A Review on Privacy Priserving Data Publishing of Social Network

Amolika N. Patil Dr. S.P. Deshpande

Abstract- **Nowadays online social network is very popular and its data are increasingly made publicly available to third parties . Sharing social network data in its raw form raises serious privacy concerns because a successful privacy attack not only compromises the sensitive information of the target victim but also the relationship with his/her friends or even their private information. So security protection of private information online has been a serious and important research topic. In recent years several anonymization techniques have been proposed to solve these issues but these are not able to fully satisfied the privacy issues. In this paper privacy of social network sites has been investigated and reviewed.**

Keywords- **privacy-protection, social network, preservation, disclosure, PPDP .**

I. INTRODUCTION

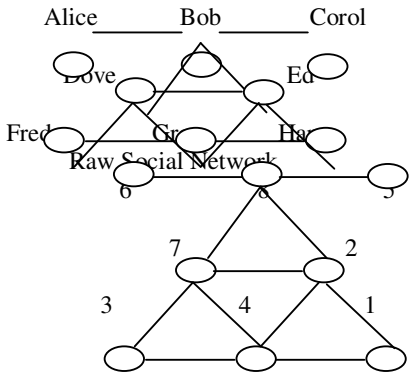
Recently social networking sites such as Facebook, Twitter, Linkld and etc have gained large popularity. Participating users of these sites form online social network, which provides sharing, organizing and finding contents and contacts.

The relation between privacy and use of social network sites is very close and delicate. Commonly, people would like personal information to be known by the small group of close friends, families and not by strangers or outsiders. In some cases users information disclosure can be helpful to other users, companies and third parties. Private information is very valuable when the information of many people gathered on social network sites. The popularity of online social network application increases serious problems about the security and privacy of their users.

Privacy associated with online social networking depend on the level of identification of the information provided to its recipients and its users. Even social networking sites that do not openly expose their users identities may provide enough information to identify the profile owner. So there is need to protect owners profile and sensitive information [1][2][12].

Online Social Network:

Online social networks are organized around users. By means of functionalities provided by these networks, users are allowed to share, organize and find content and contact very easily [3][4].



Anonymized Social Network

Many previous works in privacy preserving data publishing removing explicit identifiers is insufficient because an adversary may utilize some external knowledge to identify an individual from the data.[4][12]

Privacy Preserving Data Publishing:

A privacy principal is a set of security requirement to prevent sensitive data linkage of some attacks. As the name suggests, "Privacy preserving data publishing"(PPDP) tries to publish privacy preserved data. Given some security requirements and utility or quality of the anonymized data. The process to achieve a given privacy principle is called anonymization [5].

Definition : Privacy protection

"Access to the published data should not enable the attacker to learn anything extra about any target victim compared to no access to the database, given that the attacker has only a limited amount of background knowledge."[6]

Privacy can be divided into three categories:[7]

1) Identity disclosure: Identity of an individual who is associated with a node is disclosed.
2) Link disclosure: The sensitive relationship between two individuals are disclosed.
3) Content disclosure: The sensitive data associated with each node is compromised.

EXISTING WORK

According to literature survey background knowledge is the piece of information that is known to the adversary and can used by the adversary to infer the privacy of an individual. In social network data, the information that can be used as background knowledge to intrude into user privacy are personal attributes and structural attributes. With these two, an adversary may conduct different types of attacks against social network privacy. Therefore,

background knowledge plays an important role in modeling privacy attacks on social network data.

Categories of Privacy:

Identity Disclosure:

Identity preserving model deals with protecting individual identity from being re-identified. Formally, the problem can be defined as: Given a published social network data, if an adversary can identify the vertex of a target individual by analyzing topological features of the vertex based on his background knowledge about the individual from the social network, then the identity of target individual is disclosed. Identity disclosure occurs when an individual behind a record is exposed. This type of breach leads to the revelation of information of a user and relationship he/she shares with other individuals in the network.[7][8][10]

Link Disclosure:

The link disclosure problem is centered around the protection of the connection between vertices in a network. [7] Link disclosure occurs when sensitive link structure information is leaked as a result of social network data publication, or inferred by compromised social network users. Inferring link structure from anonymized data, the social network owner wants to publish the social network to untrusted recipients for analysis purposes in a way that sensitive relationships between users cannot be inferred from the published data. [8] Link disclosure occurs when the associations between two individuals are revealed. Social activities generate this type of information when social media services are utilized by users.[9][10]

Contect Disclosure:

Content disclosure is normally an issue when the private data associated with a user on the network is disclosed to others. Content disclosure takes place when an attacker obtains the information of a sensitive and confidential user attribute. Sensitive attributes may be linked with an entity and link relationship. The adversary use any of the structural background knowledge to identify the sensitive value of an individual if cluster of vertices anonymized together share same sensitive information[7][8][9][10].

CHALLENGES IN PRIVACY PRESERVING DATA PUBLISHING OF SOCIAL NETWORK [7][8][11]

1) Modeling of background knowledge of adversaries is difficult in social network data. Because in social network information from various sources such as labels of vertices and edges, subgraphs, and neighborhood graphs can be used to identify individuals.

2) Information loss is the metric which measures the amount of distortion. A social network is a graphical structure with a set of vertices and edges hence it is difficult to compare two social networks by comparing the vertices and edges individually. Anonymized social network and original social networks which have the same n number of vertices and edges may have very different properties like betweenness, connectivity, and diameter. Information loss and anonymization quality can be measured in different ways.

3) It is even difficult to devise graph-modification algorithms that balance the goals of preserving privacy with the utility of the data. Because the nodes and edges in a graph are all correlated. Therefore, the impact of a single change of an edge or a node can spread across the whole network.

Going through the literature survey it is observe that the process to achieve a given privacy principal is called anonymization . Existing operations of anonymization include generalization, perturbation, suppression, randomization , slicing , etc are the most widely used techniques. But not a only technique is efficient to satisfy all the three categories i.e identity disclosure, link disclosure and content disclosure. That means some techniques protect against identity disclosure attack, some protect against link disclosure attack and some of them protect against content disclosure attack [2][4][6].

To overcome these challenges, several researchers have recently proposed different types of privacy models, adversaries, and graph modification algorithms. Unfortunately, none of the work is linked to solve all the problems in one shot. Protecting against each kind of privacy disclosures may require different methods or combination of them.

CONCLUSION

In this paper we discuss PPDP Attacks and their categories . Another section is of challenges in PPDP. Existing anonimization techniques are not able to satisfy all the three privacy disclosure attacks so there is need to develop a model which is helpful to the data owner to protect their information at the time of publishing before the receipt receive it. In proposed research profile of the data owner is taken to be consider . Because when a person registered on social network his/her profile was stored on social network and service providers publish this data to third parties . At that time introducer make target to the victim and misuse of information is occure.

REFERENCES

1. Gross, R, and A. Acquisti. "Information Revelation and Privacy in Online Social Networks", the 2005 ACM Workshop on Privacy in the Electronic Society. Pp. 71-80,2005.
2. Xi Chen, Shuo Shi. "A Literature Review of Privacy Research on Social Network Sites", 2009 IEEE international Conference on Multimedia Information Networking and Security, pp. 93-97, 2009.
3. M. Hay, G. Miklau, D. Jensen, P. Weis and S. Shrivastava, "Anonymizing Social Network", Computer Science Department, University of Massachusetts Amherst, Tech. Rep. pp. 07-19, 2007.
4. Benjamin C. M. Fung, Yan'an Jin and Jiaming Li, "Preserving Privacy and Frequent Sharing Patterns for Social Network Data Publishing", 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp.479-485, 2013.
5. Qingming Tang, Yinjie Wu, Shangbin Liao and Xiaodong Wang, "Improving Strict Partition for Privacy Preserving Data

Publishing", 2010 IEEE First international Conference On Networking and Distributed Computing" pp. 207-212, 2010.

6. Andrei Manta, "Literature Survey on Privacy Preservation Mechanism in Data Publishing" Master of Science Perort, pp.1-62, Nov-2013.
7. Sri Krishna Adusumalli, Valli Kumari Vatsavayi, Jyothi Vadisala, " A Study of Privacy Attack on Social Network Data", JGRCS, vol. 5, No.7, 2014.
8. Kun Liu, Kamalika Das, Tyrone Grandison, Hillol Kargupta, "Privacy Preserving Data Analysis on Graphs and Social Network" Tech. rep., pages 1-22.
9. V.Vijayalakshmi, A.S.Arunachalam, R.Nandhakumar, "Mining Social Media-Utility based Privacy", IJCSIT, vol 5(4), pp. 5480-5485, 2014.
10. Amardeep Singh, Divya Bansal, Sanjeev Sofat, "Privacy Preserving Techniques in Social Network Data Publishing – A Review", International Journal of Computer Application, vol. 87,no. 15, 2014.
11. Amin Milani Fard, Ke Wang, Philip S. Yu, "Limiting Link Disclosure in Social Network Analysis through Subgraph-Wise Perturbation, 2012 ACM.
12. Xuan Ding, Lan Zhang, Zhiguo Wan and Ming Gu, "A Brief Survey on De-anonymization Attacks in Online Social Networks" 2010 IEEE, pp 611-615.