International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) e-ISSN: 2395-5325
Volume3, Issue 1 (January'2017)

A COMPREHENSIVE STUDY OF PRESERVING THE DATA PRIVACY IN ONLINE SOCIAL NETWORKS

Prathima.G,
Assistant Professor,
Department of Computer Science,
Global Academy of Technology, Bangalore, India

Abstract: Publishing data on a social networking site has become a part of online social activity .Social networking offers an important means for users to converse, interact and share information. Due to rapid growth of social network, data has been publicly available across in one way or another. Preserving the data Privacy of social network data has becomes a more and more important Factor. In the present paper, we focus on a brief review of the existing anonymization techniques for privacy preserving of social network data. We also try to identify the new challenges in Data privacy preserving of social network data.

Keywords: Data Publishing, Online Social Networks, Data Privacy

I. INTRODUCTION

Online social networks with the large number of people subscribe to social networks or social media has generated large amount of user data that is gathered and maintained by the social network service providers [1]. The data generated by social network services is termed as the social network data that needs to be published for others in certain situations. One of the situations is when specific analysis of the user data needs to be done and another situation is when the owner of the data has to share the data with third parties like advertising partners which is part of policies generally accepted by subscribers[1]. The data contains valuable information about users that helps third parties in better social targeting of advertisements. Social network analysis is being used in modern sociology, geography, economics, and information sciences[1]. Researchers in various fields use this data for different purposes like researchers in government institutions require social network data for information and security purposes. So, data needs to be shared or published in all above mentioned situations. Owner of data can publish it for others to analyse but it may create serious privacy threats[1]. To fulfill the demands for the network data, online social media operators have been sharing the data they gather and maintain with external third parties such as advertisers, application developers, and academic researchers like Facebook has thousands of third-party applications and there has been an exponential increase in this number.

Social network data contains sensitive information about the users. Thus sharing of this data in its raw form may breach privacy of individuals. Individual privacy is defined as "the right of the individual to decide what information about himself should be communicated to others and under what circumstances"[1]. A privacy breach occurs when private information about the user is disclosed to an adversary. So, preserving privacy of individuals while publishing user's collected data is the need of the day.

1.1. CATEGORIES OF PRIVACY BREACH: The privacy breaches in social networks can be categorized into three types:

Identity disclosure - 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.

Sensitive link disclosure - Sensitive 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.

Sensitive attributes disclosure– Sensitive attributes disclosure takes place when an attacker obtains the information of a sensitive user attributes. Sensitive attributes may be linked with an entity and link relationship.

There are many examples of accidental disclosure of private information of users' data that causes organizations to be conservative in releasing the network data, such as the AOL search data example and attacks on Netflix data.

As per the promises of social networks there is a need to address these issues. Therefore, data needs to be released to third parties in such a way that ensures the privacy of the users. Thus data should be anonymized before releasing or publishing to third parties [2].

But preserving privacy in social networks is difficult even after Data anonymization because of increasingly overlapping user-bases among social networking services and certain De-Anonymization attacks on anonymized data [12].

II. PROBLEM DEFINITION

Existing privacy protection mechanisms for social networks are only effective against very restricted adversaries and have been evaluated on small, simulated networks whose characteristics are different from real social networks.

So there needs an effective data privacy preserving algorithm for social network users.

This quantitative study examines the present problems on Privacy Preserving algorithms on social networks and finds a better advanced algorithm for Data Privacy.

- The present problems which will be addressed are re-identification of sensitive attributes should be made difficult.

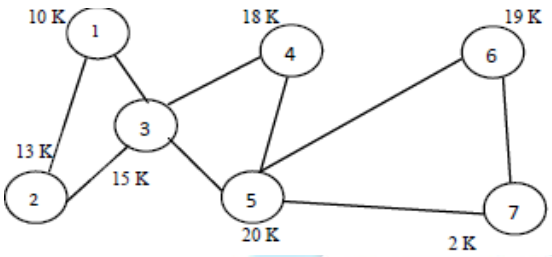
- Edge Privacy should be taken into account has all edges are Mutual friend attacks public by default.
- Mutual Friends Attacks and Seed Identification should be made difficult.

III. LITERATURE SURVEY

Existing models of preserving privacy for micro-data have been utilized for social network data. Work has been done by various researchers using k-anonymity, l-diversity and integrated approach of k-anonymity l-diversity for protecting users' data while publishing it online. Social network data is unstructured data represented as a graph where each node/vertex represents an individual and edges represent link/association between nodes.

Fig1 shows a social network structure with 7 nodes representing individuals and salaries are sensitive attributes shown by labels. Privacy preserving techniques are based on that notion.

Fig. 1



Privacy preserving techniques are developed keeping following things into consideration:

1. Adversary's knowledge 2. Utility of the data after release

So, depending upon the knowledge that an adversary uses to attack the target node following techniques have been developed by various researchers using the notion of k-anonymity . Wei et al. considered the privacy disclosure in online social network data publishing. It has been assumed that adversaries have the knowledge of the degree of a target individual and the target's immediate neighbor's.

A practical solution to defend against background knowledge attacks has been proposed. Anonymized social networks obtained by proposed method can be used to answer aggregate network queries with high accuracy. Social network has been modeled as an undirected labeled graph. k-subgraph has been proposed to reduce the risk of privacy disclosure in social network data publication.

Zou et al. proposed k-automorphism based on the assumption that the adversary has knowledge about degree, sub graph and neighbor of the target node. Tripathy et al. proposed an algorithm for graph isomorphism based on adjacency matrix. It says that a subgraph is indistinguishable from at least k-1 other subgraphs.

Cheng et al. used k-isomorphism to preserve privacy when adversary has subgraph knowledge. Wu et al. proposed k-symmetry technique to protect privacy against re-identification using subgraph information. Lan et al. developed an algorithm called KNAP against 1neighborhood attack for publishing social networks data. Skarkala et al. applied K-anonymity to weighted social networks.

Liu et al. proposed the concept of k-degree to prevent vertex re-identification through the information of vertex degree. Preserving privacy in social networks using k-anonymity protects against linking disclosure but still it may leak privacy under the cases of homogeneity and background knowledge attacks. Moreover, K-anonymity doesn't protect against attribute disclosure. So, L-diversity was developed by Machanavajjhala.

Panda et al. used a new practical and efficient definition of privacy called l-diversity on preserving privacy in collaborative social network data and the effect on the utility of the data for social network analysis has been seen.

It has been identified that l-diversity social network still may leak privacy as an adversary may have some prior knowledge about the sensitive attribute value of an individual before seeing the released table. After seeing the released table, the adversary may have a posterior knowledge.

Information gain i,e., the difference between the posterior knowledge and the prior knowledge is the factor to leak privacy. So the concept of t-closeness was suggested. Li et al. proposed to preserve relationship privacy between two users one of whom can be identified in the released social network data. l-diversity anonymization model has been defined to preserve users' relationship privacy.

Two graph manipulation algorithms, MaxSub and MinSuper, have been proposed to achieve l-diversity anonymization. Then, to preserve privacy in better way integrated approach of K-anonymity and L-diversity has been suggested by few authors as mentioned below. Kavianpour et al. proposed an integrated algorithm that takes the advantages of K-anonymity and l-diversity algorithm then evaluated the effectiveness of the combined strengths. Proposed algorithm has been able to increase the level of privacy for social network users by anonymizing and diversifying disclosed information.

Tripathy et al. proposed an algorithm which follows k-anonymity and ldiversity properties and can handle a variant of multisensitive attributes during anonymization process. Drawback of proposed algorithm is that it still needs some improvements in order to reduce the complexity so that it can be applied to large social networks. Yuan et al. defined a k-degree l-diversity anonymity model for the protection of structural information and sensitive labels of people.

Many privacy models like k-anonymity to prevent node re-identification through structure information have been proposed but an attacker may still be able to obtain private information of a person i.e. the label-node relationship is not well protected by pure structure anonymization methods. An anonymization methodology has been done by adding noise nodes into the original graph with the consideration of introducing the least distortion to graph properties.

IV. RESEARCH OBJECTIVES

The prime aim/goal of the proposed research work is to design and develop an innovative Data Privacy Preserving algorithm in Social Networks in order to achieve desired Data Privacy in

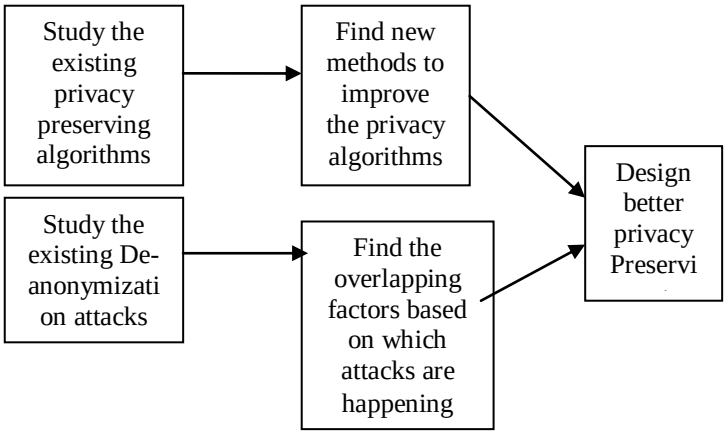
Social Networks. In order to carry out this study, following research objectives are set.

1. To preserve usefulness (utility) of anonymized data is an important aspect while applying techniques for privacy preservation. So, there is a need to develop methodologies that can quantitatively measure utility of data. There is need to evaluate various techniques in terms of tradeoff between privacy and utility.
2. Many algorithms like k-anonymity, L-diversity, integrated approach of k-anonymity & L-diversity have been developed for preserving privacy of social network user data but existing techniques leads to substantial information loss.
3. Anonymization techniques have been developed for one time released network data. But many applications require publishing data periodically so there is a need to develop techniques that can preserve privacy of dynamic releases.
4. Techniques are available for preserving privacy in case of distributed tabular data. However, in case of social network distributed privacy preserving techniques are not well reported.
5. Existing privacy preserving approaches for social networks have been evaluated using either small datasets or synthetic datasets. There is need to conduct empirical experiments on large datasets.
6. There is no existing technique which can prevent homogeneity attacks, background knowledge attacks, attacks arising due to distance between sensitive values.

It became evident from the literature that privacy of users is the main concern and need for the days. Various models proposed for tabular micro-data have been adopted for preserving privacy of social network data. Techniques like Kanonymity, L-diversity, integrated K-anonymity L-diversity have been used till now but these techniques lead to substantial information loss. So, there is a scope of improvement of the techniques that provide privacy preservation with minimum information loss and better utility of released data.

V. RESEARCH METHODOLOGY

The different steps/Methods in designing the Propose study are given below.[Refer Fig:2] steps/Methods in designing the Proposed Research are given below.



1. Study the existing privacy preserving algorithms for anonymized data.
2. Find new methods to improve the privacy algorithms.
3. Study the existing De-anonymization attacks
4. Find the overlapping factors based on which attacks are happening
5. Design a better privacy Preserving algorithm
6. The different steps/Methods in designing the Proposed Research are given below.

VI. IMPLEMENTATION PHASE

The mathematical modeling will be simulated using statistical tool such as R Tool using R Programming and Graph Database Neo4j; the proposed Data Privacy Preserving system/algorithm will be compared with the present Privacy preserving algorithms, and the same will be experimented against De-anonymization attacks in order to benchmark it. Results are expected to show that the proposed algorithm provides better Data privacy compared to existing algorithms.

VII. EXPECTED OUTCOME

The possible outcome of the proposed study is expected as follows:

- Efficient Data Privacy: The current technique is expected to provide better data privacy compared to the present Data Privacy Algorithms and gain a better utility of released data [anonymized data]
- Scalability: In order to achieve high scalability we are experimenting on larger datasets
- Computational Capability: With an anticipated less time and space complexity, fast processing speed over large scale , the current study is expected to possess lower degree of computation complexities

VIII.REFERENCES

[1]. Privacy Preserving Techniques in Social Networks Data Publishing-A Review Amardeep Singh Divya Bansal, Sanjeev Sofat, IJCAT Feb2014.

[2]. B. Zhou, Jian Pei, Wo-Shun Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data," ACM SIGKDD Explorations Newsletter, Vol. 10, pp. 12-22, 2008.

[3]. B. Krishnamurthy and C.E. Wills, "Characterizing Privacy in Online Social Networks," Proc. First Workshop Online Social Networks (WOSN), 2008.

[4]. Narayanan and V. Shmatikov, "De-Anonymizing Social Networks," Proc. IEEE 30th Symp. Security and Privacy, 20090002E

[5]. M. Hay, G. Miklau, D. Jensen, P. Weis, and S. Srivastava,"Anonymizing Social Networks," technical report, Univ. Massachusetts,Amherst, 2007.

[6]. B. Zhou and J. Pei, "Preserving Privacy in Social Networks against Neighborhood Attacks," Proc. Int'l Conf. Data Eng. (ICDE), 2008

[7]. J. Douceur, "The Sybil Attack," Proc. First Int'l Workshop Peer-to- Peer Systems, vol. 2429, pp. 251-260, 2002.

[8]. J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney, "Statistical Properties of Community Structure in Large Social and Information

Networks," Proc. 17th Int'l Conf. World Wide Web (WWW),2008.

[9]. C. Wilson, B. Boe, A. Sala, K. Puttaswamy, and B. Zhao, "User Interactions in Social Networks and Their Implications," Proc. Fourth ACM European Conf. Computer Systems (EuroSys), 2009

[10]. A Two-Stage Deanonymization Attack against Anonymized Social Networks Wei Peng, Student Member, IEEE, Feng Li,Member, IEEE, Xukai Zou, Member, IEEE, and Jie Wu, Fellow, IEEE TRANSACTIONS ON COMPUTERS, VOL. 63, NO. 2, FEBRUARY 2014

[11]. S. Sorlin and C. Solnon, "Reactive Tabu Search for Measuring Graph Similarity," Proc. Fifth IAPR Int'l Conf. Graph-Based Representations in Pattern Recognition, vol. 3434, pp. 172-182, 2005.

[12]. R. Xiang, J. Neville, and M. Rogati, "Modeling

[13]. Relationship Strength in Online Social Networks," Proc. ACM 19th Int'l Conf. World Wide Web (WWW), 2010

