Enhancing Social Network Privacy Using RSA Algorithm by Data Anonymization Methodology

Nisha Soman

Abstract -- In the social network site large amount of data can be published and shared between the users. So the major concern when publishing and sharing the data in the site is to preserve the data with privacy. The goal of preserving privacy is to prevent an attacker from re-identifying a user and finding the fact that a certain user has a specific sensitive value. So define a RSA algorithm for safely publishing a social network labeled graph, and then develop corresponding graph anonymization algorithms in the original graph. Then further propose a novel anonymization algorithm of introducing noise nodes in the original graph with the least distortion to key graph properties. The user of RSA creates and secures the graph by means of two large prime numbers.

Keywords: Data Anonymization, Privacy, Social Networks, RSA, Prime numbers

1 INTRODUCTION

THE rapid growth of social networks, such as Facebook and LinkedIn, society is experiencing exponential growth in the number and variety of data collections containing person-specific information as computer technology, network connectivity and disk storage space become increasingly affordable. The social graph in the Internet context is a graph that depicts personal relations of internet users. The following Fig 1. shows a network graph which consists of large number of nodes. These nodes represented the sensitive labels of the individuals.

All of these nodes consist of sensitive labels where each node can transfer the data from one node to the respective destination node. The sensitive labels represent the sensitive data of the individual users. Recently a variety of privacy models and anonymization algorithms have been developed such as K-anonymity, l-diversity, t-closeness etc. Even when these privacy models are enforced an attacker may still be able to infer one's private information if a group of nodes largely share the same sensitive labels. So introducing the graph with noise nodes where the data anonymization algorithms have been used. The secrecy of the data can be improved by means of applying RSA algorithm in the original graph and introduce the noise nodes in the original graph.

Nisha Soman completed Masters Of Engineering in Department of Computer Science and Engineering, under Anna University, Chennai.

Email:somannisha45@gmail.com

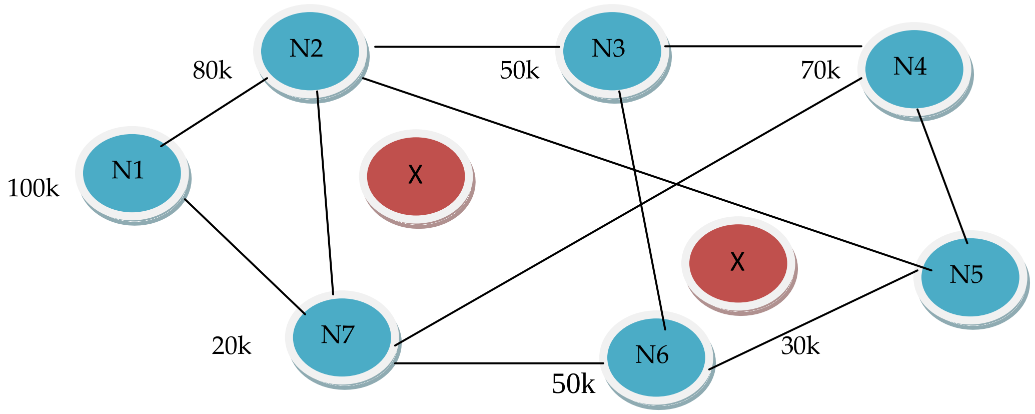

Fig1. Network Graph with nodes having labels and noise nodes

So here consider the network by graph model which consists of large number of nodes that describe the social relationship between the users. In this graph model where each vertex in the graph is associated with the sensitive label. For each distinct degree appearing in this graph, there exist at least two nodes. Moreover, for those nodes with the same degree, they contain at least two distinct sensitive labels. Thus, the attacker cannot reidentify a node or find the node-label relation with degree knowledge.

2 EXISTING SYSTEM

In the existing system privacy preserving goal is to prevent an attacker from reidentifying a user and finding the fact that a certain user has a specific sensitive value. To achieve this goal, defined a *k*-degree-*l*-diversity (KDLD) model for safely publishing a labeled graph, such as degrees and distances between nodes. Here combine a *k*-degree anonymity with *l*-diversity to prevent not only the reidentification of individual nodes but also the revelation of a sensitive attribute associated with each node. The two methods are Clustering and Edge editing was used. Clustering is to merge a subgraph to one super node, which is unsuitable for sensitive labeled graphs. Edge-editing methods keep the nodes in the original graph unchanged and only add/delete/swap edges.

2.1 Problem Description

By using the Clustering method, it is unsuitable for sensitive labeled graph and hence the node label relations may be lost. The Edge Editing may change the properties of the graph by adding new edges, swapping edges or deleting an edge from the original graph. So the secrecy of the network graph can be improved by means of using RSA algorithm by adding new noise nodes in the original graph.

3 DATA ANONYMIZATION CONCEPTS AND TECHNIQUES USING RSA

Data anonymization is the process of either encrypting or removing personally identifiable information from the data sets, so that the people whom the data describe remain anonymous. The technology that converts clear text data into a non human readable and irreversible form. The term "anonymous message" typically refers to a message that does not reveal its sender. If someone sends a file, there may be information on the file that leaves a trail to the sender. However, once the file is anonymized, data associated with it being sent cannot be traced to the sender. For that introducing the noise node in the original graph to find the default hackers are hacking the data.

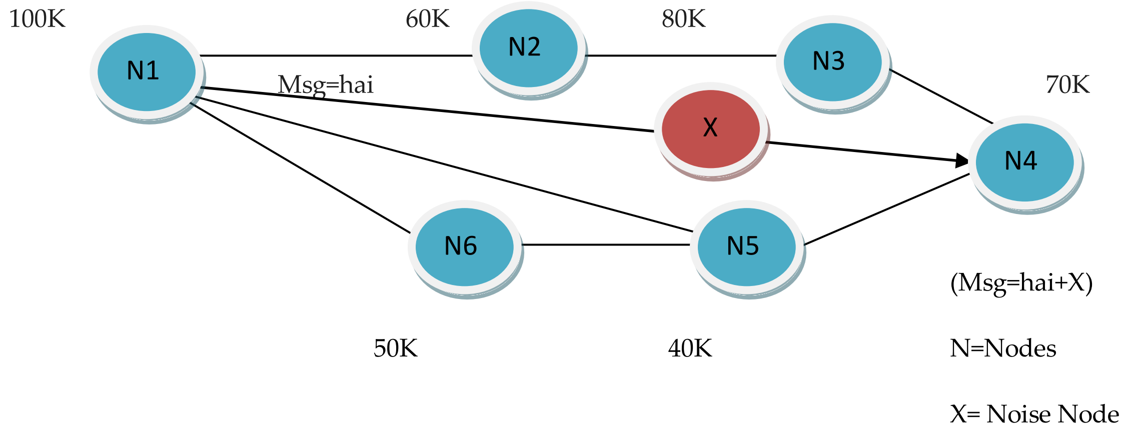

Fig 2. Data Anonymization

Anonymization is a technique that enterprises can used to increase the security of the data in the public social network site can be organized by means of introducing noise node in the social network site and still allowing the data to be analyzed and used. The data anonymization can used the key and then transferred the data. Using the RSA algorithm the public key cryptosystems and can be widely used for secure data transmissions Then provide the rigorous analysis of the theoretical bounds of number of noise nodes are added and the impact of the original graph property. Anonymized data can be stored in the social network site and processed without concern that other individuals may capture the data. Later the results can be collected and mapped to the original data in the secure area.

The message can be sending from node N1 to N4 by introducing the noise node. The use of adding the noise node can be introduced to detect the data can be hacked by the attackers. If the original data can be hacked by the

4 RESULTS

The growth of the Internet and electronic commerce has brought to the forefront the issue of privacy in social network. Large volumes of personal and sensitive information are electronically transmitted and stored every day. Encryption is the standard method for making a communication private. Anyone wanting to send a private message to another user encrypts (enciphers) the message before transmitting it. Only the intended recipient knows how to correctly decrypt (decipher) the message. Anyone who was "eavesdropping" on the communication would only see the encrypted message. Because they would not know how to decrypt it successfully, the message would make no sense to them. As such, privacy can be ensured in electronic communication. By using the data

default attackers the message can be changed to some other data. But by introducing the noise node by the knowledge of the sender can be sustained that the data cannot be changed instead of the original data with the noise node id can be received. From these the sender can conclude that the original message is not attacked by the hackers. If the original message hacked by the default hackers the data can be converted to some other data. A user of RSA creates and then publishes a public key based on the two large prime numbers along with an auxiliary value. The prime numbers must be kept secret.. RSA encrypts messages through the following algorithm, which is divided into 3steps:

1. Key Generation

2. Encryption

3. Decryption

anonymization method that converts clear text data into a non human readable so that the default attackers in the network can't find out the original message. Apart from these if the data can be hacked it can be understand and get by means of adding noise nodes in the graph. To demonstrate the effectiveness of the graph construction algorithm, compare the original graph with data anonymization algorithm by introducing noise nodes in the original graph. Here in the original graph can add more number of nodes. The noise node added can improve the properties of the graph.

5 CONCLUSIONS

Here it is proposed a data anonymization algorithm in the original graph and use the prime numbers in the RSA model for preserving privacy when published social network data. In order to achieve the requirement of secure data transmission, design a noise node adding algorithm to construct a new graph from the original graph with the constraint of introducing fewer distortions to the original graph. Then give a rigorous analysis of the theoretical bounds on the number of noise nodes added and their impacts on an important graph property. The extensive experimental results demonstrate that the noise node adding algorithms can achieve a better result than the previous work using edge editing only. The prime numbers used in the RSA only have divisors of 1 and self they cannot be written as a product of other numbers.

6 References

[1] X. Xiao and Y. Tao, "Anatomy: Simple and Effective Privacy Preservation," Proc. 32nd Int'l Conf. Very Large Databases (VLDB '06), pp. 139-150, 2006

[2] X. Ying, X. Wu, and D. Barbara, "Spectrum Based Fraud Detection in Social Networks," Proc. IEEE 27th Int'l Conf. Very Large Databases (VLDB '11), 2011.

[3] X. Ying and X. Wu, "Randomizing Social Networks: A Spectrum Preserving Approach," Proc. Eighth SIAM Conf. Data Mining (SDM '08), 2008.

[4] E. Zheleva and L. Getoor, "To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles," Proc. 18th Int'l Conf. World Wide Web (WWW '09), pp. 531-540, 2009.

[5] J. Cheng, A.W.-c. Fu, and J. Liu, "K-Isomorphism: Privacy Preserving Network Publication against Structural Attacks," Proc. Int'l Conf. Management of Data, pp. 459-470, 2010.

[6] A. Campan, T.M. Truta, and N. Cooper, "P-Sensitive K-Anonymity with Generalization Constraints," Trans. Data Privacy, vol. 2, pp. 65-89, 2010

[7] J. Cheng, A.W.-c. Fu, and J. Liu, "K-Isomorphism: Privacy Preserving Network Publication against Structural Attacks," Proc. Int'l Conf. Management of Data, pp. 459-470, 2010.

[8] G. Cormode, D. Srivastava, T. Yu, and Q. Zhang, "Anonymizing Bipartite Graph Data Using Safe Groupings," Proc. VLDB Endowment, vol. 1, pp. 833-844, 2008.

[9] S. Das, O. Egecioglu, and A.E. Abbadi, "Privacy Preserving in Weighted Social Network," Proc. Int'l Conf. Data Eng. (ICDE '10), pp. 904-907, 2010

[10] K.B. Frikken and P. Golle, "Private Social Network Analysis: How to Assemble Pieces of a Graph Privately," Proc. Fifth ACM Workshop Privacy in Electronic Soc. (WPES '06), pp. 89-98, 2006.

[11] S.R. Ganta, S. Kasiviswanathan, and A. Smith, "Composition Attacks and Auxiliary Information in Data Privacy," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 265- 273, 2008.

[12] S. Bhagat, G. Cormode, B. Krishnamurthy, and D. S. and. Class-based graph anonymization for social network data. PVLDB, 2(1), 2009.