A Review on Privacy Preserving in Social Network

S. Mayil[1], Dr. M. Vanitha[2]

[1]Research scholar, PG & Research Department of Computer Science,
JJ College of Arts and Science (Autonomous), Pudukkottai, Tamil Nadu, India

[2]Assistant Professor, Ph.D. and Research Department of Computer Application,
Alagappa University, Karaikudi , Tamilnadu.

Abstract: The development of online social network and social network publishing data has led to the risk of leakage of personal confidential information. This requires privacy protection before the data is published by the network service provider. Online data privacy in social networks has been a major problem in recent years. Therefore, this research is still in its infancy. Some published research papers propose solutions to provide privacy table micro data. But these technologies cannot be directly applied to social network data, because social networks are vertex and edge of a complex graphical structure. Technologies, such as K-anonymity, its variants, and diversity-L have been applied to social networking data. Integrated K-anonymous and L-diversity techniques have also been developed to protect the privacy of social network data in a better way.

Keywords: Anonymization, K-anonymity, L-diversity, Social Network, Privacy Attacks.

1. Introduction

Because of the increased popularity of online social networking sites [1], many subscribe to social networks or social media. This results in a large amount of user data being collected and maintained by the social network service provider. Data generated through social networking services is known as social network data being published for others in certain circumstances. In one case, when a detailed analysis of user data is required and in other cases, the owner of the data must be shared with a third party, for example, the data for the advertising partner is part of the general user acceptance policy. . These data contain other valuable information about the user who helps in positioning a better social. Social network analysis is used in modern sociology, geography, economics and information science [2]. Researchers in various fields use these data for different purposes for researchers in government agencies requiring information from social networks and safety data [3]. Therefore, data must be shared or published in all of the above cases. The owner of the data can post the analysis of others, but can cause serious privacy threats.

In order to meet the needs of network data, online social network operators have shared the data collected and maintained with third parties such as advertisers, application developers and academic researchers as Facebook has thousands of third-party applications with Has been an exponential growth in this number [4]. Social network data contains sensitive and confidential information about users [5-7]. Thus, sharing the data in its original form can break personal privacy. Personal privacy is defined as "the individual's decision on what information about himself should be communicated to others, under what circumstances" [8]. When private and confidential user information is disclosed to the opponent for privacy violations occur. Therefore, it is an important research area to preserve the privacy of individuals by publishing the data collected by the users. This work has been done by some researchers in this direction.

This document is structured as follows: Section 2 describes the categories of privacy violations; secondly, the challenges of social network privacy in which data have been reported in Section 3; Techniques for privacy in social networks are already covered in Section 5; Section 6 sets out new research directions; and, finally, Section 7 concludes with a review.

2. Categories of Privacy Breach

In the social network privacy gap can be divided into three categories [9-10]:

i. Sensitive link disclosure - When a relationship is found between two people, a sensitive connection occurs. This type of information is generated when a user uses social activities of social network services.

ii. Sensitive attribute disclosure - When the attacker access to a sensitive and confidential information on the properties of the sensitive properties of the public content. Sensitive attributes can be linked to a relationship of entities and links.

iii. Identity disclosure - Disclosure when a person behind the record is exposed to identity. This type of failure leads to the disclosure of information about users and their sharing of relationships with other individuals in the network.

Because the user wants to end your data privacy from the service provider all of these violations of privacy posing as a serious threat to track, extortion and robbery mentioned. In addition to damage the image and personal reputation. There are also many examples of organizations that have conserved private data through the release of data networks such as search data from AOL [11] and accidental disclosure of user

data from examples of Netflix data attacks [12]. According to the commitment of social networks, it is necessary to address these issues. Therefore, the data must be published to a third party to ensure the privacy of the user. Therefore, the data must be anonymous or published to a third party prior to launch. But as discussed in the next section of the privacy of the preservation of social networks is difficult.

3. Challenges in Preserving Privacy in Social Network Data Publishing

Ensuring data privacy for social networking micro format data is difficult because [13]:

i. The context of knowledge modeling is the formidable opponent of micro social networking data. With micro formats of the table, the user can associate the prospective identifiers with the social networks near the social network from various information, such as vertex and edge, sub graph and graphic tags, to identify the personal identification.

ii. The loss of information is an indicator of the amount of distortion measured. In the form of the data loss micro data can be used to determine the sum of the information loss of personal records. Because the social network is a graphical structure with a set of vertices and edges, it is difficult to compare the vertex and edge social networks by comparing them individually. Anonymous and social networks of the original social network have the same number of vertices and edges that can have very different properties of the interaction, connectivity and diameter. The information and anonymous quality loss can be measured in different ways.

iii. The development of privacy technology for data protection in social networks is difficult to relational data. Flake Anonymous uses divide-and-conquer technology, while social network nodes and edge structures, any change in labels or edges can have effects on vertices and other blocks of edges.

The proposed method for chip micro can be applied directly to the social network data due to the connection between vertices of the graphics network compared to the independent nodes in the tabular data. In micro data, each tuple is independent, but the vertices of a social network are associated with edges. An adversary may use information about the network structure to violate the privacy of the user. Therefore, there is a need to develop a technology that secures the privacy of entities that publish social network data.

4. Equations Privacy Preserving Techniques – Micro Data

It does important work to form micro data to protect privacy. Models such as K-anonymous [14] [20] L-diversity [15], T-close [16] have been proposed to have shown good results in anonymity, after fig.1 summarizes the three models, their nature and shortcomings.

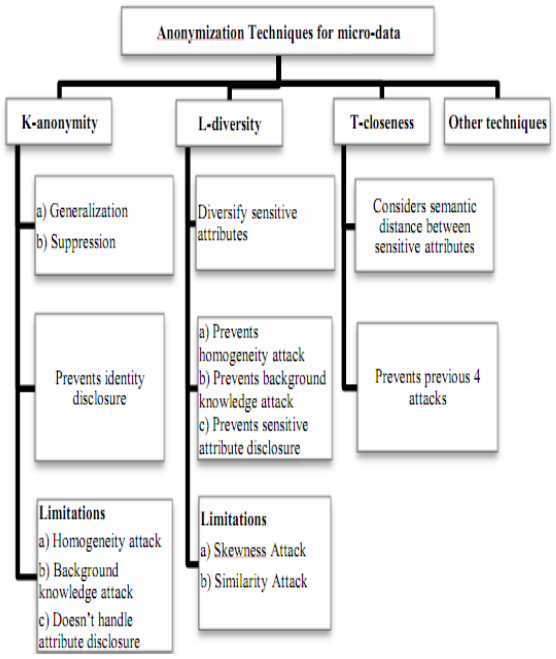

Figure 1: Existing Privacy Preserving Techniques for Tabular micro-data

5. Other recommendations Privacy Preserving Techniques – Social Networks

Existing models maintain their privacy for being used for social micro data. This work has been done using K-anonymity, L-diversity and integrated K-L-diversity anonymity, with the emphasis being on-line published to protect user data by some researchers [17]. The social network data is represented as where each node / vertex represents a single graph. The unstructured data and the edges represent the links / associations between the nodes. Fig2 shows the structure of the social network with the individual and wage 7 nodes representing the sensitive attributes shown by the label [18-20]. The technology to protect privacy is concept-based.

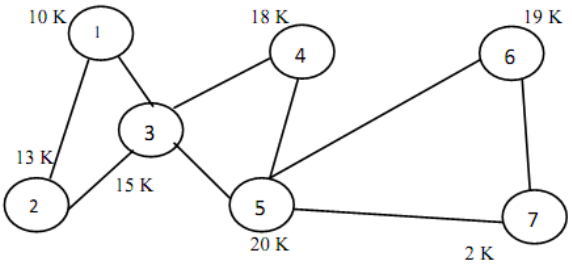

Figure 2: Social network graph with 7 nodes and sensitive attributes as salaries

The privacy of the preservation technology is developed taking into account the following points:

1. Knowledge of opponents
2. After the application data is published

Thus, according to the technique used by the opponent, the knowledge of the attacking target node has been developed by some researchers using the concept of K-anonymity [21].

The disclosure of online social networking data disclosure privacy. It has been assumed that the opponent has knowledge of the individual whiteness and proximity of the target. It has put forward a practical solution for attacking background knowledge defense. The proposed obtained anonymous social network can be used for web queries that are added in a highly accurate manner. The social network is modeled as an undirected marker. K-sub has been proposed to reduce the risk of disclosure of privacy disclosures from social network data. A K-structure based on the hypothesis that the adversary has the knowledge of the degree, the subgraph and the neighbor of the target node. proposed a graph isomorphic algorithm based on adjacency matrix. A child with at least K-1 sub-graph is no different [22-23]. K-isomorphism is used to preserve privacy when the opponent has knowledge of the child. A K-symmetric technique to prevent re-authentication using information sub-privacy. Developed an algorithm called KNAP to attack a neighborhood committee's data published on social networks. K-anonymity is added to the weighted social network. The concept of K-degree to prevent re-recognition of vertices by information vertex degree [24].

Use K-anonymous social network privacy protection to prevent disclosure of links, but still can lose the homogeneity of attacks and the privacy of the context of knowledge. In addition, K's anonymity cannot be defended by the disclosure of attributes. Thus, L-diversity was developed [25]. The use of privacy is called 'diversity of utility models, efficient definition of privacy in the details of the protection of social networks and collaborative observation of data for social network analysis of the usefulness of the impact. It has been determined that social networks L-diversity can still lose privacy as opponents can have some prior knowledge of individual sensitive attribute values before seeing the release table. See the table after the release, the opponent may have a better understanding. Gain information I, E [26]. The difference between post-knowledge and predictive factors is the loss of privacy. Therefore, it has been proposed that the input is close to the concept of proposes to keep two users, one of which can find the privacy of the relationship in the released social network data. L-type diversity has been defined as preserving the privacy of user relationships. Both algorithms are proposed for image processing, and MaxSub and MinSuper achieve L-diversity anonymity.

Then, in order to protect privacy, in order to better integrate anonymity and K-L-diversity approaches have been suggested by some authors as described later.

An integration algorithm that takes advantage of the K-anonymous and l-diversity algorithms and then evaluates the strength of the combined strengths. The algorithm has been able to increase the level of anonymity and diversification of social network users' privacy disclosure information [27]. The following anonymity and K-L-diversity properties and can deal with variations of the multisensibles anonymous period attribute of the algorithm. The algorithm corresponds to the modified form of the micro-algorithm and also depends on some modified algorithm development for the near-anonymous attack. The disadvantage of this algorithm is that it still needs some improvement to reduce the complexity so that it can be applied to large social networks [28]. The definition of the model anonymity for k-L-diversity protects sensitive tags from structural information and people [29]. The privacy for k-anonymous Many models to avoid re-recognition through the information structure of the node have been raised, but the attacker may still be able to obtain information about a person, that is, private information, the login node relationship is not well protected by anonymous means Pure structure [30]. An anonymous method has been proposed to add noise nodes by taking into account the graph-to-graph property that introduces the least distortion.

In addition to the above procedures, protection of privacy, they have proposed and developed other technologies shown in table 1.

Table 1: Various Other Privacy Preserving Techniques in Social Networks

Year	*Author*	*Brief*
2016	Bernardo Cuenca Grau et al.	Lay the foundations of privacy-preserving data publishing (PPDP) in the context of Linked Data.
2016	Zaobo He et al.	Consider a publish–subscribe system with secure proxy decryption (PSSPD) in mobile social networks.
2016	Akshaya Tupe et al.	Focus to maintain the privacy for distributed data, and also overcome the problems of M-privacy and secrecy approach with new anonymization and slicing technique
2016	Qian Wang et al.	Consider continuous publication of population statistics and design RescueDP an online aggregate monitoring framework over infinite streams with w-event privacy guarantee.
2016	Shouling Ji et al.	Present a US-based De-Anonymization (DA) framework, which iteratively deanonymizes data with accuracy guarantee. Then, to de-anonymize large-scale data without knowledge of the overlap size between the anonymized data and the auxiliary data.
2016	Shengshan Hu et al.	Describe an effective and practical privacy-preserving computation outsourcing protocol for the prevailing scale-invariant feature transform (SIFT) over massive encrypted image data.
2015	V.Vijeya Kaveri et al.	Explores the existing anonymization techniques for privacy preserving publishing of social network data.
2015	Franco Callegati et al.	Reports the early results of the privacy analysis which is being undertaken as part of the

		analysis of the clearing process in the Emilia-Romagna region, in Italy, which will compute the compensations for tickets bought from one operator and used with another.

6. Research Directions

And then from the literature to introduce:

i. Use of anonymously stored data (practical) It is very important to protect the privacy of applications. Therefore, it is necessary to develop a useful method for quantitatively measuring data. You need privacy and utility between the compensation conditions to evaluate the various technologies.

ii. Many methods such as K-anonymity, L-diversity, integrated K-anonymity, and L-diversity have been developed to maintain the privacy of user data in social networks, but the existing techniques result in significant loss of information.

iii. Anonymous technology has developed a published data network. However, many applications require regular publication of data, so it is necessary to develop security techniques that can preserve dynamic versions.

iv. They are available in distributed form data, P, to protect privacy. However, in distributed preservation of social networks, they cannot be well documented in the literature except the case of technology.

v. Social networks that are close to existing privacy have used the preservation of small data sets or aggregated data into assessments. It is necessary to carry out empirical experiments in large data sets.

vi. There is a prior art technique that can prevent attacks of uniformity of attack, background knowledge, that is, attacks due to the distance between sensitive values.

7. Conclusion

It becomes clear from the literature that user privacy is a major concern and research topic these days. Several models have maintained the privacy of the social network data through tabular micro data. Technologies such as K-anonymity, L-diversity, and K-anonymous integration of L-diversity have so far been used, but these techniques result in significant loss of information. Thus, there exists an improved preservation technique that provides privacy with minimal loss of information and better utilization of the published data.

References

[1]. Bernardo Cuenca Grau and Egor V. Kostylev by "Logical Foundations of Privacy-Preserving Publishing of Linked Data", Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016.

[2]. Zaobo He, Zhipeng Cai, Qilong Han, Weitian Tong, Limin Sun, Yingshu Li by "An energy efficient privacy-preserving content sharing scheme in mobile social networks",Springer, 2016.

[3]. Akshaya Tupe, Amrit Priyadarshi by "Data Mining with Big Data and Privacy Preservation", IJARCCE, 2016.

[4]. Qian Wang, Yan Zhang, Xiao Lu, Zhibo Wang, Zhan Qin and Kui Ren, Fellow "Real-time and Spatio-temporal Crowd-sourced Social Network Data Publishing with Differential Privacy", IEEE Transactions on Dependable and Secure Computing", 2016.

[5]. Shouling Ji, Weiqing Li, Mudhakar Srivatsa, Jing Selena He, Raheem Beyah by "General Graph Data De-Anonymization: From Mobility Traces to Social Networks", 2016.

[6]. Shengshan H, Qian Wang, Jingjun Wang, Zhan Qin, and Kui Ren, Fellow by "Securing SIFT: Privacy-preserving Outsourcing Computation of Feature Extractions over Encrypted Image Data", IEEE Transactions on Dependable and Secure Computing", 2016.

[7]. V.Vijeya Kaveri and Dr.V.Maheswari by "Cluster Based Anonymization for Privacy Preservation in Social Network Data Community", Journal of Theoretical and Applied Information Technology, 2015.

[8]. Franco Callegati, Aldo Campi, Andrea Melis, Marco Prandini, Bendert Zevenbergen by "Privacy-Preserving Design of Data Processing Systems in the Public Transport Context", Pacific Asia Journal of the Association for Information Systems Vol. 7 No. 4, pp.25-50 / December 2015.

[9] Kun Liu, Kamalika Das, Tyrone Grandison, Hillol Kargupta, "Privacy-preserving data analysis on graphs and social networks," In: Next Generation of Data Mining, pp. 419-437, 2008.

[10] E. Zheleva, L. Getoor, "Preserving the privacy of sensitive relationships in graph data," In: Privacy, Security, and Trust in KDD, Lecture Notes in Computer Science, Vol. 4890, pp 153-171, 2008.

[11] S. Hansell, "AOL removes search data on vast group of web users," New York Times, 2006.

[12] Facebook (2013, Facebook Statistic. Available: http://www.facebook.com/press/info.php/statistics

[13] Benjamin C. M. Fung, Ke Wang, Rui Chen, Philip S. Yu, "Privacy-preserving data publishing: A survey of recent developments," In: ACM Computing Surveys (CSUR), Vol. 42, pp 1-53, 2010.

[14] P. Samarati, L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," In: IEEE Transactions on Knowledge and Data Engineering, 2001.

[15] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, "l-diversity: Privacy beyond k-anonymity," In: ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 1, 2007.

[16] Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian, t-closeness: Privacy beyond k-anonymity and l-diversity," In Proc. of 23rd International Conference on Data Engineering ICDE 2007, IEEE, Istanbul, pp 106-115, 2007.

[17] Xiaoxun Sun, Hua Wang, Jiuyong Li, Traian Marius Truta, "EnhancedP-sensitive -anonymity models for privacy preserving data publishing," In: Transactions on Data Privacy, vol. 1, pp 53-66, 2008.

[18] X. Xiao, Y. Tao, "M-invariance: towards privacy preserving re-publication of dynamic datasets," In Proc. of International conference on Management of data SIGMOD '07, ACM, New York, NY, USA, pp 689-700, 2007.

[19] Bee-Chung Chen, Kristen LeFevre, Raghu Ramakrishnan, "Privacy skyline: Privacy with multidimensional adversarial knowledge," In Proc. of 33rd International conference on Very large data bases VLDB '07, pp 770-781, 2007.

[20] L. Sweeney, "k-anonymity: A model for protecting privacy," In: International Journal of Uncertainty Fuzziness and Knowledge Based Systems, Vol. 10, pp. 557-570, 2002.

[21] Qiong Wei, Yansheng Lu, "Preservation of Privacy in Publishing Social Network Data", In Proc. of International Symposium on Electronic Commerce and Security, Guangzhou City, pp 421 - 425, 2008.

[22] L. Zou, L. Chen, M. T. Ä Ozsu, "K-automorphism: A general framework for privacy preserving network publication", In Proc. of 35th International Conference on Very Large Data Base, Vol. 2, pp 946-957, 2009.

[23] B. K. Tripathy, G. K. Panda, "A New Approach to Manage Security against Neighborhood Attacks in Social Networks", In Proc. of International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Odense, pp 264 – 269, 2010.

[24] J.Cheng, AdaWai-cheeFu, Jia Liu, "K-isomorphism: privacy preserving network publication against structural attacks," In Proc. of the 2010 ACM SIGMOD International Conference on Management of data, pp. 459-470, 2010.

[25] W. Wu, Yanghua Xiao, Wei Wang, Zhenying He, Zhihui Wang "k-symmetry model for identity anonymization in social networks," In Proc. of the 13th International Cojucbybgnference on Extending Database Technology, ACM, New York, USA, pp 111-122, 2010.

[26] Lihui Lan, Hua Jin, Yang Lu, "Personalized anonymity in social networks data publication", In Proc. of IEEE International Conference on Computer Science and Automation Engineering (CSAE), Shanghai, pp 479 - 482, 2011.

[27] Maria Eleni Skarkala, Manolis Maragoudakis, Stefanos Gritzalis, Lilian Mitrou, Hannu Toivonen, Pirjo Moen, " Privacy Preservation by K-Anonymization of weighted Social Networks", ASONAM, pp 423-428. In IEEE Computer Society, 2012.

[28] K. Liu, E. Terzi, "Towards identity anonymization on graphs," In Proc. of 2008 ACM SIGMOD International conference on Management of data, Vancouver, Canada, 2008.

[29] G.K.Panda, A. Mitra, Ajay Prasad, Arjun Singh, Deepak Gour, "Applying l-Diversity in anonymizing collaborative social network" In: International Journal of Computer Science and Information Security, Vol 8, Issue 2, pp 324 - 329, 2010.

[30] Na Li, Nan Zhang, Sajal K. Das, "Relationship Privacy reservation in Publishing Online Social Networks", In Proc. of IEEE International Conference on Privacy, Security, Risk, and Trust, Boston, MA, pp 443-450, 2011.