Privacy Protection on Social Network Data Using Anonymization Methodology

[1] **D. Shalini Reddy**, [2] **K. Narayana**

[1] Post-Graduate Student,
Department of Computer Science and Engineering,
SIT, PUTTUR, India

[2] Head & Associate Professor,
Department of Computer Science and Engineering,
SIT, PUTTUR, India

Abstract - Privacy is amongst the major concerns when publishing or sharing social network data for social science research and also business analysis. Recently, researchers have developed privacy models just like k-anonymity in order to avoid node *reidentification* through structure information. However, even though these privacy models are enforced, an attacker can always have the capacity to infer one's information that is personal if the list of *nodes largely shares* the identical sensitive labels (i.e., attributes). To put it differently, the label-node relationship just isn't thoroughly protected by pure structure *anonymization* methods. We present privacy protection algorithms which facilitate graph data to become published within a form in a way that an adversary who possesses information regarding a *node's* neighborhood cannot safely infer its identity as well as sensitive labels. To the present aim, the algorithms transform an original graph in to a graph during which *nodes* are sufficiently indistinguishable. The algorithms are created to accomplish that while losing only small amount information and even though preserving the maximum amount of Utility as possible.

***Keywords* - Author Guide, Article, Camera-Ready Format, Paper Specifications, Paper Submission.**

1. Introduction

While using rapid growth of social networks, including Face book and LinkedIn, increasingly more researchers found that it's a great possibility to obtain useful information from all of these social network data, including the user Behavior, community growth, disease spreading, etc. However, it can be paramount that published social network data should never reveal information that is personal of an individual. Thus, how you can protect individual's privacy including duration preserve the utility of social network data becomes a challenging topic. Social network data contains sensitive and private information in regards to the users. Thus sharing with this data in their raw form may breach privacy of an individual. Individual privacy is described as "the proper of the individual to make the decision what details about himself really should be communicated to others and under what circumstances" [8]. A privacy breach occurs when private and private information regarding the user is disclosed for an adversary.

So, preserving privacy of an individual while publishing user's collected data is an essential research area. Work continues to be carried out by various researchers within this direction. In this particular paper, a graph model where each vertex within the graph is assigned to a sensitive label. Recently, much work may be done on anonym zing tabular micro data. A various privacy models and also anonymization algorithms are actually developed. In tabular micro data, many of the nonsensitive attributes, called quasi identifiers, may be used to reidentify individuals and their sensitive attributes. When publishing social network data, graph structures may also be published with corresponding social relationships. As a result, it might be exploited as being new methods to compromise privacy.

1.1 Grouping of Privacy Breach

The privacy breaches in social networks are usually categorized into three types:

i. Identity disclosure - Identity disclosure occurs when anyone behind accurate documentation is exposed. This kind of breach causes the revelation of data of any user and relationship he/she explains to other individuals within the network.

ii. Sensitive link disclosure - Sensitive link disclosure happens when the associations between two individuals are revealed. Social activities generate this kind of information when social networking services are used by users.

iii. Sensitive attribute disclosure – Sensitive attribute disclosure happens when an attacker obtains the information of the sensitive and confidential user attribute. Sensitive attributes might be related to an entity and link relationship

Every one of these mentioned privacy breaches pose severe threats like stalking, blackmailing and robbery because users expect privacy in their data through the Service provider end. Besides so it damages the image and standing of an individual. Depending on the promises of social networks there exists a have to address these issues. Therefore, data must be released to third parties in a way that ensures the privacy with the users. Thus data needs to be anonymized before releasing or publishing to third parties

2. Related Work

2.1 Attacks on Anonymized Social Networks

In this particular paper present both active and passive attacks on anonymized social networks, showing that both varieties of attacks enable you to reveal truth identities of targeted users, even from simply a single anonymized copy from the network, along with a surprisingly small investment of effort from the attacker. It describe active attacks through which an adversary chooses an arbitrary group of users whose privacy it wishes to violate, creates only a few new user accounts with edges to those targeted users, and helps to create a pattern of links one of several new accounts with all the goal of which makes it be noticed within the anonymized graph structure.

The adversary then efficiently finds these new accounts along with the targeted users inside the anonymized network that may be released. At the theoretical level, the coming of O (p log n) nodes through the attacker within the n-node network can start compromising the privacy of arbitrary targeted nodes, the type with the attacks.

The social network is definitely an n-node graph G = (V, E), representing interactions within the on-line system. Nodes correspond to user accounts, and an edge (u, v) points too u has communicated with v (again, think about the example Illustration showing an e-mail or instant messaging network). The attacks become much better to perform when the released graph data is directed; for the majority of in the paper we are going to therefore think about the harder case of undirected graphs, in which we assume which the curator in the data—the agent that releases the anonymized network — eliminates the directions within the edges. The active attacks will always make utilisation of the following two kinds of operations.

First, an individual can make a new user account within the system; this adds a brand new node to G. Second, a node u can choose to communicate with a node v; this adds the undirected edge (u, v) to G. The purpose of the attack is always to take an arbitrary pair of targeted users w1, . . . ,wb , for each set of them, to utilize the anonymized copy of G to understand whether or not the edge (wi ,wj) in point of fact exists. This is actually the sense the location where the privacy these users will probably be compromised. (Other privacy compromises, including learning the degree of the targeted user, also occur, but we focus our attention on understanding about edges.)

2.2 Anatomy Simple and Easy and Effective Privacy Preservation

This paper presents a deliberate study with the anatomy technique. First, to formalize the brand new methodology, depending on the privacy requirement of l-diversity. Every set of QIT and ST helps to ensure that the sensitive worth of any individual mixed up in micro data might be correctly inferred by an adversary with probability at most of the 1/l. A larger l causes stronger privacy protection.

3. Framework of Anonymization

A whole new group of various procedures for anonymizing social network data according to merging the entities into classes and also by mapping the entities and nodes often denoted them in anonymized target graph. Anonymization techniques are increasingly being challenging to attackers with larger background information. Yet turned unsatisfied since it has lower utility and less graph structure will be here revealed.

The framework presentation for analyzing privacy preservation develops a brand new reidentification algorithm for target of numerous anonymized social network graphs. Our Deanonymizing algorithm will be based upon network topology, doesn't contain creation of more number of dummy “Sybil” nodes.

Existing defenses works involving the overlapping of target network and adversary’s information. A common reidentification algorithm established that it might successfully monitors and de-anonymize large amount of users in anonymous social network graph. Since human names has never unique identity, this algorithm having overlap condition in membership.

The state analyzation of privacy protection in social network graphs describes effective anonymization attacks to protect from hackers. In this particular paper, starclique, a minimum graph required k-anonymity, where user is identified for all possible contributions of information objects.

The identification of social intersection attack can compromise users to recognize shared objects relying on social graph topology

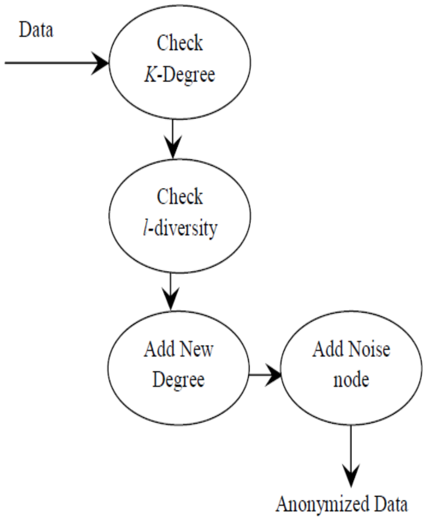

Fig 1: Architecture

Preserving privacy within social networks using k-anonymity protects towards linking disclosure however it might leak privacy within the cases of homogeneity and background information attacks. Moreover, K-anonymity doesn't control attribute disclosure. So, L-diversity got its start by Machanavajjhala Panda et al. used a novel practical and efficient concept of privacy called l-diversity on preserving privacy in collaborative social network data along with the impact on the utility from the data for social network analysis may be seen. It is often identified that l-diversity social network most likely will leak privacy just as one adversary might have some earlier knowledge regarding the sensitive attribute importance of an individual before seeing the released table. After looking at the released table, the adversary might have a posterior knowledge. Information gain i,e., the main difference relating to the posterior knowledge along with the earlier knowledge could be the answer to leak privacy. Therefore the idea of t-closeness may be suggested to become introduced. Li et al. proposed to preserve relationship privacy between two users considered one of whom is usually identified within the released social network data. L-diversity anonymization model have been defined to preserve users' relationship privacy. Two graph manipulation algorithms, MaxSub and MinSuper, are already proposed to accomplish l-diversity anonymization.

4. Conclusion

In this particular paper, k-degree-l-diversity model has implemented for privacy preserving social network data publishing. Implementation of both distinct l-diversity and recursive (c, l)-diversity also happened. So that you can attain the dependence on k-degree-l-diversity, a noise node adding algorithm to make A whole new graph through the original graph while using the constraint of introducing fewer distortions for the original graph. Rigorous analysis of the theoretical bounds within the variety of noise nodes added and their impacts by using an important graph property. Extensive experimental results demonstrate how the noise node adding algorithms can performs a much better result compared to the previous work of edge editing method. In a very distributed environment, data publication satisfy certain privacy requirements, a hacker can still collapse privacy by connecting the information by different users. Similar Protocols needs to be developed to conserve the data publishers to assure the Privacy preservation.

References

[1] C wang, Sherman S. M. Chow, Q. Wang, K Ren and W.Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage",IEEE Trasaction on Computers I, vol. 62, no. 2, pp.362-375 , February 2013.

[2] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy Preserving Public auditing for storage security in cloud computing," in Proc.of IEEE INFOCOM'10, March 2010.

[3] Wang Shao-hu, Chen Dan-we, Wang Zhi-weiP, Chang Su-qin, "Public auditing for ensuring cloud data storage security with zero knowledge Privacy" College of Computer, Nanjing University of Posts and Telecommunications, China, 2009

[4] KunalSuthar, Parmalik Kumar, Hitesh Gupta, "SMDS: secure Model for Cloud Data Storage", International Journal of Computer applications, vol56, No.3, October 2012

[5] AbhishekMohta, Lalit Kumar Awasti, "Cloud Data Security while using Third Party Auditor", International Journal of Scientific & Engineering Research, Volume 3,Issue 6, ISSN 2229-8 June 2012.

[6] Q. Wang, C. Wang,K.Ren, W. Lou and Jin Li "Enabling Public Audatability and Data Dynamics for Storage Security in Cloud Computing", IEEE Transaction on Parallel and Distributed System, vol. 22, no. 5, pp. 847 – 859,2011.

[7] D. Shrinivas, "Privacy-Preserving Public Auditing in Cloud Storage security", International Journal of computer science nad Information Technologies, vol 2, no. 6, pp.2691-2693, ISSN: 0975-9646, 2011

[8] K Govinda, V. Gurunathprasad and H. sathishkumar, " Third Party Auditing for Secure Data Storage in Cloud Through Digital Signature Using RSA", International Journal of Advanced science and Technical Research, vol 4,no. 2, ISSN: 2249-9954,4 August 2012

[9] S. Marium, Q. Nazir, A. Ahmed, S. Ahthasham and Aamir M. Mirza, "Implementation of EAP with RSA for Enhancing The Security of Cloud Computig", International Journal of Basic and Applied Science, vol 1, no. 3, pp. 177- 183, 2012

[10] XU Chun-xiang, HE Xiao-hu, Daniel Abraha, "Cryptanalysis of Auditing protocol proposed by Wang et al. for data storage security in cloud computing", http://eprint.iacr.org/2012/115.pdf, and cryptologyeprintarchieve: Listing for 2012.

[11] B. Dhiyanesh "A Novel Third Party Auditability and Dynamic Based Security in Cloud Computing" ,

International Journal of Advanced Research in Technology, vol. 1,no. 1, pp. 29-33, ISSN: 6602 3127, 2011

[12] C. Wang, Q. Wang and K. Ren, “Ensuring Data Storage security in Cloud Computing”, IEEE Conference Publication, 17th International Workshop on Quality of Service (IWQoS), 2009

[13] Balkrishnan. S, Saranya. G, Shobana. S and Karthikeyan. S, “Introducing Effective Third Party Auditing (TPA) for Data Storage Security in Cloud”, International Journal of computer science and Technology, vol. 2, no. 2, ISSN 2229-4333 (Print) | ISSN: 0976-8491(Online), June 2012.

[14] Sushmita Ruj, Milos Stojmenovic, Amiya Nayak, "Decentralized Access Control with Anonymous Authentication for Secur-ing Data in Clouds,"IEEE Transactions on Parallel and Distributed Systems, pp. 1045-9219, 2013.

[15] S. Ruj, M. Stojmenovic and A. Nayak, “Privacy Preserving Access Control with Authentication for Securing Data in Clouds”, IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, pp. 556–563, 2012.

[16] C. Wang, Q. Wang, K. Ren, N. Cao and W. Lou, “Toward Secure and Dependable Storage Services in Cloud Computing”, IEEE T. Services Computing, vol. 5, no. 2, pp. 220–232, 2012.

[17] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, “Fuzzy keyword search over encrypted data in cloud computing,” in IEEE INFOCOM. , pp. 441–445, 2010.

[18] S. Kamara and K. Lauter, “Cryptographic cloud storage,” in Financial Cryptography Workshops, ser. Lecture Notes in Com-puter Science, vol. 6054. Springer, pp. 136–149, 2010.

[19] H. Li, Y. Dai, L. Tian, and H. Yang, “Identity-based authentication for cloud computing,” in CloudCom, ser. Lecture Notes in Computer Science, vol. 5931. Springer, pp. 157–166, 2009.

[20] C. Gentry, “A fully homomorphic encryption scheme,” Ph.D. dissertation, Stanford University, 2009, http://www.crypto.stanford.edu/craig.

[21] A.-R. Sadeghi, T. Schneider, and M. Winandy, “Token-based cloud computing,” in TRUST, ser. Lecture Notes in Computer Science, vol. 6101. Springer, pp. 417–429, 2010.

[22] R. K. L. Ko, P. Jagadpramana, M. Mowbray, S. Pearson, M. Kirchberg, Q. Liang, and B. S. Lee, “Trustcloud: A framework for accountability and trust in cloud computing,” HP Technical Report HPL-2011-38. Available at http://www.hpl.hp.com/techreports/2011/HPL-2011-38.html.

[23] R. Lu, X. Lin, X. Liang, and X. Shen, “Secure Provenance: The Essential of Bread and Butter of Data Forensics in Cloud Computing,” in ACM ASIACCS, pp. 282–292, 2010.

[24] D. F. Ferraiolo and D. R. Kuhn, “Role-based access controls,” in 15th National Computer Security Conference, 1992.

[25] A B Lewko and B Waters, “Decentralizing attribute based encryption”, springer 2011.