."

Add contradictory information. Another possibility is to actually introduce new data within the profile that directly contraindicates the presence of the sensitive value. For example, if

we add the possession of a driver license to a profile, as discussed above, this contradicts the hypothesis that the user is underage.

Remove certain information. Instead of directly adding contradictory information, another possibility is to simply eliminate some of the information causing the inference, thus breaking the linkage. In the context of the running example, removing the user's group affiliation from the "NewYearBornParty" group will break the inference link. The user's name may also need to be removed from the group page.

Make some more information private. Instead of making any modifications to the user profile, an alternative is to just make the attributing information itself private. In this fashion, it cannot be used to make the sensitive inferences. Of course, this only applies as long as the newly private made information itself cannot be inferred, thus possibly requiring multiple runs of the inference prevention process.

Or a combination of all of the above. In certain cases, a single modification of any of the above kinds may not be sufficient to ensure privacy, in which case, we can use a combination of the above to enforce it.

In order to guarantee that the above-instrumented modifications have made the user free of inferences, it may be necessary to run inference detection module again. This is due to the possibility of creating new inferences by the modifications.

Ontology

As discussed above, breaking inference may require generalizing some of the attribute values in user profile. For this, the advisory plug-in can utilize a relevant ontology which may be provided by a third party, such as VIVO ontology repositories (VIVO, 2015) and DAML ontology (DAML, 2015). The generalizable attributes in a user profile can be categorized into multiple types based on their semantics and the domains they take their value from. There can be different attribute categories such as address and locations, organizational affiliations, etc. that are generalizable and are commonly used in the online profiles of users. For each category, ontology from appropriate domains needs to be used for attribute value generalization.

Minimizing modifications to preserve utility

In order to break the detected inferences, we may need to modify several attribute values in the user profile. Ideally, we would like to minimize this modification while still breaking the inferences. That is, usability should be enhanced as much as possible while satisfying privacy protection requirement.

Recall that, to determine the inference source for a user, the Attribution plug-in may get a set of candidate rules from the information repository. Based on these rules, modification suggestions are given. However, among the candidate rules, there may be overlaps on the attribute values being affected. That is, some of the attribute values may appear in multiple candidate rules. Given our rulebased approach, minimizing the modification of user profile is thus equivalent

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 2258-2267

to finding a minimal set of attribute values being affected such that all the candidate rules contain at least one of these attribute values. In other words, we need to find a set of attribute values of the smallest size which would still break all the candidate inferences.

For example, suppose we have the following three candidate rules.

R_1: $AttrVal_1 \wedge AttrVal_2 \wedge AttrVal_3 = => TargetValue_1$

R_2: $AttrVal_2 \wedge AttrVal_4 = => TargetValue_2$

R_3: $AttrVal_1 \wedge AttrVal_4 = => TargetValue_3$

Then, one minimal set of the attribute values is {$AttrVal_1$, $AttrVal_2$}, which makes the conjunctive forms of all the candidate rules false. This essentially breaks the inferences. Note that the minimum solution is not unique. For instance, some other minimal sets for this example include {$AttrVal_1$,

$AttrVal_4$}, {$AttrVal_2$, $AttrVal_4$}, {$AttrVal_3$, $AttrVal_4$}.

However, it turns out that finding such a minimal set of attribute values is actually NP-hard (Bovet, Crescenzi, & Bovet, 1994). That is, it is a computationally hard problem. To show this, we now provide a brief reduction from the minimum set cover problem that can be reduced to the problem of finding the minimal set of attribute values. More formally, given a universe U and a family S of subsets of U, a cover is a subfamily C that is a subset of S of sets whose union is U. Given the input as a pair (U, S), the minimum set cover problem is to find a set covering C that uses the fewest subsets, which is known as an NP-hard optimization problem.

Reduction: minimum set cover = = > our problem

We first introduce some notations. Let a family S of subsets be $S_1, \ldots, S_m$, $U = \{e_1, \ldots, e_n\}$ ($|U| = n$), and S_i is a subset of U ($1 \leq i \leq m$).

Now, we relate it to our problem. Let each set S_i correspond to an $AttrVal_i$ ($1 \leq i \leq m$), i.e., $S_1 \equiv AttrVal_1, \ldots, S_m \equiv AttrVal_m$. Then, based on whether an element e_i is in $AttrVal_j$ ($1 \leq i \leq n$, $1 \leq j \leq m$), we form n clauses $C_1, \ldots, C_n$ in the following form: $C_i = \wedge_j AttrVal_j$, where $e_i \in AttrVal_j$. That is, each clause corresponds to the conjunctive form of an inference rule.

Given a set of inference rules with their respective conjunctive forms as $C_1, \ldots, C_n$, where $C_i = \wedge_j AttrVal_j$ ($1 \leq i \leq n$, $1 \leq j \leq m$), our problem is to find a minimum set of AttrVals that make the conjunctive form of each rule false. With the above-defined relations between S_i and C_i, we can see that finding the minimum set cover is essentially equivalent to finding the minimal set of AttrVals for breaking all of the rules.

The following example illustrates how the above reduction works. We first have the minimum set cover instance as follows. Let $U = \{e_1, e_2, e_3\}$, and a family S of five subsets: $S_1 = \{e_1, e_3\}$, $S_2 = \{e_1, e_2\}$, S3 = {e1}, S4 = {e2, e3}. That is, we have n = 3 and m = 4. Obviously, the minimum set cover for this instance includes {S_1, S_2}, {S_1, S_4}, {S_2, S_4}, {S_3, S_4}.

Now, let each set S_i correspond to an $AttrVal_i$. Thus, $AttrVal_1 = \{e_1, e_3\}$, $AttrVal_2 = \{e_1, e_2\}$, $AttrVal_3 = \{e_1\}$, $AttrVal_4 = \{e_2, e_3\}$. Then, based on whether an element e_i is in $AttrVal_j$ ($1 \leq i \leq 3$, $1 \leq j \leq 4$), 3 clauses are formed: $C_1 = AttrVal_1 \wedge AttrVal_2 \wedge AttrVal_3$; $C_2 = AttrVal_2 \wedge AttrVal_4$; $C_3 = AttrVal_1 \wedge AttrVal_4$. Indeed, each clause C_i is exactly the conjunctive form of each rule R_i in our previous example. Then, each e_i is equivalent to the corresponding C_i ($1 \leq i \leq 3$). With the abovedefined relationship between S_i and $AttrVal_i$, it is clear that finding the minimum set cover for input (U, S) is the same as finding the minimum set of AttrVals to make all the three clauses false.

We can also verify the solutions. For example, the minimal sets for making the three clauses false include {$AttrVal_1$, $AttrVal_2$}, {$AttrVal_1$, $AttrVal_4$}, {$AttrVal_2$, $AttrVal_4$}, {$AttrVal_3$, $AttrVal_4$}. These correspond to the given minimal set cover solutions listed earlier.

Reduction: our problem = = > minimum set cover

Conversely, we now show that our problem of finding the minimal set of attribute values can be reduced to the minimum set cover problem. Let the number of clauses in the candidate rules be n, and the number of AttrVals be m. For each clause C_i, we create an equivalent element e_i (i.e., $e_1 \equiv C_1, \ldots, e_n \equiv C_n$). For each $AttrVal_i$ in the clauses, we create a set S_i ($1 \leq i \leq m$). Now, based on the condition whether $AttrVal_i$ is included in element e_j ($1 \leq j \leq n$, $1 \leq i \leq m$), we assign the element e_j to set S_i. That is, S_i = {e_j, where e_j includes $AttrVal_i$}. As a set cover problem, S consists of all S_i, and the union of all S_i forms U. With the above-defined relations between C_i and S_i, finding the minimal set of AttrVals is consequently reduced to finding the minimum set cover of input (U, S).

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 2258-2267

Input: a universe $\tilde{U}$ and a family $\hat{S}$ of subsets of $\tilde{U}$.
Output: set cover C of input ($\tilde{U}$, $\hat{S}$).

1: Assign $\tilde{U}$ to U;
2: Assign an empty set to C;
3: **while** U is not equal to empty set
4: Select an S from $\hat{S}$ that maximizes $|S \cap U|$;
5: Remove U from S;
6: Add S to C;
7: **end while**
8: Return C

Figure 1. The method of greedy set cover.

Given the three rules in the preceding example, we now have three equivalent elements: $e_1 = AttrVal_1 \wedge AttrVal_2 \wedge AttrVal_3$, $e_2 = AttrVal_2 \wedge AttrVal_4$, $e_3 = AttrVal_1 \wedge AttrVal_4$. With the 4 AttrVals in the rules, we create four corresponding sets: S_1, S_2, S_3, and S4. Then, we assign each element into these sets. For example, since $AttrVal_1$ is included in e_1 and e_3, we have $S_1 = \{e_1, e_3\}$.

Similarly, we get $S_2 = \{e_1, e_2\}$, $S_3 = \{e_1\}$, $S_4 = \{e_2, e_3\}$.

Since the problem is NP-hard, clearly we cannot design an optimal polynomial-time solution for it. However, since our problem is directly reducible to the set cover, we can use the greedy algorithm designed for it which guarantees a ln(n) approximation. Then, based on the greedy solution, we can subsequently minimize the modifications to user profile.

Figure 1 provides the steps on how the greedy method works (Cormen, Leiserson, Rivest, & Stein, 2009). Specifically, let U be the set of remaining uncovered elements and C be the cover being constructed. At each iteration, a subset S is chosen that covers as many uncovered elements as possible (with ties broken arbitrarily). After S is selected, its elements are removed from U, and S is placed in C. When the algorithm terminates, it returns the set covering C. This greedy algorithm can easily be implemented to run in time polynomial in $O(|\tilde{U}||\hat{S}|\min(|\tilde{U}|, |\hat{S}|))$.

Discussion

While OSNs are becoming an increasingly popular platform for individuals to connect, share, and communicate with other people around the world, the extensive disclosure and display of personal data have made privacy concerns particularly salient in recent years. It is especially noteworthy that the privacy risks posed to OSNs users are often unexpected and unaware of. Some recent studies have focused on uncovering privacy exploits and inference attacks that could unknowingly occur to OSNs users. However, few have taken a holistic view to examine and address the inference attacks that users may not be well aware of. In this study, we have proposed a data-driven and holistic framework to alleviate the rule-based inference problem by detecting and breaking the detected inferences. In our framework, not only OSNs service providers play a proactive role in ensuring privacy protection, also a mechanism is offered for OSNs users to have more control over their personal information through the client side plug-in tools. Such collaborative format shall entail tangible benefits to both sides in the long run. We subsequently discuss the implications of our research from both academic side and practical perspective.

Implications for research

This article's focus on proposing a holistic framework for mitigating the inference attacks complements the stream of research that has focused more on identifying the attacks in the context of OSNs. For example, a comprehensive empirical study was conducted to assess the feasibility and accuracy of inference attacks that are launched from the extension API of OSNs (Ahmadinejad & Fong, 2014).

Chaabane, Acs, and Kaafar (2012) suggest that it is possible to infer users' undisclosed personal particulars from public shared interests and public personal particulars of other users who have similar interests. Recent scientific results have shown that social network Likes, such as the "Like Button" records of Facebook, can be used to automatically and accurately predict even highly sensitive personal attributes (Kosinski, Stillwell, & Graepel, 2013). Li et al. (2015) discover a series of privacy exploits and find that most of these exploits are inherent due to the conflicts between privacy control and OSNs functionalities. Also, a user's social graph can be inferred from their public listings (Bonneau, Anderson, Anderson, & Stajano, 2009).

All the above findings demand greater stress on searching for proactive solutions to address the privacy issues in OSNs. Recently, research efforts have been made to specifically mitigate specific privacy risks arisen in OSNs. For

instance, Stern and Kumar (2014) suggest to use a wheel interface to improving privacy settings control in online social networks. Buccafurri, Fotia, Lax, and Saraswat (2016) propose a protocol that is able to keep Likes un-linkable to the identity of their authors, in such a way that the user may choose every time she expresses a Like, those non-identifying (even sensitive) attributes she wants to reveal. However, few have taken a holistic view and data-driven approach as we have done in this study. As mentioned earlier, our framework involves both sides of an OSNs platform, i.e., the providers and the users. Such collaborative set-up creates a win-win situation.

This research also contributes to the general literature in examining the challenges and opportunities of addressing privacy issues that arise in OSNs. As noted by Zhang, Sun, Zhu, and Fang (2010), there are inherent design conflicts in existence between some security and privacy goals and traditional design goals such as usability and sociability of OSNs. An ultimate solution may require interdisciplinary expertise and collaborative efforts, for instance, including experts from the social science and network security communities, industry, regulatory bodies, and all other relevant communities to make decisions on both mechanisms and policies. This research is intended to provide one of the starting points toward the goal of privacy-preserving OSNs. Hopefully, this study will prompt future research to further identify effective and practical ways to achieve such goal.

Implications for practice

From a practical perspective, this study provides meaningful implications. Past research has suggested that online users consider it most important to (1) be aware of and (2) have direct control over personal information stored in service providers' databases (Malhotra, Kim, & Agarwal, 2004). In addition, users should be allowed to control, i.e., add, delete, and modify at will, their information. These organizational efforts can jointly mitigate an individual user's privacy concerns (Stewart & Segars, 2002). Such mitigation may subsequently lead to important benefits for OSNs providers, such as continuously growing their user base and generating revenue.

First, as argued by Yeung, Liccardi, Lu, Seneviratne, and Berners-Lee (2009), the future of online social networking lies in a decentralized approach. The researchers point out privacy is one of the two major problems facing OSNs. They point out that decentralization provides a mechanism that allows users to have more control over their own data so as to address their privacy concerns. The lack of user confidence in the mechanisms of privacy control can be one of the major problems hampering the growth as well as the real usage of OSNs (Tucker, 2014). Conversely, providing users the mechanism of privacy control can attract more users to sign up and stimulate further growth in real usage. Therefore, it would be in the best interest of OSNs providers to work with other stakeholders including user community to actually support and implement such mechanism.

Second, recent study shows that an increased perception of control over personally identifiable information can help the user of a social network website to more readily engage with click-through advertising (Tucker, 2014). Given that OSNs providers heavily rely on advertising to generate revenue, this suggests that publicly giving users control over their private information can financially benefit social networking providers. In addition, extant research suggests that if OSNs providers are successful at reassuring consumers that they are in control of their privacy, firms can use personalization of ads to generate higher click-through rates (Aguirre, Mahr, Grewal, De Ruyter, & Wetzels, 2015). In fact, it not only benefits OSNs providers, but also advertisers that do advertisement on social networking sites (Tucker, 2014). Thus, giving users the appropriate tools for controlling their privacy may be an important way of ensuring that the advertising-supported OSNs can continue to thrive.

Conclusion and future work

In this study, we have focused on the inference problem arising in the popular online social networking sites. We propose a framework to alleviate the rule-based inference problem by detecting and breaking the inferences that are represented as rules of attributes and/or attribute values. To prevent privacy inferences, user profiles may need to be modified. We also demonstrate that how this modification can be minimized so that usability can be enhanced.

For future work, we plan to eventually implement and release proposed privacy plugins for the users of social networking platforms such as Facebook. To this end, we intend to first develop an adhoc prototype in the context of e-learning, where college students will be the primary users of the prototype. We can solicit useful feedback from the community of these users. In particular, with the wide use of smart phones on college campus, each of the proposed privacy plug-ins will be initially developed as a mobile application (app). We hope that, with the apps conveniently available on the mobile phones, it will help attract large number of users to try and use them. This may lead to a wide array of constructive feedback about the functionality and efficacy of the apps running as part of the prototype. Then, based on the comments and feedback from the users, we will refine and further improve the overall framework as well as its modular plug-ins. Finally, the ad-hoc prototype will be enhanced and extended to be a fully functional implementation of the proposed framework.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 2258-2267

While our framework protects users from common privacy inferences, background knowledge when added to the user's posts and interactions may still suffice to breach privacy. We plan to examine this issue in the future.

4 Framework overview

In our proposed framework, the online social network data are used to develop an information repository consisting of an inference attribution repository and a contradictory information repository. This can be done either by the social network service provider or any other involved parties. The inference repository is created by applying classification methods such as decision tree learners, for example, C4.5 (Quinlan,

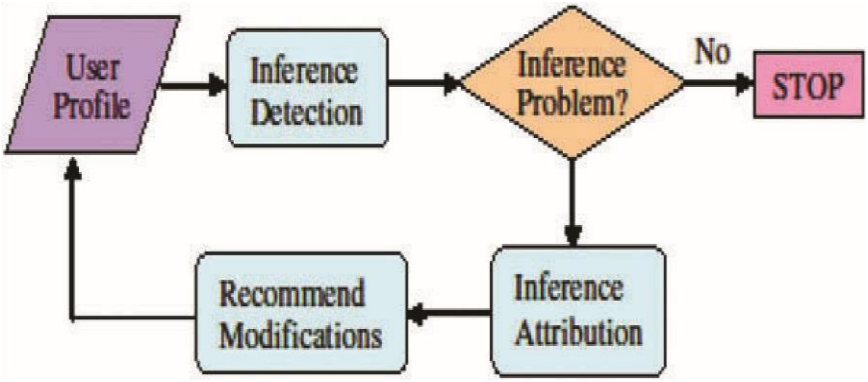

Figure 2. The flow chart for inference detection and prevention.

2014), on the online social network data to discover the attribution rules for sensitive attributes. The contradictory information repository is also a set of rules where the attributes/attribute values in the conditional part of the rule contraindicate (inversely relate to) the sensitive attribute value.

The framework also includes a set of plugins including inference detection plugin, attribution plugin, and advisory plugin. All these plugins can be downloaded by a user and run at his/her site to check any privacy breach or change the profile to avoid such breach. The inference detection plugin is primarily a classifier that given a user profile predicts what sensitive attributes of the user can potentially be leaked. This can be based on any combination of the state of the art detection techniques that have been developed in the literature (e.g., De Vel, Anderson, Corney, & Mohay, 2001; Koppel & Schler, 2004; Staddon, Golle, & Zimny, 2007).

The process flow for inference prevention is shown in Figure 2. This process is initiated by the user whenever the user creates a new online profile or updates an existing profile. This profile is taken as an input to the inference detection plugin running locally at the user site. Based on the privacy preferences of the user, the inference detection plugin detects if the user profile is consistent with the privacy preferences or could possibly leak sensitive attribute values. As a simple running example, suppose on January 4, 2010, user Alice has the following attribute values in her profile: Location = "Sunshine State," Group Affiliation = "NewYearBornParty," and Recent Activity = "Just qualified for full driver license." Since "Sunshine State" implies that she is in Florida, "NewYearBornParty" is a group for people who are born on January 1, and the age limit for obtaining full driver's license in Florida is 18, thus it can be inferred that Alice's birthdate is January 1, 1992. The inference detection plugin detects the following inference: Location ("Sunshine State") and GroupAffiliation ("NewYearBornParty") and RecentActivity("Just qualified for full driver license") infers BirthDate("January 1, 1992").

Once an inference problem is detected, the attribution plugin is activated to determine the source of the problem (i.e., what causes the detected inference). Specifically, the attribution plugin interacts with the inference attribution repository in the knowledge base to identify the possible attribute value(s) contributing to the detected inference problem. Based on the relevant inference rules retrieved from the inference attribution repository, the advisory plugin makes suggestions to the user on what needs to be modified in the profile to break the detected inference.

We note that online social network is a dynamic environment, thus the contents of the information base as well as the inference detection plugin should evolve over time. In other words, in order to reflect the most current state of potential inferences, the rules in the information repositories need to be updated or added. This evolving process may also involve learning from the past experience. The goal is to make it better capture the emergent inferences as much as possible.

4. FUNCTIONALITY DESCRIPTION

We now discuss the detailed functionality of each component of the framework and the methods employed to support such functionality.

5. REFERENCES

[1] Aguirre, E., Mahr, D., Grewal, D., De Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. Journal of Retailing, 91 (1), 34–49. doi:10.1016/j.jretai.2014.09.005

[2] Ahmadinejad, S. H., & Fong, P. W. (2014). Unintended disclosure of information: Inference attacks by third-party extensions to Social Network Systems. Computers & Security, 44, 75–91. doi:10.1016/j.cose.2014.04.004

[3] Backstrom, L., Dwork, C., & Kleinberg, J. (2007). Wherefore art thou r3579x?: Anonymized social networks, hidden patterns, and structural steganography. Proceedings of the 16th international conference on World Wide Web (pp. 181–190), Banff, AB, Canada: ACM.

[4] Balduzzi, M., Platzer, C., Holz, T., Kirda, E., Balzarotti, D., & Kruegel, C. (2010). In Jha, S., Sommer, R., and Kreibich, C. (Eds.), Abusing social networks for automated user profiling. In Recent advances in intrusion detection (pp. 422–441). Berlin Heidelberg: Springer.

[5] Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. First Monday, 11 (9). doi:10.5210/fm. v11i9.1394

[6] Berry, M. J., & Linoff, G. (1997). Data mining techniques: For marketing, sales, and customer support. New York, NY: John Wiley & Sons, Inc.

[7] Bilge, L., Strufe, T., Balzarotti, D., & Kirda, E. (2009). All your contacts are belong to us: Automated identity theft attacks on social networks. Proceedings of the 18th international conference on world wide web (pp. 551–560), Madrid, Spain: ACM.

[8] Bonneau, J., Anderson, J., Anderson, R., & Stajano, F. (2009, March). Eight friends are enough: Social graph approximation via public listings. Proceedings of the Second ACM EuroSys Workshop on Social Network Systems (pp. 13–18), Nuremberg, Germany: ACM.

[9] Bonneau, J., & Preibusch, S. (2010). The privacy jungle: On the market for data protection in social networks. In Moore, T., Pym, D., and Ioannidis, C. (Eds.), Economics of information security and privacy (pp. 121–167). Boston, MA, US: Springer.

[10] Bovet, D. P., Crescenzi, P., & Bovet, D. (1994). Introduction to the theory of complexity (pp. 220–229). Englewood Cliffs, NJ: Prentice Hall.

[11] Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. Journal of Computer- Mediated Communication, 13, 210–230. doi:10.1111/j.1083-6101.2007.00393.x

[12] Buccafurri, F., Fotia, L., Lax, G., & Saraswat, V. (2016). Analysis-preserving protection of user privacy against information leakage of social-network Likes. Information Sciences, 328, 340–358. doi:10.1016/j.ins.2015.08.046

[13] Chaabane, A., Acs, G., & Kaafar, M. A. (2012, February). You are what you like! Information leakage through users' interests. Proceedings of the 19th Annual Network & Distributed System Security Symposium (NDSS), San Diego, CA.

[14] Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). Introduction to algorithms. Cambridge, MA: MIT press.

[15] DAML. (2015). Ontologies by Keyword. Retrieved October 22, 2015, from http://www.daml.org/ontologies/keyword. html

[16] Davis, C. A., Jr, Pappa, G. L., De Oliveira, D. R. R., & De L Arcanjo, F. (2011). Inferring the location of twitter messages based on user relationships. Transactions in GIS, 15 (6), 735–751. doi:10.1111/j.1467-9671.2011.01297.x