Privacy in Online Social Networks: A Survey

1Subhashri C 2Nisha M 3Sindhiya R 4Karpagam V
1,2,3M.E Scholars 4Assistant Professor
1,2,3,4Department of Computer Science Engineering
1,2,3,4K.L.N. College of Engineering, Pottapalayam, Sivagangai 630612, India

Abstract

Knowledge and Data Engineering is the entire process of applying computer-based methodology, including new techniques for knowledge discovery, from data. It refers to the discovery of new information in terms of patterns or rules from vast amounts of data and also maintains privacy of data. Online social network have qualified immense growth in recent years and used by millions of people also.OSN is a platform which creates social relationship among users to share their interests, activities and real life connections but also raising a many privacy and security issues. While OSNs allow users to restrict access to shared data, they currently do not provide any mechanism to enforce privacy concerns over data associated with multiple users. This paper survives many techniques used in OSN privacy issues. Also discuss the pros and cons of these approaches.

Keyword- Online social networks, data privacy, social networking, privacy protection algorithms

I. INTRODUCTION

Data warehousing and Data mining used for decision making systems, Knowledge Discovery process and various techniques. Data mining tools perform data analysis and may uncover important data patterns, contributing greatly to business strategies, knowledge bases, and scientific and medical research. Data warehousing used to information processing by providing a solid platform of consolidated, historical data for analysis.

- Data mining is the process of using “raw” data to infer important “business” relationships.
- Despite a consensus on the value of data mining, a great deal of confusion exists about what it is.
- Data Mining is a collection of powerful techniques intended for analysing large amounts of data.
- There is no single data mining approach, but rather a set of techniques that can be used stand alone or in combination with each other.

Social networking sites are involved the use of internet to connect users with friends, family and acquaintances. Social media sites have a variety of privacy settings you can adjust. This means you can control who sees your profile page and other information you share on the site. Some people do not mind having their personal information available for anyone to view online. Researchers from different computer science disciplines have handled some of the problems that arise in OSN and propose a different range of privacy solutions, including software tools and design principles. Each of these solutions is developed with a specific type of user and privacy problem in mind. To tackle this problem we have to aware of the privacy issues. Private matching protocols are helpful to provide different levels of privacy to users. Access control mechanism is used to provide privacy to the data viewed by specific user .Recent technique for providing privacy is labelling which classifies the sensitive and non-sensitive labels. So labelling method is efficient and best suited for privacy problems in OSN network.

II. APPLICATIONS OF KNOWLEDGE AND DATA ENGINEERING

In recent years, Knowledge and data Engineering has been viewed as threat to privacy because of the widespread proliferation of electronic data maintained by corporations.

This has led to increased concerns about the privacy of the underlying data. In recent years, a number of techniques have been proposed for modifying or transforming the data in such a way so as to preserve privacy. Privacy-preserving data mining finds numerous applications in surveillance which are naturally supposed to be “privacy-violating” applications. The key is to design methods which continue to be effective, without compromising security.

Most methods for privacy computations use some form of transformation on the data in order to perform the privacy preservation. Typically, such methods reduce the granularity of representation in order to reduce the privacy. This reduction in granularity results in some loss of effectiveness of data management or mining algorithms. This is the natural trade-off between information loss and privacy. Some methods used for privacy issues are,

A. The randomization method
The randomization method is a technique for privacy-preserving knowledge and data engineering in which noise is added to the data in order to mask the attribute values of records.
The noise added is sufficiently large so that individual record values cannot be recovered. Therefore, techniques are designed to derive aggregate distributions from the perturbed records. Subsequently, knowledge and data engineering techniques can be developed in order to work with these aggregate distributions.

B. Distributed privacy preservation
In many cases, individual entities may wish to derive aggregate results from data sets which are partitioned across these entities. Such partitioning may be horizontal (when the records are distributed across multiple entities) or vertical (when the attributes are distributed across multiple entities).
While the individual entities may not desire to share their entire data sets, they may consent to limited information sharing with the use of a variety of protocols.
The overall effect of such methods is to maintain privacy for each individual entity, while deriving aggregate results over the entire data.

C. Downgrading Application Effectiveness
In many cases, even though the data may not be available, the output of applications such as association rule mining, classification or query processing may result in violations of privacy.
This has led to research in downgrading the effectiveness of applications by either data or application modifications. Some examples of such techniques include association rule hiding, classifier downgrading, and query auditing.

D. The k-anonymity model and l-diversity
The k-anonymity model was developed because of the possibility of indirect identification of records from public databases. This is because combinations of record attributes can be used to exactly identify individual records. In the k-anonymity method, we reduce the granularity of data representation with the use of techniques such as generalization and suppression.
This granularity is reduced sufficiently that any given record maps onto at least k other records in the data. The l-diversity model was designed to handle some weaknesses in the k-anonymity model since protecting identities to the level of k-individuals is not the same as protecting the corresponding sensitive values, especially when there is homogeneity of sensitive values within a group.

III. LITERATURE SURVEY

A. Fairness
Aware And Privacy –Preserving Friend Matching Protocol In Mobile Social Networks:
This paper proposed the model of secure friend discovery process privacy preserving interest and profile matching problem.
It introduces a new security threat rising from existing friend discovery protocols which leads to runaway attack and introduce a serious unfairness issue. To overcome this issue it introduces a novel Blind Vector Transformation technique which is used to hide correlation between the original vector and transformed results.
Based on this technique it proposes privacy preserving fairness-aware interest and profile matching protocol which allows one party to match its interest with the profile of another without revealing its real interest and profile and vice versa. It introduced another concept that is Light Weight Verifier Checking technique.
It used to enable the verifier to check the matching at the minimized overhead and prevent any participant from beginning the runaway attack.

![Fig. 1: Friend discovery in mobile social networks](image-url)
In fig1 Alice’s interest profile is matching with Bob’s profile but they cannot know about personal information of each other.

B. Multiparty Access Control for Online Social Network
This paper proposed an approach to enable the protection of shared data which is associated with multiple users in online social network.

It introduces an Access Control model to capture the core of multiparty authentication requirement along with Multi Party Policy Specification scheme and Policy Enforcement mechanism. The use of multiparty access control mechanism can greatly enhance the flexibility for regulating data sharing in online social network.

If user posts a comment in a friend’s space, he/she can specify which users can view the comment. It uses interface for advanced query which is used to give privacy details about data which he/she posted in their space.

Interface for owner query allows the data either private or public or viewed by a specific person. Interface for showing violated assessors used to show the persons who are all trying to view the specific data.

C. A Trust-Based Privacy-Preserving Friend Recommendation Scheme for Online Social Networks
This paper proposed trust-based privacy-preserving friend recommendation scheme for OSNs where OSN uses apply their attributes to find matched friends and create social relationships with strangers via a multi-hop trust chain.

In this paper they design a light weighted privacy preserving friend recommendation scheme for OSNs by utilizing both user’s social attributes and their existing trust relationships to create a multi-hop trust chain between strangers.

During the recommendation process they consider the privacy leakages preservation approaches regarding social attributes, identity and their trust relationships.

D. Privacy in Online Social Networks Using Labels
This paper proposed the social networking is modelled as graphs in that Users are nodes and features are labels. Labels are denoted either as sensitive label or non-sensitive label.

It handles node labels both as background knowledge an adversary may possess, and as sensitive information which has to be protected it also gives presence privacy protection algorithms which allow for graph data that to be published in a form such that an adversary who possesses information about a nodes neighbourhood cannot safely infer its identity and its sensitive labels.

Graph-Based Noisy Node Detection algorithm is used to group two nodes which is having maximum similarity of their neighbourhood labels are grouped together.

Sensitive Label Privacy Protection method used to hide the sensitive data and protect data from the unknown users.

IV. COMPARISON TABLE ON DIFFERENT PRIVACY ISSUES

<table>
<thead>
<tr>
<th>Author Name</th>
<th>Title</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haojin Zhu, Suguo Du, Muyuan Li, Zhaoyu Gao.</td>
<td>Fairness- Aware And Privacy -Preserving Friend Matching Protocol In Mobile Social Networks.</td>
<td>Matching friends without revealing the private details.</td>
<td>It is not suited for fine-grained interest/profile matching methods.</td>
</tr>
<tr>
<td>Hongxin Hu, Gail-Joon Ahn, Jan Jorgensen.</td>
<td>Multiparty Access Control For Online Social Network.</td>
<td>It gives the protection of shared data associated with multiple users in OSNs.</td>
<td>It did not provide the services for collaborative management of shared data in OSNs.</td>
</tr>
<tr>
<td>Linke Guo, Chi Zhang, Yuguang Fang.</td>
<td>A Trust-Based Privacy-preserving Friend Recommendation Scheme For Online Social Networks.</td>
<td>It establishes social relationships with strangers via a multi-hop trust chain.</td>
<td>There is possible to enter the intruders in the network.</td>
</tr>
<tr>
<td>Amarjagdaonkar, Hasib M. Shaikh</td>
<td>Privacy In Online Social Networks Using Labels.</td>
<td>It is used to hide the sensitive data from unknown users.</td>
<td>Non sensitive data viewed by all users.</td>
</tr>
</tbody>
</table>

V. CONCLUSION

In this paper we have surveyed different privacy issues arise in online social networks. The comparison between different approaches and pros and cons has also been made. Thus this work provides required information on privacy issues in online social network and techniques to handle issues.
REFERENCES