5. Privacy in social Network

A. Privacy Defined

There is a need to define a perfect privacy definition for the safety of user's profile from the inference attack. Privacy is/was defined by each individual's word. Various privacy definitions can be put forward. An attacker can formulate attacks by various possible methods. It can be a passport number, voter id etc. The attacker can acquire background knowledge by various methods. Thus we need to hide the sensitive data and decrease the classification accuracy and probability while preserving the essence of social networking.

Any set of classifiers, *C*, the classification accurateness of any random classifier *c' € C* when trained on *K* and this is used to classify data set *G* to predict sensitive data $P_{c(K)}^{j}$.$P_{c(G,K)}$ denotes the prediction accuracy .Another identifier is used to identify the additional accuracy of the attacker. The possible attacks would be predicting the death rate and predicting disease types by analyzing the revealed private details along with the public details.

B. Perturbation and Anonymization

For protecting against inference attacks on online social networks, we need to modify and remove certain parameters. The details residing on social network data can be deployed in three ways: adding details, modifying details and removing details from nodes. These methods are called Perturbation and Anonymization.Introducing noise into *D* to decrease classification can be considered as perturbation. Removing nodes is considered as anonymization. Here consider two graphs which are sanitized versions of the original graph. When artificial details are added to the data set, the accuracy is minimized and the sanitized versions of the graphs cannot reveal the favorite activity. So the classification accuracy is decreased. In short the details should be removed to decrease the classification accuracy on sensitive attributes.

C. Details on the Social Network

The main component of online social network is the details that are linked to each individual. We need to choose which details are to be removed. Suppose a social network data set *G*and a list of sensitive details *I*,here we need to determine which of the details should be removed that helps to decrease the classification accuracy. Consider that a user has the class value C_2out of all the set of classes of C and this person has the public details D_i

$$arg\max_{1\le x\le p}\left(P(C_x^i) \times P(D_i^1|C_x^i) \dots P(D_i^m|C_x^i)\right.$$

We are removing the most correlated details with that of private. When we remove these details the building up of a classifier to predict the sensitive details accuracy is decreased.

D. Links on Social Network

Another method to decrease the classification is the anonymization technique. This process of manipulating link information.Anonymization is technique left with only two choices adding or removing links. We are considering the case of removing friendship links. Suppose a user belongs to two classes of friendship links and the true link is C_1.This technique helps to remove links in C_1 to decrease the classification accuracy. A node to be in class C_2 if the below equation is positive.

$$B(i)=p(C_2,N_i)-p(C_1,N_i)$$

We need to maximize the value of B(i) by removing links.

E. Detail Generalization Hierarchy and Detail Value Decomposition

Another important feature of detail anonymization is the implementation of detail generalization hierarchy (DGH) which is the process of generating a hierarchical ordering of the details within a given category. It is represented as a tree structure. For example a user likeskathakali.We can replace it with art forms .Furthermore we can also specify and replace it with the traditional one.

Another technique is the detail value decomposition in which it is a process of dividing an attribute into a series of representative tags. At each step we generalize each detail type.

6. Security Concerns For online Social Network

A. Two-Step Verification and Authentication

The two factor verification and authentication consists of three steps verifying username and password authenticate security token, if a user doesn't have a security token: logs user in right after he passes Step 1 only. Most of the online social networks fail to keep trust on the users. If an attacker uses a brute force attack and he can take the password of the user. The relevance of two factor authenticationcomes in, Two-factor authentication is a security process in which the user provides two means of identification, one of which is typically a physical token, such as a card, and the other one is typically something memorized, such as a security code.According to proponents, two-factor authentication could drastically reduce the incidence of online identity theft, phishing expeditions, and other online fraud, because the victim's password would no longer be enough to give athiefthe access to their information.

B. Blow Fish Cipher Encryption

Blowfish is a fast block cipher, except when changing keys. Each new key requires pre-processing equivalent to encrypting about 4 kilobytes of text, which is very slow compared to other block ciphers. This prevents its use in certain applications, but is not a problem in others. In one application Blowfish's slow key changing is actually a benefit: the password-hashing method used in OpenBSD uses an algorithm which is derived from Blowfish that makes use of the slow key schedule; the idea is that the extra computational effort required gives protection against dictionary attacks. Blowfish has a memory footprint of just over 4 kilobytes of RAM. This constraint is not a problem even for older desktop and laptop computers, although usage in the smallest embedded systems such as early smartcards is prevented. Blowfish was one of the first secure block ciphers which is not subject to any patents and therefore is freely available for anyone to use. This benefit has contributed to its popularity in cryptographic software. Users' uploaded

Licensed Under Creative Commons Attribution CC BY

images are encrypted with this algorithm and stored with the server. So pictures are securely stored in the server. Blowfish has a 64-bit block size and a variable key length from 32 bits up to 448 bits. It is a 16-round Feistel cipher and uses large key-dependent S-boxes .In blowfish algorithm a 64-bit plaintext message is first divided into 32 bits. Each line represents 32 bits. The algorithm keeps two sub key arrays: the 18-entry P-array and four 256-entry S-boxes. The S-boxes accept 8-bit input and produce 32-bit output. One entry of the P-array is used every round, and after the final round, each half of the data block is XORed with one of the two remaining unused P-entries.

C. Filter Unwanted Messages

This is the first proposal of a system to automatically filter unwanted messages from OSN user walls on the basis of both the message content and the message creator relationships and characteristics.

D. GROUP Priority

We assign priority to each group created by the user, so each group has a unique priority with one another. This priority helps to view the posts and allows them to appear on the user walls.

E. Trusted Friends

We will be assigning our friends as trusted entity and this will help to recover the account whenever it is necessary. Only these friends can help when the account is hacked.

7. Conclusion

Online social networking is one of the emerging trends in today's world. Most of the people are involved in surfing the OSN, but most of them are unaware of the attacks from social networks. People will usually publish their personal details in these sites and this will eventually cause privacy issues. Social networking also becomes the marketing media. When companies rely on social networking site, they have various intentions. Users need to take care while publishing the details inside the social network. Security features of these sites should be improved in order to protect from various attacks.

References

[1] Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham, "Preventing Private Information Inference Attacks on Social Networks," IEEE Trans. Knowledge and Data Engineering, vol. 25, no. 8, Aug 2013, pp.1849-1861.
[2] M. Hay, G. Miklau, D. Jensen, P. Weis, and S. Srivastava, "Anonymizing Social Networks," Technical Report 07-19, Univ. of Massachusetts Amherst, 2007.
[3] J. He, W. Chu, and V. Liu, "Inferring Privacy Information from Social Networks," Proc. Intelligence and Security Informatics, 2006.
[4] L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore Art Thou r3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography," Proc. 16th Int'l Conf. World Wide Web (WWW '07), pp. 181-190, 2007
[5] K. Liu and E. Terzi, "Towards Identity Anonymization on Graphs," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '08), pp. 93-106, 2008.
[6] S.A. Macskassy and F. Provost, "Classification in Networked Data: A Toolkit and a Univariate Case Study," J. Machine Learning Research, vol. 8, pp. 935-983, 2007.
[7] C. Zhang, J. Sun, X. Zhu, and Y. Fang," Privacy and Security for Online Social Networks: Challenges and Opportunities",‖ *IEEE Netw.*, vol. 24, no. 4, pp. 13–18,Jul./Aug. 2010.
[8] L. A. Cutillo and R.Molva, ―Safebook: A privacy preserving online social network leveraging on real-life trust,‖*IEEE Commun. Mag.*, vol. 47, no. 12, pp. 94–101, Dec. 2009.

Licensed Under Creative Commons Attribution CC BY