International Journal of Emerging Trend in Engineering and Basic Sciences (IJEEBS)
ISSN (Online) 2349-6967 *Volume 2 , Issue 1(Jan-Feb 2015), PP1-5*

Privacy Preserving in Online Social Network

Mr. Swapnil Patil[1], Mr. Nihalahmad R.Shikalgar[2]
[1]*(UG Student, Department of Computer Engineering,*
Adarsh Institute of Technology & Research Centre, Vita, India)
[2]*(Assistant Professor, Department of Computer Engineering,*
Adarsh Institute of Technology & Research Centre, Vita, India)

Abstract :- The Social networking have become an important part of the online activities on the networking for publishing or sharing social data on social networking, research and business analysis Social networks offer to web users new interesting means to communicate, interact, and socialize. Privacy-related issues are very important in social networking because, the ultimate challenge is how to prevent privacy attack when much personal information is available. This paper we present a rigorous a rigorous technique AES algorithm to encrypt sensitive attributes and attribute names. We used unique token (key) per user, therefore, prevents potential leaks of sensitive labels and information associated with them. Sensitive labels or user information is secured it will be in encrypted format.

Keywords: - Secure communication in OSN, Protecting privacy, protecting sensitive information.

I. INTRODUCTION

In today's internet determined the people we have witnessed the rapid growth of online social networking sites (OSN) as well as their integration into our everyday life. OSN such as Facebook (FB), Twitter, LinkedIn, MySpace etc. now represent a fundamental shift in the way that we communicate in our personal and working live. With the sharing nature of OSN's and the sites' control of posted information and personal relationships, concerns have developed regarding trust and privacy issues within social networking. Mainly, the data may contain sensitive information about individuals that cannot be disclosed without compromising their confidentiality. This paper we use AES algorithm to encrypt sensitive attributes and attribute names. We used unique token (key) per user, therefore, prevents potential leaks of sensitive labels and information associated with them. Because it will be publish in encrypted format.

We consider a graph model in every vertex of graph is linked with sensitive labels or private information. We develop a new algorithm (heuristic search) by adding noise nodes into the original graph without change original graph drastically, and provide security of each user & its sensitive data.

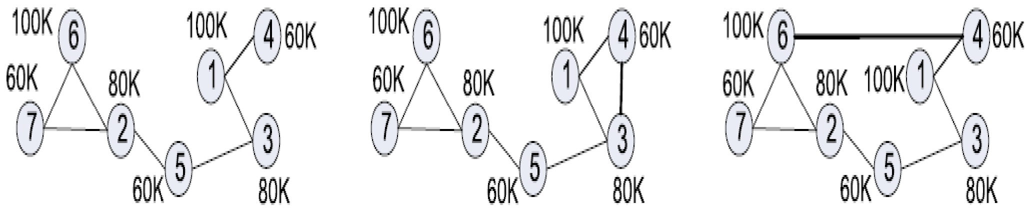
Fig. 1a shows an example of a possible structure attack using degree collect the information. If an adversary knows that one person has three friends in a graph, he can know that node 2 is that person and the related attributes of node 2 are revealed. K-degree anonymity can be used to prevent such structured attacks in SN. However, in many applications in, a social network where each node has sensitive attributes should be published. For example, a graph may contain the user salaries which are sensitive. In this case, k-degree alone is not sufficient to prevent the inference of sensitive attributes of individuals. Fig. 1b shows a graph that satisfies 2-degree anonymity but node labels are not consider in a graph. In it, nodes 2 and 3 have the same degree 3, but they both have the label "80K." If an attacker knows someone has three friends in the social networks, he can conclude that this person's salary is 80K without exactly re-identify the node. Therefore, when sensitive labels are considered, the l-diversity should be adopted for graphs. Again, the l-diversity concept here has the same meaning as that defined over tabular data.

II. LITERATURE REVIEW

Online social Networks have always been an important component of our daily life, but currently that more and more people are connected to the Internet, and their online counterpart is satisfying an increasingly vital role. Consider a graph model where each vertex in the graph is associated with as the sensitive label or [private information]. Many varieties of privacy models and anonymization algorithms have been developed (like as k-anonymity, l-diversity, t-closeness). In tabular micro-data, some of the non-sensitive attributes, called quasi identifiers, can be used to identify individuals and their sensitive labels. Use in this paper graph properties with adding noise node into the graph. We conducted intensive experiments to evaluate the impact of anonymization on the classification on future data. Experiments on real life data show that the quality of classification can be preserved even for highly preventive anonymity requirements of OSNs Privacy Benjamin. Any node, there exists at least other k! 1 node has the same degree as this node.

Growing must needs to address privacy concern when social network data is released for mining purposes has recently led to significant interest in various techniques for graph anonymization in social network. We are using AES encryption technique for protecting sensitive labels in social network data anonymization it is more secure privacy preserving approach in social network as compare to other research. Proposed a Line is Programming-based method to protect the edge weights while preserving the path of shortest paths. These two works focused on the protection of edge weights instead of nodes, which are different with our work.

- In previous research in edge editing and node re-identification of social networking graph using graph properties.
- It protects social network data and provides privacy to the user.
- But it change original graph.
- It tested on certain limit on nodes in social network.
- It does not work with distributed environment in network.
- Our new research is work with distributed environment.



a) Original graph SN b) 2-degree anonymous graph SN c) 2-degree 2-diversity graph SN

- Adding nose only makes it harder for attackers to gain info about of sensitive labels of user.

Fig.1 Publishes A Graph With Degree And Label Anonymity.

- It is not privacy-preserving approach

III. PROBLEM DESCRIPTION

In this paper, a social network graph is defined as follows:

A. Definition 1: Social Network Graph- in a social network graph is a four tuple G (V, E, σ, ⋋) where V is a set of vertices in graph, and each vertex represents a node in the social network. E ⊆ V × V is the set of edges between vertices, σ is a set of labels that vertices have. ⋋: V→ σ maps vertices to their table.

B. In this paper, we use the words “node” and “vertex” interchangeably. In a published (privacy preserving) social networking graph, an attacker could reidentify a node by degree information and further infer sensitive labels into graph. To prevent this possible leakage, we define “k-degree-l-diversity” principle for published graphs, which have the same spirit of k − l diversity in relational data

Definition 2 (KDLD) For each vertex in the graph, there exists at least k -1 other vertices having the same degree in a graph. Moreover the vertices with the same degree contain at least l distinct sensitive labels. We use distinct l-diversity to ensure that there exist at least l-distinct labels in each equivalent class (group), i.e., a group of nodes having the same degree. We use distinct l-diversity to demonstrate the basic working procedure of our method for SN. We give the detailed discussion about how to handle more complex l-diversity such as recursive diversity. We call a graph a KDLD graph if it satisfies the k-degree-l-diversity constraint. A KDLD graph protects two aspects of each user when an attacker uses degree information to attack: 1) The probability that an attacker can correctly re-identify this user is at most 1/k; 2) The sensitive labels of this user can at least be related with l- Different values. Since each equivalent class contains at least k nodes, when an attacker uses the degree to re-identifies a node, the probability he correctly re-identifies this user is at most 1/k in graph. Furthermore, since there are at least l distinct labels in each equivalent class in node, a user’s sensitive label is related to at least l values. Our goal is to protect each user’s privacy by adding certain edges nodes to transfer a social network graph to a K-Degree L-Diversity graph. We refer the added edges/nodes as noise edges/nodes.

IV. IMPLEMENTATION DETAILS

Anonymization is a clustering problem one or more nodes are connected each other in various graph in social network and sharing information and resources in social networking business as well as personal relations. In figure shows Spearman coefficient correlation it is described by monotonic function there are repeated data values, a perfect spearman correlation of +1 or -1 occurs when each of variables is monotone function. Groups of columns based of spearman are used for columns and total no. of records.

L-diversity should be checked in the social network graph. Portion table is used for variables name, variables position in the ascending order and rank of the variables. Main approach is to preserving social network individual privacy.

A. MATHEMATICAL MODEL DESIGN:

Spearman’s correlation based L-diversity privacy preserving in social network

- Let, D is with social network database with total P records & S columns.
- Let, {R1, R2, R3 ….…. R p.} Be total number of P records.
  Let {C1, C2, C3……….5.} be total no. of S columns.

B. **EQUATIONS:**

Spearman’s correlative coefficient is given by,

$$F = \frac{\sum_i (C_{1i} - \bar{c_1}) - (c_{2i} - \bar{c_2})}{\sqrt{\sum_i (c_{1i} - \bar{c_1})} \sum_i (c_{2i} - \bar{c_2})} \quad (1)$$

Assigned a rank equal to average of their positions in the ascending order of the values of Spearman’s rank correlation coefficient is a reliable and simple method of testing both the strength and direction (positive or negative) of any correlation between two variables. The value should be between -1 (*perfect negative correlation*) and +1 (*perfect positive correlation*). This is not the end, however as you must now test to see how likely it is that your calculation is not just the result of chance this is called as significance testing. It considering your result in relation how much data you had (these are the degree of freedom).

Following figure shows Spearman’s correlative coefficient in positive, negative and no relation states.

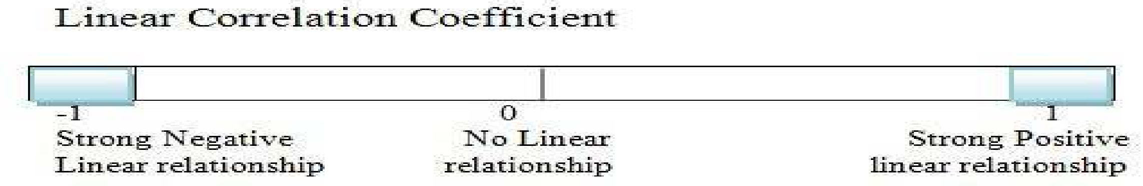


Fig.2. Spearman’s Correlative coefficient in positive, negative and no relation states

C. **PROBLEM DEFINITION:**

Protecting privacy in social networking individual and public using anonymization and diversity technique with AES algorithm, graph searching etc.

D. **ALGORITHMIC STRATEGY:**

- **Step1:** Find the total correlation between various columns of set ‘S’ using equation 1.
- **Step2:** Select the value of K randomly.
- **Step3:** Assign the K random values from set S to K cluster such that every cluster gets one value.
- **Step4:** Classify columns in cluster based on (1-p) values.
- **Step5:** Compute new centroids and update member accordingly.
- **Step6:** Continue steps 4 & 5 until centroids don’t move.
- **Step7:** Thus we break the correlation among columns now problem the tables based on the cluster.
- **Step8:** Calculate L-diversity of those centroids & check for privacy.

To generate a KDLD sequence is calculate, the triples P in P should be divided into groups in a graph. All the corresponding nodes in the same group shall be adjusted to have the same degree in a graph.

E. **ALGORITHM TWO: - STEPS**

We are using the following algorithm to construct the published graph which preserves the AP L. The algorithm contains five steps are as follows: Adding, editing and deleting extra noise nodes in original graph.

Step 1: Neighborhood Edge Editing ()

We add or delete some edges if the corresponding edge-editing operation follows the neighbourhood rule. By doing this, the sensitive degree sequence P of original graph G is closer to P new in case AP L is preserved;

Step 2: Adding Node Decrease Degree ()

For any node whose degree is larger han its target degree in P new , we decrease its degree to the target degree by making using of noise nodes;

Step 3: Adding Node Increase Degree ()

For any node whose degree is smaller than its target degree in P new, we increase its degree to the target degree by making using of noise nodes;

Step 4: New Node Degree Setting ()

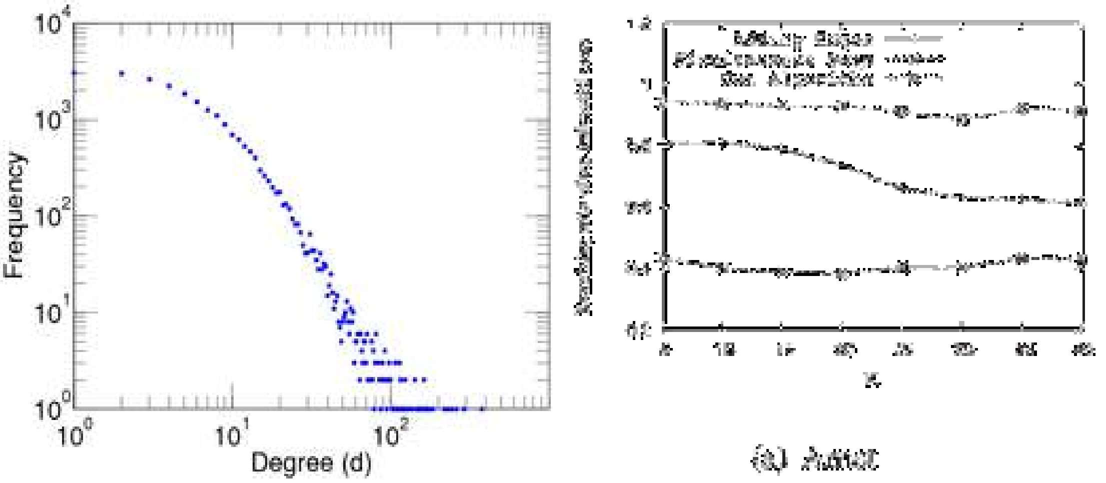
For any noise node, if its degree does not appear in P new, we do some adjustment to make it has a degree in P new. Then, the noise nodes are added into the same degree groups in P new;

Step 5: New Node Label Setting ()

We assigning sensitive labels to noise nodes to make sure all the same degree groups still satisfy the requirement of the distinct l -diversity. It is obvious that after Step 4 and Step 5, the sensitive degree sequence P' of the published graph G0 is a KDLD sequence. In Steps 2, 3, and 4, we carefully connect the noise nodes into the graph to make the change of AP L as less as possible in a graph.

V. EXPERIMENTAL RESULTS

The effectiveness of our Anonymization AES algorithm, we compare our work with two pure edge-editing graph construction algorithms: adding edges and graph construction algorithm [1]. We generate the KDLD graph for each data set using the K-L-BASED sensitive degree sequence generation algorithm. Note here we use different graph construction algorithms to generate anonym zed graphs for the same KDLD sequence. In this experiment we exam our algorithm on three real data set – Arnet (Nodes and Edges), Cora data set (Nodes and Edges), and DBLP data set (Nodes and Edges) Details of these data set can be found in online supplement material and results. With the help of noise node adding algorithm



(a) Arnet

Fig.3 Cora and Arnet Dataset

VI. CONCLUSIONS

We propose a k-degree-l-diversity model for privacy preserving social network data publishing. We implement distinct K-degree, l-diversity and Anonymization. We design an algorithm to preserving privacy of user on social network. We are using heuristic search strategy that will search the input phase with minimum overhead. With give approximate answer within polynomial time. We give a rigorous analysis of the theoretical bounds on the minimum number of noise nodes added. Extensive experimental results demonstrate that the add minimum noise node AES algorithms and heuristic strategy can achieve a better result than the previous work using edge editing only and noise node adding attractive direction to study clever algorithms which can reduce the reduction of noise nodes with Anonymization and diversity. Privacy is key matter when sharing social network data for organization and personal. It is necessary of today's large use of social network to provide privacy and security of private information. We present new technique that will reduce noise nodes in our model.

- Add minimum no of nodes & improve Anonymization technique.
- We implementing privacy-preserving approach.

It is designed to help out these publishers publish an integrated data together to certification the security and privacy.

REFERENCES

[1] K. Le-Fevre, D. DeWitt, R. Ramakrishnan. *Mondrian multidimensional k-anonymity* In International Conference on Data Engineering 2006.

[2] M. Hay, G. Miklau, D. Jensen, D. Towsley, P. Weis, "*Resisting Structural re-Identification in the Anonymized Social Networks*," Proc. VLDB Endowment, vol. 1, pp. 102-114, 2008.

[3] B. Zhou and J. Pei, *"Preserving Privacy in Social Networks Against Neighborhood Attacks,"* Proc. IEEE 24th Int'l Conf. Data Eng. (ICDE '08), pp. 506-515, 2008. 08), 2008.

[4] Hay, Michael; Miklau, Gerome; Jensen, David; Weis, Philipp; and Srivastava, Siddharth, "*Anonymizing Social Networks*" (2007).Computer Science Department Faculty Publication Series paper 180.

[5] K. Le-Fevre, D. DeWitt, R. Ramakrishnan. *Mondrian multidimensional k-anonymity In International Conference on Data Engineering 2006*

[6] A. Meyerson and R. Williams. *On the complexity of optimal k-anonymity in ACM Symposium on Principles of Database Systems 2004*

[7] B.S. Hettich and C. Merz. *UCI repository of machine learning databases, 1998*

[8] P. Samarat,- Protecting respondent's privacy in micro data release IEEE Transactions on Knowledge and Data Engineering, 13, 2001.

[9] L. sweeney, achieving k-anonymity privacy protection using generalization and suppression. International journal on uncertainty, Fuzziness and knowledge based system, 2002.

[10] A.-L. Baraba´ si and R. Albert, "Emergence of Scaling in Random Networks," Science, vol. 286, pp. 509-512, 1999.

[11] Bruce Kapron, Gautam Srivastava, S. Venkatesh -IEEE international Conference 2011, *Social Network anonymization via Edge Addition.*

[12] Benjamin C. M. Fung, Ke Wang, and Philip S.Yu, Fellow, IEEE Data Engineering 2007 *Anonymizing Classication Data for Privacy Preservation.*

[13] Ping Xiong, Tianqing Zhu management of e-Commerce and e Government (ICMeCG), 2012 Conference on An Anonymization *Method Based on Tradeoff between Utility and Privacy for Data Publishing.*

[14] Gionis A.; Tassa, T, IEEE Knowledge and data engineering 2009.*K anonymization with minimal loss of information.*

[15] Shapiro, S S. (SysCon) IEEE Knowledge and data engineering 2012, *Situating Anonymization within a Privacy risk model.*