I J C T A, 9(41) 2016, pp. 1044-1052 ISSN: 0974-5572

A NOVEL APPROACH TO PREVENT INFORMATION LEAKAGE ON OSN BASED ON TRUST CLEARANCE

Divyani Joshi, Urjita Thakar and Vandan Tewari

Abstract: Online Social Networking (OSN) sites provide a platform to build social relationships amongst users. The wide adoption of OSNs raises concerns due to sharing of sensitive data online. A number of mechanisms have been employed in popular OSNs where users can control discretionary viewing of their personal information. However, this sensitive information could still be leaked even when security features are properly configured. Typically, users have friends in their friend list with whom generally information sharing is done. Some of these friends can be more trustworthy than others. The information to be shared with friends can be of different sensitivity levels - (i) 'Public' - shareable with all, (ii) 'Sensitive' - shareable with trusted friends, (iii) 'Highly Sensitive' – shareable with highly trusted friends only. Thus friends with whom information with different level of sensitivity is to be shared need to be assigned the corresponding level of trust clearance.The existing techniques to control information flow are inefficient as they do not consider the most important channel for information leakage, i.e. through sharing. A post may traverse to an untrusted user if a trusted user who is friend to an untrusted user shares that post with him. To check the trustworthiness of the users in an OSN, most of the existing approaches consider only interaction based attributes such as tagging, messaging etc. In the proposed approach, user indicates the sensitivity level of his information. Both sensitive and highly sensitive information are shared with users as per their trust clearance. Trustworthiness is calculated considering interaction based attributes such as 'Tagging' and 'messaging', and a non-interaction based attribute which is 'mutual friend list'. To evaluate the proposed approach, a social network has been created named as 'Friendsbook' (http://45.127.101.22/friendsbook/). It has been observed that by hiding the sensitive information from an untrusted user, the risk of information leakage is minimized to a great extent.

Key Words: Online Social Network, Social Graph, Information Leakage, Sensitive information, Trusted user

I. INTRODUCTION

Online Social Networks (OSNs) allow people to virtually connect with existing friends and make new friends. People can share contents like messages, images, videos etc. and interact with friends or create various communities of friends etc. At present about 82% of total population uses atleast any one of OSNs such as Facebook, Google+, Twitter, Linkedln etc. which facilitate building relationships, sharing personal experiences and information. Through OSNs, a large amount of personal data is published online which is accessed by users from all over the world. In OSN users can share information with the users in their friend list who can further share the information with users in their friend list. Thus there is risk of the information getting spread to the users with whom users do not want to share. In an OSN, certain privacy rules are predefined such as restricting access to personal information and checking for authorization. Yet, there is no restriction on sharing information received from other people. Any trusted user who has authority to view and share sensitive data can simply use this ability to inadvertently share that data with other untrusted users who are not supposed to access it. This untrusted user can share sensitive information further in the network, which results in information leakage. In an initial attempt to

Department of Computer Engineering, S.G.S.I.T.S,Indore , M.P., India
Emails- divyanijoshi25@gmail.com, uthakar@sgsits.ac.in, vandantewari@gmail.com

handle this information leakage problem, Facebook [2] implemented a new function to customize privacy for each user while sharing messages. Using this option, a user can choose a range of friends to share the information with and also to hide the message from specific users. Later on, Google+ [3] developed a concept of grouping users into circles such that a user can select a specific circle to share the information with whenever he starts to share a message. It appears superficially that the problem of information leakage has been overcome in Facebook and Google+ by tracking the message-ID to hide it from those users who are not permitted to share the data. However, Facebook and Google+ have neglected an important channel of information propagation, that is extended sharing of messages through trusted users. Past studies [8] have shown that people with weak interactions are far more liable for information leakage. Some methods that can be used to calculate interaction between the users are based on various parameters such as comment length [14], a user action [17] etc. The methodologies proposed so far are found insufficient in many ways to prevent information leakage e.g. they do not consider extended sharing as privacy breach and also many approaches consider only interaction based attributes to determine trustworthiness. Hence there is a need of proper solution which overcome the deficiencies found in the previous approaches. Therefore, in this paper, a mechanism is presented for prevention of information leakage upto two hops in a social network. Information is categorized as ‘sensitive’ or ‘highly sensitive’ apart from ‘normal’ and a trust value is calculated, for calculating trustworthiness of of a user. The users with trust value greater than a threshold value are said to have trust clearance. Depending on category of the information, it is shared with users with required trust clearance. An alert is generated whenever user shares his sensitive information with the untrusted user. Rest of the paper is organized as follows: In the next section background details are given. In section III related work is discussed. In section IV proposed approach is presented. Results and discussions are presented in section V and the paper is concluded in section VI.

II.BACKGROUND STUDY

In OSN, users can share information with either friends in his ‘friend list’ or publicly. Sometimes user shares some sensitive information about himself or others. This sensitive information can be such that through it a user can be harmed directly or indirectly. A study [16] showed that about 95.8% participants shared some sensitive information through their OSN accounts. Location Leakage is a special case of information leakage. Modern OSN services includes integration with mobile devices which encourages sharing of location information [16] with other users. Many times users intentionally share their private or sensitive information with other users. There are many incidents caused because of location leakage through OSN. It is possible that users unknowingly share, their location information, when on uploading media such as photos or videos, which may have geotagging enable. Facebook and Twitter are popular Online Social Networking sites. Their details are as given next.

A. Facebook Facebook [24] allows registered users to create and update their profiles, upload pictures and videos, send messages and keep in touch with friends, family and colleagues. A user may post some sensitive information on Facebook which can give rise to security threat. To protect the information, users are allowed to make some privacy settings to indicate which other users are allowed to view the information.

B. Twitter Twitter [25] is another free social networking micro blogging service that allows registered users to broadcast short messages called tweets. Twitter members can also follow or retweet tweets of other twitter user. The default setting of tweets is to enable public viewing of all messages. Unlike Facebook or LinkedIn, where members needs approval for social connections, in twitter any user can follow any other. Also Twitter is open for users of any age. In twitter, a user can protect his tweets by allowing only some people to see the tweets. Next section describes the security features provided by OSNs to protect user's information.

C. Security Features Provided by OSN In order to protect the privacy of user, modern OSN provides some level of privacy control mechanism that allows user to control "Who can view what in his profile".

- Profile Privacy: It controls the user's profile privacy, i.e. who can view the profile or the consisting personal information. Users can select different access control for each personal information.
- Application Privacy: It controls what information can be provided to the installed application by the user. Some application may ask the user to post some information on its behalf. It is upto the user to allow or deny the request.
- News Feed Privacy: It controls the news feed published on the user's timeline.
- Search Privacy: It controls whether the other user can search a user on OSN and how a person can contact him. It is the feature provided by the Facebook.

Next, the research work related to information sharing on OSNs is presented.

III. RELATED WORK

In this section, work carried out by earlier researchers pertaining to protection of privacy and information leakage related challenges and general security issues in OSNs have been presented.

Nilothpal et al [1] proposed a privacy protection tool called Privometer. Privometer measures amount of sensitive information leakage based on relationship information by accessing private information from friends. It assumes that malicious applications run an inference algorithm to get user's sensitive information. It hides the user's information which matches with any sensitive attribute of his/her friend. The main drawback of this approach is that it prevents the leakage of only mapped attributes; the unmapped attributes still remain insecure. Ahmad et al [10] proposed a mechanism to improve the privacy on OSNs which uses friendship intensity to calculate the trust level of the user determined through data mining technique. Though they have considered interactive metrics to calculate friendship intensity but non-interactive metrics have been ignored. This data mining model is validated on synthetic data which may fail for real Facebook dataset. Huina et al [3] characterized the nature of privacy leaks in twitter OSN. They built automatic classifiers to identify the users who leak the information and their method of information leakage. They could only identify the information leaks but could not give any prevention measure.

An information model which defines practical implementation of existing privacy settings has been proposed by Nigusse et al [17]. It takes care of users personal information such as name, address etc. without considering the issue of information sharing and posting. Multi-hop diffusion of information is

also not considered in the proposal. Safebook [22] is a Social Network formed to tackle the security and privacy problem. It is based on decentralized architecture. The privacy of user data stored and managed centrally is prone to access by malicious service providers. To mitigate the risk decentralized architecture has been proposed. However, itresults in to communication delays. Lam et al [19] showed that, in existing OSNs, by examining the public interaction of user with his friends, a malicious user can infer the personal information, no matter how tight privacy settings are.

Zhang et al [23] proposed evaluation method for trustworthiness of the user for information sharing based on comment length, time difference in comment as a indicator of interaction quality. Among these, the comment length was used to determines trustworthiness of the user. Assumption was that a person giving long message can be trusted more. Podobnik et al [9] proposed a trust calculation method based on the tags, likes, comments on the post of the given user. The author also proposed a method named as 'ego user' method to find 10 closest friends of a user in OSN. An OSN user usually interacts with 15% of the friends in the friend list on average [8]. Hence noninteraction based attributes should also be considered.

There are various security issues which cannot be resolved even by enforcing all the security features described in previous section. Some important issues are as follows:

1. One of the privacy problems is that OSN doesn't usually warn the users about risk of disclosing personal information.
2. Privacy tools provided by most of the OSNs fail to protect user's data as they are inflexible. They either make the users profile public(visible to whole audience) or private(available to only friends). Facebook is among the few OSNs which provides more finer level of privacy. However, the interface provided by the Facebook is too complex for most of the normal users.
3. OSN users can only control access to their own profile but they have no control over what others post about them. Hence it is possible that user is not aware of the information posted about him.

In the next section, the method proposed has been discussed.

IV. PROPOSED METHOD

In the proposed method the trust value of each user is calculated and based on the trust clearance, information of different sensitivity levels is shared. Following are the steps to enable information sharing only with trustworthy users:

- Data Preprocessing: In this step, the OSN dataset is processed to extract the values of required attributes.
- Trust Calculation: The trust value of each user is calculated using 'ego user' method [9] applied on the fetched user attribute values.
- Finding Trust Clearance: Trust clearance of different users is determined and trusted users are identified.
- User Elimination: List of identified trusted users is presented to the user. It is also indicated whether a trusted friend has other untrusted friends in his friend list. The user has option of eliminating such friends from the list of trusted users with whom the information can be shared.

Detailed description of each of the steps is given next.

A. Data Preprocessing: In this step required attributes are extracted from the dataset. Following attributes are selected for the trust value calculation:

1) Mutual Friends: This attribute corresponds to common friends of the users between whom information is to be shared. Large number of common friends indicates that individuals are either strongly connected with each other, or they have the same context. The significance of mutual friends as a factor of rustworthiness is highlighted in a survey [10] where 46% of individuals add strangers in their network only if they have mutual friends.

2) Tag: When information is tagged, a link is created tothat users profile. According to a survey [10], among the various ways to interact with friend, 43% of the users preferred tag as a method of interaction with friend.

3) Messages: Messages are chat messages which are sent to a particular person. According to a survey [20], most users prefer chat as their means to interact with their close friends. The values of selected attributes are used for trust calculation as elaborated next.

B. Trust Calculation: Trust value of each user is calculated to determine whether a user can be trusted for sharing the information or not. Following is the formula used for this purpose.

$$\text{Trust Value } t = M_n * W_1 + T_n * W_2 + MF_n * W_3 \quad \ldots\ldots(1)$$

Where M_n = Number of message sent to a particular user, T_n = Number of times the particular user is tagged, MF_n = Number of mutual friends. W_1 is the weight given to messages, W_2 is the weight given to tag, W_3 is the weight given to mutual friends.

Weights are assigned for the attributes by using ego user method. In this method, ego user is the user whose trust value is maximum. Weight assigned as follows:-

1) Calculate the value of trust for every user as per the following equation.
   $$\text{Trust Value } TV = M_n + T_n + MF_n \quad \ldots\ldots(2)$$
2) Find the user with maximum trust value; such a user is ego user.
3) Next, the values of weights are calculated based on the values for each attribute of the ego user. This is done as given below:

$$W_1 = \frac{M_e}{TV_e} * 100 \quad \ldots\ldots(3)$$

Where, M_e = Number of maximum number of messages sent to a particular user by ego user, TVe is the trust value of ego user.

$$W_2 = \frac{T_e}{TV_e} * 100 \quad \ldots\ldots(4)$$

Te = Maximum number of time the particular user is tagged by the ego user.

$$W_3 = \frac{MF_e}{TV_e} * 100 \quad \ldots\ldots(5)$$

MFe = Maximum number of Mutual Friends ego user can have with a particular user.

Trust value is calculated for every user using equation 1.

C. Finding Trust Clearance: Since information to be shared by the user has different levels of sensitivity, the information has been categorized into two sensitivity levels Sensitive and Highly Sensitive as decided

by the user himself. Different threshold trust values for sensitive and highly sensitive information are calculated. Users with trust value higher than the threshold, have the required trust clearance.

• Threshold Calculation for Sensitive Information For receiving sensitive information threshold is calculated by taking average of trust values of all the users. The users with trust value greater than Ths are considered as trusted, thus eligible to receive information of category 'sensitive'.

$$Th_s = \frac{\sum Trust\ Value}{n_a} \quad \text{.....(6)}$$

where na is the number friends a user have.

• Threshold Calculation for Highly Sensitive information: For receiving highly sensitive information, the threshold calculation is done considering the average of trust value of users who had trust clearance to receive sensitive information.

The equation given below is used for calculation of threshold for highly sensitive information.

$$Th_h = \frac{\sum Trust\ Value}{n_t} \quad \text{.....(7)}$$ where, Trust Value is trust value of users eligible to receive sensitive information, nt is the number trusted friends a user have. The Highly trusted users identified like this are subset of trusted users having high trustworthiness. These trusted and highly trusted users are least likely to leak the information. To prevent information leakage following step is followed:

D. User Elimination: If a trusted user unintentionally shares this trusted post to his own profile, and in case, the trusted user has an untrusted user in his friend list then it become visible to untrusted users. To prevent information leakage through such indirect links, the trusted user who have connection with untrusted users are to be eliminated. The user is prompted by showing a popup which shows the names of trusted users who have connection with untrusted users. It allows the user to eliminate those trusted users with whom the given information is not to be shared. The information is shared with remaining trusted users.

V. EXPERIMENT AND RESULT

For the realization of proposed approach a prototype an online social networking site "Friendsbook" (http://45.127.101.22/friendsbook/) with features similar to facebook have been developed. 106 users were registered on this OSN. To determine the effectiveness of proposed approach, some test cases are discussed next. • Test Case 1 The user selects the information to be shared with 'trusted' user having option such as 'public', 'only me' and 'trusted' as shown in Fig 1 .

Fig. 1 Posting of an information to 'Trusted'

OSN next asks user to enter the sensitivity level of the post. Here the user selects 'sensitive' as sensitivity level of information. Next, the user clicks on post to share the image. Since the post is 'sensitive', it is to be shared with the trusted users.

Fig. 2: Selecting Sensitivity level of Post

To identify trusted users, trust value is calculated. Table I shows friends of 'K J' along with their trust values which have been calculated as per formula given in equation 1 of section IV

TABLE I: Friends of K J and their Trust values

Username	Trust Value
NM	1000
PK	533.33
DJ	500
NJ	131.33
SKP	566.66
AW	566.67
KD	200
AS	666.67
KP	566.67
BK	266.67
BL	133.33
Gk	666.67
SM	566.67
PrK	200
NiM	533.33
SM	700
VA	566.67
ZK	533.33
SS	533.33

Threshold value is calculated using equation 6 which evaluates to 498.72. Based on this threshold value, following is the list of trusted users from which the user ‘K J’ can either select some or all trusted friends to receive the information are:

TABLE II. Trusted Friends of ‘K J’

Username
NM
PK
DJ
SKP
AW
AS
GK
SM
NiM
VA
ZK
SS
KP
SM

The test case that shows sharing of Highly Sensitive information is presented next.

Test Case: To post ‘highly sensitive’ information, user has to select ‘highly sensitive’ as the sensitivity level of the post as shown in Fig 2. For the user ‘Khushboo Jain’ threshold value is calculated using equation 6 of section IV. The calculated threshold value is 607.14.

TABLE III: Highly Trusted Friends of user Khushboo Jain

Username
NM
AS
SM
GK

The information categorized as ‘highly sensitive’ is shared with these users listed in Table III. It has been observed from testing that the proposed approach reduces the risk of information leakage for both direct and indirect links for the information of each sensitivity level. In case 1, it is observed that sensitive information is shared with only trusted users, while highly sensitive information is shared with highly trusted users only.

IV. CONCLUSION

In this paper, an important issue of prevention of leakage of information through direct and indirect links in an OSN has been discussed. A method has been proposed that enables the user to determine trustworthy users with whom sensitive information can be shared with least risk of leakage of such information. The user is also able to categorize the information based on its sensitivity. From the test cases, it has been shown that the sensitive information can only be shared with adequately trusted friends identified on their profile attributes. Inclusion of this approach in OSNs will increase user’s faith and will encourage more users to register on OSNs. In future, soft computing techniques can be applied to determine trustworthy users.

REFERENCES

[1] Nilothpal Talukder, Mourad Ouzzani, Ahmed K. Elmagarmid, Hazem Elmeleegy, and Mohamed Yakou, "Privometer: Privacy protection in social networks",Data Engineering Workshops(ICDEW), IEEE 26th International Conference, pp. 266-269, 2010.

[2] Yilin Shen, Yu-Song Syu, Dung T. Nguyen, My T. Thai, "Maximizing Circle of Trust in Online Social Networks",ACM, pp. 155-164, 2012.

[3] Huina Mao, Xin Shuai, Apu Kapadia, "Loose Tweets: An Analysis of Privacy Leaks on Twitter", ACM, pp. 111, 2010.

[4] M. Granovetter. "The Strength of Weak Ties". American Journal of Sociology, 1973. [5] Amin Ranjbar, Muthucumaru Maheshwaran, "Using community structure to control information sharing in online social networks",Computer Communications 41, pp. 11-21, 2014.

[6] XI Chen, Katina Michael, "Privacy Issues and Solutions in Social Network Sites",IEEE Technology and Society Magazine,IEEE, pp. 43-53, 2012.

[7] Samah Al-Oufi, Heung-Nam Kim , Abdulmotaleb El Saddik, "A group trust metric for identifying people of trust in online social networks" , Expert Systems with Applications pp 13173–13181, 2012.

[8] Lerone Banks, Shyhtsun Felix Wu, "All Friends are NOT Created Equal: An Interaction Intensity based Approach to Privacy in Online Social Networks", Computational Science and Engineering, CSE '09. IEEE Conference, PP 970 – 974, 2009.

[9] Podobnik, Vedran; Striga, Darko; Jandras, Ana; Lovrek, Ignac, "How to Calculate Trust between Social Network Users?," Proceedings of the 20th International Conference on Software, Telecommunications and Computer Networks SoftCOM, 2012.

[10] Waqar Ahmad, Asim Riaz, Henric Johnson, Niklas Lavesson, "Predicting Friendship Intensity In Online Social Networks", Tyrrhenian Workshop on Information Technology, 2010.

[11] Imrul Kayes, Adriana Iamnitchi, "A Survey on Privacy and Security in Online Social Networks", ACM Computing Surveys, 2015.

[12] Balachander Krishnamurthy, Craig E. Wills, "On the Leakage of Personally Identifiable Information Via Online Social Networks" Proceedings of the 2nd ACM workshop on Online social networks , Pages 7-12, 2009.

[13] Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, Daniel Starin, "Persona: An Online Social Network with User-Defined Privacy" Proceedings of the ACM SIGCOMM conference on Data communication Pages 135-146, 2009.

[14] Y Li, Y Li, Q Yan, RH Deng, "Privacy Leakage Analysis in Online Social Networks" Computers & Security, Elsevier, 2015.

[15] I.F.Lam, K.T.Chen, L.J.Chen, "Involuntary Information Leakage in Social Network Services". IWSEC , pages 167-183, Berlin, Heidelberg, 2008.

[16] T. H. Ngoc, I. Echizen, K. Komei, and H. Yoshiura, "New Approach to Quantification of Privacy on Social Network Sites". AINA , pages 556-564, Washington, DC, USA, IEEE Computer Society, 2010.

[17] G. Nigusse and B. D. Decker. "Privacy Codes of Practice for the Social Web: The Analysis of Existing Privacy Codes and Emerging Social-Centric Privacy Risks". In AAAI Spring Symposium Series, 2010.

[18] Abdul Molok, Nurul Nuha, Ahmad, Atif, and Chang, Shanton, "Understanding the Factors of Information Leakage through Online Social Networking to Safeguard Organizational Information" ACIS Proceedings. Paper 62, 2010.

[19] Ieng Fat Lam, Kuan Ta Chen, and Ling Jyh Chen, "Involuntary Information Leakage in Social Network Services," In Proceedings of IWSEC 2008.

[20] Paul, T., Stopczynski, M., Puscher, D., Volkamer, M., and Strufe, T., "C4ps helping facebookers manage their privacy settings". In Social Informatics, pages 188–201, 2012.

[21] Giunchiglia, F., Zhang, R., and Crispo, B., "Relbac: Relation based Access Control". In Fourth International Conference on Semantics, Knowledge and Grid, pages 3 –11, 2008.

[22] "Safebook: A Privacy-Preserving Online Social Network Leveraging on Real-Life Trust", IEEE Communications Magazine, December 2009.

[23] Lizi Zhang, Cheun Pin Tan, Siyi Li, Hui Fang, Pramodh Rai, Yao Chen, Rohit Luthra, Wee Keong Ng, and Jie Zhang, "The Influence of Interaction Attributes on Trust in Virtual Communities". UMAP Workshops, LNCS 7138, pp. 268 –279, 2012.

[24] http://www.facebook.com.

[25] https://twitter.com/?lang=en.