BBN-Based Privacy Management Sytem for Facebook

Enkh-Amgalan Baatarjav and Ram Dantu
Department of Computer Science and Engineering
University of North Texas
Denton, Texas, 76203, USA
Email: {eb0050, rdantu}@unt.edu

Yan Tang and João Cangussu
Department of Computer Science
University of Texas at Dallas
Richardson, Texas, 75080, USA
Email: {yxs055100, cangussu}@utdallas.edu

***Abstract*—Online social networking sites (SNSs) has changed our lifestyle and become a main medium of communication among young adults to stay in touch with their friends, to organize events, to make friends, to promote themselves, to date, etc. To create content rich environment, SNSs make their platform available for third-party developers. The developers can build their applications based on users' social graph containing their personal and social information. Unfortunately, any information users posted on their profile can be harvested and used for unethical purposes due to Facebook's lack of application privacy configuration. In this paper we propose a privacy-management system for Facebook applications. The system can take advantage of the correlation between some profile features and network privacy settings, in this way it can automatically configure a users privacy settings. Our preliminary result show promising result.**

I. Introduction

Online social networking sites (SNSs) such as Facebook, MySpace, Orkut, Twitter, have become an integral part of our communication medium. SNSs have rich functionalities to attract all demographics, especially teenagers. According to December 2008 survey of Pew Internet & American Life Project, number of American adult internet users who joined in an online SNS were quadrupled in last four years – from 8% to 35%, and 75% of them were age range of 18 to 24 [1]. Online SNSs are more popular among teenagers than any other demographics – 65% of teenagers use one or more SNSs [1][2]. Online SNSs attract not only teenagers but also adults.

Users join in online SNSs to keep up with friends, organize events with friends, make new friends, or flirt. One of the main features of SNSs is "profile" where users post information about themselves. The profile can include real name, e-mail, physical address, phone number, academic classification, major, hometown, birthdates, sexual orientation, relationship status, interests, job history, favorite music/movies/books, etc [3]. Revealing personal information provides credibility (through identifying credentials) and suggests areas of compatibility between parties. Online SNS users post their profile for a number of reasons: 89% use it to keep up with friends, 57% use it make plans with friends, and 49% use it to make new friends [1]. How much and what kind of information is revealed depend on users' privacy concern and trust in SNS and its members. Facebook users are more willing to share personal information than MySpace users [4][3]. Sharing physical location can explain Facebook users trust in the network i.e, when users create a profile on Facebook network, they have to join in a network. The network can be based on a county, a region of a country, an academic and business institution[1]. Therefore, it is more likely that a Facebook user has met a person before accepting the person as a friend.

Third-party applications are essential part of SNSs. Online SNSs provide their platform with its users' social graph to the application developers[2]. The applications make SNS more engaging to users. For example, friends on Facebook can organize an event, play poker together, share photo, give electronic gift to each other, etc. 95% of 150 million Facebook users install at least one or more applications on their profiles. In addition, there are more than 660,000 developers and entrepreneurs and more than 52,000 available applications on Facebook[3]. To ensure quality and engagement, the Facebook platform provides users' personal (name, date of birth, hometown, current residency, political preference, religion, relationship status, interests, activities, movies, music listing, books, high school information, education history, etc) and social information (friends list, friendship connections, groups they joined, etc) to third-party developers. There is unfortunately no mechanism to ensure users privacy from malicious applications to harvest users' information.

Main contribution of this paper is to develop a privacy-management system (PMS) to protect Facebook users from third-party applications. PMS can also be used for profile privacy setting to eliminate opt-out privacy model. The system consists of sampling a network, building Bayesian Belief Network (BBN) to find relationship between profile features, and configuring the most suitable application-privacy setting.

II. Background

Collecting users' data such as behavior, interest, demographic is valuable for personalized systems and web applications. The data can be used to adapt to user needs. Even though personalized systems are beneficial to provide relevant contents, targeted emails, e-commerce, the Internet

[1] www.facebook.com/networks/networks.php
[2] developers.facebook.com/ & developer.myspace.com/community/
[3] www.facebook.com/press/info.php

users express significant concerns about their privacy [5]. In general context of the Internet, many researches and projects have been done to address protecting privacy. For example, Platform for Privacy Preferences (P3P) project is created by W3C for privacy standard [6] by having proposed user agents integrated in browsers check for compatibility of users' privacy preferences and website privacy policy. If there is a violation, users are notified. Due to lack of support from browser vendors, integrating P3P into application has unfortunately been slow. There have been several improvements made in the following researches [7][8][9].

Social Network Sites are the next big step toward invading users' privacy because users willingly post personal information either without considering the consequences or because they believe their information is protected [4]. Number of surveys have been done to find users' information revelation behavior. In 2007, Acquisti and Gross (Carnegie Mellon University) report their research into privacy concerns arising during Facebook social networking [10]. Their survey of 40 questions relating to Facebook privacy was taken by 506 respondents. Acquisti and Gross analyze survey-takers behavior on Facebook based on before and after learning information revealed on Facebook. They also pointed out misconception of members' profile visibility in their network based on the survey. In addition, a study on use of privacy settings shows that majority of Facebook users do not change their default privacy settings even though they can limit visibility of their profile information from strangers [11]. Large amount of personal information in easily harvestable environment (Facebook, MySpace, Orkut) can lead to social phishing attack [12]. Phishers can impersonate as victims' friend or use personal and social information.

Social networking has become increasingly popular among teenagers who can easily become victim of privacy invasion due to the lack of privacy concern [2][13]. So, members of this group reveals more information on their profile sites than older users. This lack of concern can lead to physical or online attacks. To address this issue, Susan B. Barnes proposes three approaches to solve this problem: social, technical, and legal [14]. Danah Boyd's study show four properties: persistence, search-ability, exact copy-ability, and invisible audiences that SN sites have, but conventional face-to-face interaction does not [2]. These properties have been changing the way people interact, especially for young people.

A. Facebook Privacy Issue

Social factors such as age, education level, and wealth can influence level of privacy concern [13]. Because developers can use a platform to access to users' personal and social information, users' privacy raises special concern. Facebook platform unfortunately comes with a privacy hole. The platform is available to anyone, and it provides developers with social and personal information except for contact information. Without knowing this vulnerability, many users reveal too much information. This can lead to invasion of their privacy. To find out how much information Facebook users reveal on their profile, we analyzed sample of 4,919 Facebook users on University of North Texas (UNT) network (UNT network has 37,800 registered members). Gender ratio is 35% female and 65% male. Our research shows that 75% of the users reveal their education history after high school; 70% disclose their high school's name; more than 60% post their favorite movies, music preferences, interests, relationship status, and age; 57% list books they like; between 45-51% reveal the city where they grew up, their favorite TV shows, activities, and political preference which can be one of libertarian, apathetic, very conservative, conservative, moderate, liberal, or very liberal; and, 21% disclose the current city where they currently reside.

We believe the most serious vulnerability of Facebook platform is that third-party application's source code is not reviewed or approved by Facebook. At front end, malicious applications may appear legitimate, but they can harvest users' information without users' approval. Figure 1 shows the harvesting process. Once user A installs a malicious application, the platform provides user A's personal and social information. The information includes the user's name, date of birth, religion, political preference, friends, groups, etc. Each user and group is identified by a unique ID (uid and gid). Once the harvesting program recovers these IDs, the application can easily collect the user A's friend information.

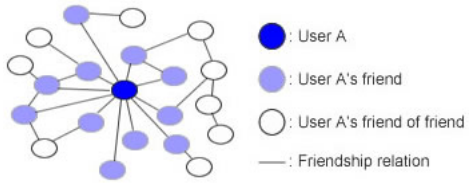

Fig. 1. Once a user installs a malicious application, any information associated with the user can be compromised, which consists of personal and social information.

III. Methodology

In this research study, we construct the privacy-management model using Bayesian Belief Network (BBN). Figure 2 shows the architecture of our model, using three components: profile information (PI), privacy manager (PM), and profile zoning (PZ). The model makes the decision (PM) on what profile information/feature (PZ) should be revealed and should not be revealed based on given profile information (PI).

Constructing BBN requires two steps. In step 1, we create the structure of the BBN using either TPDA or B-Course algorithm. TPDA algorithm has three phases: drafting, thickening and thinning. It is proved to be efficient in learning and doesn't need ordering of the nodes [15][16]. On the other hand, B-Course uses a combination of stochastic and greedy search heuristics to explore the very high-dimensional Model spaces [17]. The result of the B-Course algorithm is given in Figure 3. The next step is training the BBN, which creates the conditional probability tables for each arc in the BBN.

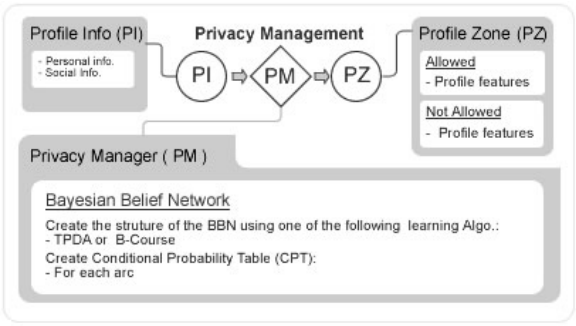

Fig. 2. Privacy Management model is based on three components: PI, PM, and PZ. PM configures user's profile information into allowed and not allowed to access categories.

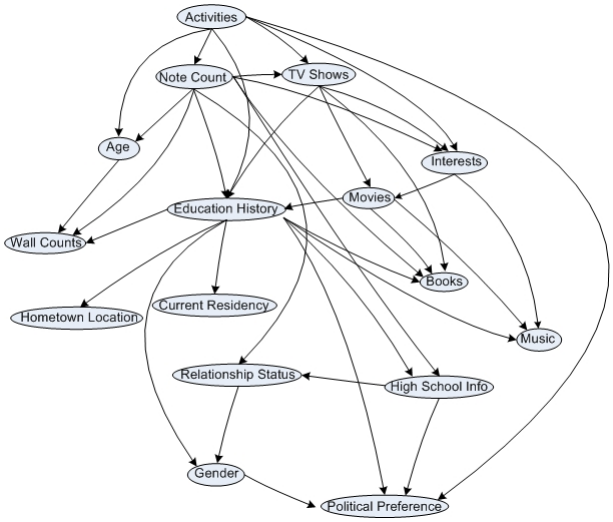

Fig. 3. Complete graph of the learned structure of B-Course

IV. Preliminary Results and Conclusion

In order to provide an indication of the accuracy of the proposed approach, a sample of ten queries were conducted to compute the probabilities for a given set of events. The results were compared with the actual values. Two examples of the queries are listed next: I) If age = 20, gender = male(1), relationship status = in a relationship(2), and political preference = moderate(3), what is the probability of revealing the rest of the features (hometown location, current residency, activities, interest, music, TV shows, movies, books, high school info, education history, note counts, and wall counts). The result of the query predicts that, for example, more than 37% of the users will not reveal their current residence; the actual value from the data set presents that 41% indeed did not reveal this feature for the specified group (90% accuracy). The query also indicates that, respectively 95.1% and 96.1% of the users for the query would have revealed their high school information and their education level; for the selected group all the users have revealed such information. The estimates for the remaining features as well as for the other queries present similar results. These results are a clear indication of a correlation between the exposures of some features for a given group of users and indicate the potential of the proposed approach. However, we are aware that a comprehensive evaluation of the proposed approach needs to be conducted. The next step is to compute the accuracy of the predictions by using all possible queries and comparing the probabilities with the actual values from the data set.

Acknowledgments

This work is supported by the National Science Foundation under grants CNS-0627754, CNS-0619871and CNS-0551694. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We would like to thank our colleague Santi Phithakkitnukoon and anonymous reviewers for their insightful and helpful comments and suggestions.

References

[1] A. Lenhart, "Adults and social network websites," PEW Interent & American Life Project, Tech. Rep., January 2009.

[2] D. Boyd, *Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life*. Cambridge, MA: Massachusetts Institute of Technology, 2008.

[3] F. Stutzman, "An evaluation of identity-sharing behavior in social network communities." *iDMAa Journal*, vol. 3, no. 1, 2006. [Online]. Available: http://www.ibiblio.org/fred/pubs/stutzman_pub4.pdf

[4] C. Dwyer, S. R. Hiltz, and K. Passerini, "Trust and privacy concern within social networking sites: A comparison of facebook and myspace," in *Proceedings of the Thirteenth Americas Conference on Information Systems (AMCIS)*, 2007. [Online]. Available: http://csis.pace.edu/ dwyer/research/DwyerAMCIS2007.pdf

[5] M. Teltzrow and A. Kobsa, "Impacts of user privacy preferences on personalized systems: a comparative study. in: Designing personalized user experiences for ecommerce." Kluwer Academic Publishers, 2004, pp. 315–332.

[6] L. Cranor, B. Dobbs, S. Egelman, and et al, "The platform for privacy preferences 1.1 (p3p1.1) specification," W3C, Tech. Rep. [Online]. Available: http://www.w3.org/TR/P3P11/

[7] S. Preibusch, B. Hoser, S. Gürses, and B. Berendt, "Ubiquitous social networks : Opportunities and challenges for privacy-aware user modelling," Tech. Rep., 2007.

[8] S. Preibusch, "Spontaneous privacy policy negotiations in pervasive environments," 2007, pp. 814–823. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-76890-6_7

[9] ——, "Privacy negotiations with p3p," in *W3C Workshop on Languages for Privacy Policy Negotiation and Semantics-Driven Enforcement*, 2006.

[10] A. Acquisti and R. Gross, "Imagined communities: Awareness, information sharing, and privacy on the facebook," in *Privacy Enhancing Technologies*, 2006, pp. 36–58.

[11] R. Gross, A. Acquisti, and H. J. Heinz, III, "Information revelation and privacy in online social networks," in *WPES '05: Proceedings of the 2005 ACM workshop on Privacy in the electronic society*. New York, NY, USA: ACM, 2005, pp. 71–80.

[12] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," *Commun. ACM*, vol. 50, no. 10, pp. 94–100, 2007.

[13] T. Zukowski and I. Brown, "Examining the influence of demographic factors on internet users' information privacy concerns," in *SAICSIT Conf.*, 2007, pp. 197–204.

[14] S. B. Barnes, "A privacy paradox: Social networking in the united states," *First Monday*, vol. 11, no. 9, August 2006.

[15] J. Cheng, D. A. Bell, and W. Liu, "Learning bayesian networks from data: an information-theory based approach," *Artificial Intelligence*, vol. 137, no. 1, pp. 43–90, 2002.

[16] C. K. Chow and C. N. Liu, "Approximating discrete probability distributions with dependence trees," vol. 14, pp. 462–467, 1968.

[17] P. Myllymki, T. Silander, H. Tirri, and P. Uronen, "B-course: A web-based tool for bayesian and causal data analysis," *International Journal on Artificial Intelligence Tools*, vol. 11, no. 3, pp. 369–387, 2002.