IDENTITY DISCLOSURE PROTECTION IN DYNAMIC NETWORKS USING K^W – STRUCTURAL DIVERSITY ANONYMITY

Gowthamy.R[1*] and Uma.P[2]

[*1]M.E.Scholar, Department of Computer Science & Engineering Nandha Engineering College, Erode, Tamil Nadu, India and
[2]Assistant Professor, Department of Computer Science & Engineering, Nandha Engineering College, Erode, Tamil Nadu, India

ABSTRACT

The data mining figures out accurate information for requesting user after the raw data is analyzed. Among lots of developments, data mining face hot issues on security, privacy and integrity. Data mining use one of the latest technique called privacy preserving data publishing (PPDP), which enforces security for the digital information provided by governments, corporations, companies and individuals in social networks. People become embarrassed when adversary tries to know the sensitive information shared. Sensitive information is gathered through the vertex and multi community identities of the user. Vertex identity denotes the self-information of user like name, address, mobile number, etc. Multi community identity denotes the community group in which the user participates. To prevent such identity disclosures, this paper proposes K^W-structural diversity anonymity technique, for the protection of vertex and multi community identity disclosure. In K^W-structural diversity anonymity technique, k is privacy level applied for users and W is an adversary monitoring time.

KEYWORDS

Privacy preservation, vertex disclosure and multi community disclosure, K^W-structural diversity anonymity technique.

1.INTRODUCTION

Data mining is about finding and filtering new information from millions of data.Data mining finds out the hidden data of particular user database and tries to explore it with additional information as a technology for the enterprises using data warehouse. Analyzing and extracting useful information from database is necessary for further use in different fields like market analysis, fraud detection, science exploration, etc. Data cleaning, data integration, data transformation, pattern evaluation, data presentation also should be done along with extracting information. Once the retrieval of perfect information which is in expected quality data mining process exhibits.

Due to the development of networking websites, online communities, online shopping, and telecommunications the number of network data grows rapidly today. Once the data is explored in the social network it has been used in different concern like application development, creating

advertisements and for preparing contents for research works. Even though rapid data established in the network, privacy plays a wide role in organizing and those information which becomes hot issue in this decade.

As Figure 1 show, the proposed techniques that completely anonymize user identities, communication security which preserves privacy of different users present over the social network. Variation occurs according to the methods used to hide user identity. Topology-preserving techniques used to preserve user identities by modifying the graph against topology-based attacks. On the other hand, Vertex classifying and relabeling techniques don't involve modifying the graph but none other than cluster vertices and relabel them to protect user identities.

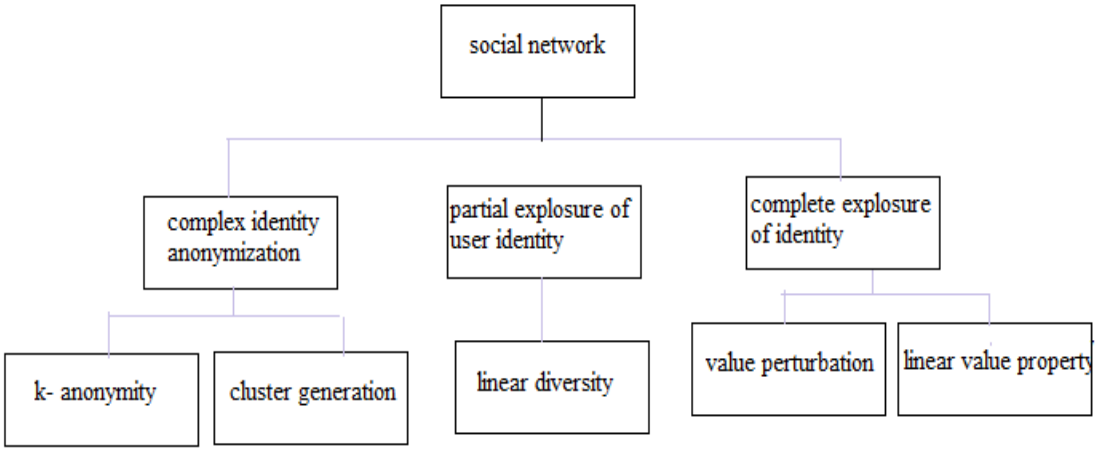

Figure 1.Continuum of relation privacy preservation.

S. Bhagat et al.[1], proposed the concept of protecting labels in the directed graph, without addressing the identity protection problem. Due to protection is established a particular protection model is not designed for preserving identity protection. To overcome this issue in the social network, the designed network should be anonymized at each snapshot of the updation done at the network like connecting with other nodes as friends and entering into the community groups enrolled in the social network as constant groups. Due to the privacy lacking in non-development techniques the attacker can hack the needed info from the online sites. The hacking process not at all ends with stealing the information but the hacker compares it with multiple data at different snapshots in different online sites. This is the problem project in this paper about the disclosure of identities in dynamic network. Identity denotes the vertex or node information and the community groups enclosed in the social networks. Data which is considered as a private data can be retrieved based on the relationship of user with other users (degree of users calculated according to the number of nodes connected with a particular user) at particular timestamp.

In this paper, disclosure problem like revealing of personal data of a user and the group participation is plotted. The considerations of multi community identity and sequential publications demand a new dynamic privacy model, since the existing privacy model such as k-structural diversity cannot ensure privacy in both situations. For time stamp problem, at each actions done at the social site the attacker watches the user to gather the information wantedat the

particular snapshot. For protecting against such an adversary, introducing a new dynamic privacy scheme, named dynamic K^w-structural diversity anonymity.

2.OVERVIEW

In Social networks activities are modeled between individuals when time changes continuously. Analyzers, customers and collaborators faces a continuous demand for privacy-preserving while data sharing in social network. This project projects the privacy risks of identity disclosures in social network while sequential releases occur.

Anonymization technique plays the major role in the protection of sensitive information of an individual. This technique is used for detecting adversaries and provides privacy for sensitive information. Privacy can be provided by encrypting or removing personally identifiable information from the dataset. Anonymization can be applied in the fields of online shopping, online communication, telecommunication and social network carts.

This paper focus on safe guarding of disclosure problem happening in social network. Disclosure problem denoted here are of two from the lot, which is users particular information revealing and the hacking of social network groups in which the user participate in. To overcome the mentioned problem, anonymization process should be done at each and every second in the social network for each action performed like adding or deleting friends and like participating in different groups or exiting from the group. Due to the fixed timestamp problem, attacker can easily retrieves the data which he tries to steal, and can comparison with other snapshot data collected at different releases done at a time.

To prevent the privacy problems, this project proposes novel k^w-structural diversity anonymity, where k denotes number of friend user connected with the particular user and w denotes the snapshot of action done at every time period of updation in the network. This project also presents a heuristic algorithm for generating releases satisfying k^w-structural diversity anonymity due to this knowledge about the victim cannot utilized by the adversary for re-identification and to take advantages.

3.RELATED WORKS

3.1.The seed and grow system

In social networking service, users loose some digital information even after performing anonymization process[2]. To make the user alert when this type attacks occurs an algorithm known as seed and grow is introduced. In seed and grow algorithm sub graph is considered as the seed which is placed by the adversary and it grows larger tree according to the information gained previously from different social sites.

3.2.Anonimos

For preserving linear properties[3] of graphs that are present in social data a linear programming based technique called anonimos is introduced to solve the shortest path problems**.**

3.3.Closeness

To overcome the limitations of l-diversity model t-closeness technique is used which make close relationship of distributed sensitive attribute in any equivalence class with the other attribute in overall table[4].

3.4.kACTUS

It is K-anonymity of Classification Trees Using Suppression(kACTUS)[5] model which performs multidimensional suppression of certain records which depends on other attribute values, without any need of domain hierarchy trees that are produced manually.

3.5.k-join-anonymity (KJA)

KJA is used for effective generalization in public database to reduce the information loss[6]. KJA contains two methodologies, first generalization of micro table and public database. The second anonymization of micro table, and finally refines the resulting groups using public database.

4.PROBLEM FORMULATION

Social network is the only place which performs updation of information each and every second according to the latest trends follows up. At each updation performed in online groups, the information becomes more powerful and useful for the business people who are in the form of sellers, customers and the person who links both sellers and customers in the field of shopping carts. These data's can be used in different applications of social activity, telecommunication, advertisement companies, etc. As long as the social activities grow, privacy for the information is needed a lot.

Privacy preservation plays major role in social network for protecting sensitive information of each individual in the network. Social network is the place where tons and tons of information present, at some situations this becomes tedious to preserve from attackers who launch attack and tires to capture preserved data. Adversaries are the person who attacks the user's information shared over the network and victims are the users who are targeted to attack [7]. There is no solution to preserve victim's vertex identity and community identity which is attacked by the adversary. Vertex identity denotes the personal information and identification of the user. Multi community identity denotes the community groups in which the user participated.

A fine-tuned approach for solving the disclosure problem is to apply anonymization algorithms for each snapshot recorded when modification is done. Every login of the user in the network is updated sequentially at every timestamp which is known as release. Each release is anonymized before it is published over the network to provide privacy for the users.

In this project, the vertex identity and community identity disclosure problem present in a social network is considered and studied. The considerations of multi community identity and sequential publications demand a new dynamic privacy model, since the existing privacy model such as k-structural diversity cannot ensure privacy in both situations. For protecting against such an adversary, the proposed system introduces a new dynamic privacy scheme, named dynamic K^w-structural diversity anonymity. This technique assumes, attacker can watch the user within time

snot by knowing number of friends connected with the particular user, which is termed as degree attack

5.PRIVACY PRESERVATION MODELS

Social network is modeled as the graph[8] with vertices, edges and relationship between the vertices. According to dynamic network, relationship between the individuals changes when time changes literally. Therefore in this paper the dynamic network is modeled as a time-stamped graph. Dynamic network is specified as $G^t(V^t,E^t,C^t)$ at time instance t, where $V^t$is the set of vertices of corresponding individuals, E^t denotes set of edges which represents relationship between the individuals, C^t denotes the community of each individuals belongs to.

Before proposing privacy model[9] for sequential releases of dynamic graph, assumed knowledge of adversary is defined. An adversary can monitor a victim at a time period and can extract knowledge includes releases of graph data and degree sequence of a victim at a time period.

For problem of timestamp consider the least time stamp as the starting time(w=1) which provides common privacy level for the collected information, for constructing more privacy levels use the most largest timestamp created with the data collected(w>1) for achieving high privacy level for both vertex and multi community identities.

6.IDENTITY PROTECTION

There are two kinds of identity mentioned in this paper, First one is vertex identity another one is multi community identity [10][11]. For protecting vertex identity at a time period number of friend nodes attached with the particular user node should be extended, to make an attacker tedious to steal information. Along with this each user node should share same degree sequence for protection. Next is to protect multi community identities, here each user node is added at disjoint groups at each time period. Based on these concerns, k-shielding consistent group is defined first and then propose the new privacy model, dynamic k^w-structural diversity.

Multi community identity can be protected against degree attack using two ways, first one is every user should share the same degree (number of friends) for ensuring protection. Second one is each and every vertex should spread privacy for all the nearby users. In other words, users with same number of friends and also involved in same community groups can confuse the attacker by diversing the degree of the users. Based on these observations, k-shielding group is computed, which defines that each user have same number of friend user and all involved in the same community groups.

7.PROPOSED SYSTEM

Anonymization plays a major role in privacy preservation. It is the process making the data or information general or includes fake data to hide the sensitive information. There are two approaches present for anonymizing graph. They are named as graph generalization[12] and graph modification. These two techniques are used to prevent privacy leakage while publishing in a network. Graph generalization is a process of clustering similar vertices into super vertices and

provides general description for numbering the vertices. The generalized graph is used for random graph construction using general description.

One of the main anonymization process is, once a graph is created with the information present over the network, it is modified according to the privacy level needed against the attack models (models like degree attack and neighborhood attack). Although the sensitive data is preserved from the attacker, originality of the data needs to be (degree distribution or average shortest path length). The graph anonymization can be obtained by performing AddingEdge, RedirectingEdge and AddingVertex process to increase the utility of the sensitive data.

Before anonymization		
Age	Sex	Zipcode
42	Female	638052
48	Male	638103
50	Male	638156
46	Female	638562
32	Female	638002
26	Male	638521
20	Male	638559
31	Female	638221

After anonymization		
40-50	People	6*****
40-50	People	6*****
40-50	People	6*****
40-50	People	6*****
30-40	People	6*****
30-40	People	6*****
20-30	People	6*****
20-30	People	6*****

Table 1.Comparison of anonymization table before and after anonymization

As Table 1.shows the anonymization of private data of a peoples address and gender from various area. Attacker who are in the need of information about the particular person can easily retrieved therefore generalization process is carried out to group the people based on their age groups. This helps to hide the age and location of particular person.

7.1. Dynamic K^W-Structural Diversity Anonymity

A new privacy model is introduced to protect vertex and multicommunity identity called k^w-structural diversity anonymity[13][14], where k is an appreciated privacy level and w is a time which an attacker can trace the victim. The model K^w – SDA makes the common vertex and multi community information in generalized form of 1/k probability. This is for securing the data from the attacker. After that, a scalable heuristic algorithm is developed to provide dynamic kw-SDA. The proposed algorithm has capability to anonymize the graph based on the previous w-1 releases and minimize the graph alterations. While performing this anonymization algorithm, much of the

dynamic network characteristics can be retained with low data loss by entering the vertex and multi community details in CS table. dynamic network characteristics can be retained with low data loss by entering the vertex and multi community details in CS table.

7.1.1. The CS-Table

The CS-table is a cluster sequence table which contains three columns with vertex, degree sequence and sequence of multicommunity identities. After the CS-Table is constructed it is dynamically updated according to the time slots when the user enters the communication network

7.1.2.CS-Table Construction

CS-Table is used to collect the degree sequences and multi-community identities of vertices in the previous releases of anonymization. K^w-SDA requires that every vertex belongs to a k-shielding consistent group. Main purpose for building CS table is to collect all the vertex and multicommunity information in the form of table for avoiding repeated scanning of same release in the social network. Due to this data utility is increased by limiting information loss.

K-Shielding Groups

Let a group contains set of vertices with same degree at a time. Each vertex node present in the social network graph can have multicommunity identity of its own node. The set of communities of a vertex participate at a time t. A group is said to be k-shielding if it has a subset of it with size k, such that the multicommunity identities of any two vertices should be disjoint of each other.

7.1.3.CS-TableUpdation

Once the cluster sequence table is created by ordering the vertex information of the user(unique id), degree sequence of number of friend user connected with the particular user and multicommunity information about the user involved group details. At each action carried out in the social network, the information present in the CS- table is automatically updated to forbid the continuous scanning of the releases.

K-Shielding Consistent Group

A group of user nodes uses same degree sequence at a particular time t at a snapshot period w. This group is denoted as k- shielding at the snapshot period w, at each time t in w period, the user node forms the subset of k size, then multicommunity identity of two user nodes present in subset becomes disjoint sets.

7.1.4. Anonymization Process

To summarize the vertex information and to generate a privacy-preserving release from the constructed CS-Table need to anonymize the vertices according to their ranking in CS-Table.
Operations for anonymization process [15], three operations are used to adjust the degree of a vertex.

1. **AddingEdge**- This operation used to collaborate two user nodes present in same community group.

2. **RedirectingEdge-** This operation is used to increment the degree of the user node by not-yet-anonymized end-point of before added relationship of the vertex.
3. **AddingVertex-**Make a vertex which is a user connected along with the duplicate vertex for preservation.

8.CONCLUSIONS

Privacy issues become more important in social network when data sharing and communication plays a wide role in the network. Main problem faced by the user of social network is identity disclosures. Two types of disclosure mainly occurred in social network are vertex identity disclosure which reveals the vertex information of user and multi community identity disclosure which reveals the group information of the user participated. To overcome these disclosures a new algorithm k^w- structural diversity anonymity(k^w- SDA) is introduced for protecting user node and multi community groups in which the user participate in sequential snapshot of a social network site. In To achieve k^w-SDA,large-scale dynamic networks are anonymized with limited information distortion. In k^w- structural diversity anonymity, k denotes the number of user in social network for applying privacy level and w denotes the tracing time period of an adversary. In addition, a summary table named CS-Table is created, to summarize the vertex information to improve efficiency.

REFERENCES

[1] S. Bhagat, B. Krishnamurthy, G. Cormode, and D. Srivastava, "Prediction Promotes Privacy in Dynamic Networks," Proc. Third Conf. Online Social Networks (WOSN), 2010.

[2] Wei Peng, Student Member, IEEE, Feng Li, Member, IEEE, XukaiZou, Member, IEEE, and Jie Wu, Fellow, IEEE," A Two-Stage Deanonymization Attack against Anonymized Social Networks", IEEE Transactions On Computers, Vol. 63, No. 2, February 2014.

[3] Sudipto Das, omerEgecioglu, and Amr El Abbadi, Senior Member IEEE,"Anonimos: An LP-Based Approach for Anonymizing Weighted Social Network Graphs", IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No. 4, April 2012.

[4] Ninghui Li, Member, IEEE, Tiancheng Li, and Suresh Venkatasubramanian , " Closeness: A New Privacy Measure for Data Publishing" IEEE Transactions On Knowledge And Data Engineering, Vol. 22, No. 7, July 2010.

[5] SlavaKisilevich, LiorRokach, Yuval Elovici, Member, IEEE, and BrachaShapira, "Efficient Multidimensional Suppression for K-Anonymity," IEEE Transactions On Knowledge And Data Engineering, Vol. 22, No. 3, March 2010.

[6] DimitrisSacharidis, KyriakosMouratidis, and DimitrisPapadias," k-Anonymity in the Presence of External Databases" IEEE Transactions On Knowledge And Data Engineering, Vol. 22, No. 3, March 2010.

[7] K. Liu and E. Terzi, "Towards Identity Anonymization on Graphs," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2008.

[8] X. Wu, X. Ying, K. Liu, and L. Chen, "A Survey of Privacy- Preservation of Graphs and Networks," Managing and Mining Graph Data, vol. 40, pp. 421-453, 2010.

[9] B. Zhou, J. Pei, and W. Luk, "A Brief Survey on Anonymization Techniques for Privacy Preserving Publishing of Network Data,"ACM SIGKDD Explorations, vol. 10, pp. 12-22, 2008.

[10] K. Liu and E. Terzi, "Towards Identity Anonymization on Graphs," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2008.

[11] C.-H. Tai, P.S. Yu, D.-N. Yang, and M.-S. Chen, "Structural Diversity for Privacy in Publishing Social Networks," SIAM Int'l Conf. Data Mining (SDM), 2011.

[12] M. Hay, G. Miklau, D. Jensen, D. Towsley, and P. Weis, "Resisting Structural Re-Identification in Anonymized Networks," Proc. VLDB Endowment, vol. 1, pp. 102-114, 2008.

[13] L. Tang, H. Liu, J. Zhang, and Z. Nazeri, "Community Evolution in Dynamic Multi-Mode Networks," Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 677-685, 2008.

[14] R. Kumar, J. Novak, and A. Tomkins, "Structure and Evolution of Online Networks," Proc. 12th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), 2006.

[15] Ninghui Li, Member, IEEE, Tiancheng Li, and Suresh Venkatasubramanian, " Closeness: A New Privacy Measure for Data Publishing" IEEE Transactions On Knowledge And Data Engineering, Vol. 22, No. 7, July 2010.

Authors

R.Gowthamy completed her B.E. degree in Computer Science and Engineering from Velalar College Of Engineering And Technology Erode, Indian 2014. She is currently doing her M.E(Computer Science and Engineering) in Nandha Engineering College(Autonomous), Erode, India.

P.Uma completed her B.Sc degree in Computer Science from Erode Arts college for Women, Erode, India in 2000. She completed her M.Sc degree in Computer Science from Navarasam Arts And Science College For Women, Arachalur,India in 2002. She completed M.E. degree in Computer Science and engineering from Kongu Engineering college, Perundurai, India in 2008. Presently she is working as Assistant Professor in Computer Science and Engineering Department in Nandha Engineering College (Autonomous), Erode, India.