www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 4 Issue 9 Sep 2015, Page No. 14182-14186

Paradigm of Privacy Preserving Techniques in Social Network Data Publishing

Ms.Shashikala N.Havanurkar, Mr.Abhijit V. Mophare

Department of Computer Science and Engineering,
N.B.Navale Sinhgad College of Engineering, Kegaon, Solapur, India
Email-Id: shashi7h@gmail.com

Department of Computer Science and Engineering,
N.B.Navale Sinhgad College of Engineering, Kegaon, Solapur, India
Email-Id: mophare_abhi@yahoo.com

Abstract: ***In our modern life, social networks are pervasive and omnipresent. Activities like tweeting, uploading photos on Facebook, finding job by using LinkedIn, are nothing but the use of social networks. Even the average computers are used to social networks. The incredible use of social networks in counter-terrorism and marketing has increased the need for accurate classification techniques for hiding important data and connections. However, users of admissible social networks have reasonable concerns about the use of their private data. Also, nowadays many social networks that have been partly driven by Web 2.0 applications have made publicly availability of the data. This data is analyzed in many different ways. Publishing the social network data by maintaining the privacy is a task of concern. In this paper, we present a brief review of the existing techniques for privacy preserving publishing of social network data.***

Keywords: Online social Network, Private data, Data Sharing, Access control.

1. Introduction

Due to the increase in popularity of online social networks on the Web [1], large number of people subscribe to social networks or social media. This has generated large amount of user data that is gathered and maintained by the social network service providers. The data generated by social network services is termed as the social network data that needs to be published for others in certain situations. One of the situations is when specific analysis of the user data needs to be done and another situation is when the owner of the data has to share the data with third parties like advertising partners which is part of policies generally accepted by subscribers. The data contains valuable information about users that helps third parties in better social targeting of advertisements. Social network analysis is being used in modern sociology, geography, economics, and information sciences [2]. Researchers in various fields use this data for different purposes like researchers in government institutions require social network data for information and security purposes [3]. So, data needs to be shared or published in all above mentioned situations. Owner of data can publish it for others to analyze but it may create serious privacy threats. To fulfill the demands for the network data, online social media operators have been sharing the data they gather and maintain with external third parties such as advertisers, application developers, and academic researchers like Facebook has thousands of third-party applications and there has been an exponential increase in this number [4]. Social network data contains sensitive and confidential information about the users [5-7]. Thus sharing of this data in its raw form may breach privacy of individuals. Individual privacy is defined as “the right of the individual to decide what information about himself should be communicated to others and under what circumstances” [8]. A privacy breach occurs when private and confidential information about the user is disclosed to an adversary. So, preserving privacy of individuals while publishing user’s collected data is an important research area. Work has been done by various researchers in this direction

This paper is organized as follows: Section 2 describes challenges while publishing social network data; followed by Access control in social networking sites which have been briefed in Section 3; Section 4 presents exiting techniques for preserving privacy; Section 5 gives research directions for new researchers; finally Section 7 concludes the review.

2. Challenges in Preserving Privacy in Social Network Data Publishing

Ensuring privacy for social network data is difficult than the tabular micro-data because:

1) Modeling of background knowledge of adversaries is difficult in social network data than tabular micro-data. In tabular micro-data, users are identified by linking quasi-identifiers from whereas in social network information from various sources such as labels of vertices and edges, subgraphs, and neighborhood graphs can be used to identify individuals.

2) Information loss is the metric which measures the amount of distortion. In tabular micro-data information loss can be measured using the sum of information loss in individual records. Since, a social network is a graphical structure with a set of vertices and edges hence it is difficult to compare two social networks by comparing the vertices and edges individually. Anonymized social network and original social

networks which have the same number of vertices and edges may have very different properties like betweenness, connectivity and diameter Information loss and anonymization quality can be measured in different ways.

3) Development of privacy preserving techniques in social network data is difficult than for relational data. Tabular micro-data is anonymized using divide-and-conquer techniques whereas social network is a structure of nodes and edges, any changes in labels or edges may have an effect on the neighborhoods of other vertices and edges.

3. Access Control of Social Networks

Past research on OSN security has mainly focused on privacy-preserving techniques to allow statistical analysis on social network data without compromising OSN members' privacy (see [9] for a survey on this topic). In contrast, access control for OSNs is a relatively new research area.

3.1 Body paragraphs

As far as we are aware, the only other proposals of an access control mechanism for online social networks are [10], [11] and [12]. The D-FOAF system [10] is primarily a Friend of a Friend (FOAF) ontology-based distributed identity management system for social networks, where access rights and trust delegation management are provided as additional services. In D-FOAF, relationships are associated with a trust level, which denotes the level of friendship existing between the users participating in a given relationship. Although [10] discusses only generic relationships, corresponding to the ones modeled by the FOAK: knows RDF property in the FOAF vocabulary [13], another D-FOAF-related paper [14] considers also the case of multiple relationship types. As far as access rights are concerned, they denote authorized users in terms of the minimum trust level and maximum length of the paths connecting the requester to the resource owner. In [11], authors adopt a multi-level security approach, where trust is the only parameter used to determine the security level of both users and resources. In [12], a semi-decentralized discretionary access control model and a related enforcement mechanism for controlled sharing of information in OSNs is presented.

The model allows the specification of access rules for online resources, where authorized users are denoted in terms of the relationship type, depth, and trust level existing between nodes in the network.

Compared to existing approaches, we use semantic web technologies to represent much richer forms of relationships among users, resources and actions. For example, we are able to represent access control rules that leverage relationship hierarchies and by using OWL reasoning tools, we can infer a "close friend" is also a "friend" and anything that is accessible by friend could be also accessible by a "close friend". In addition, our proposed solution could be easily adapted for very different online social networks by modifying the underlying SNKB. A further discussion on the differences between the proposed framework and the access control mechanism in [12] is provided in Section 7.4.

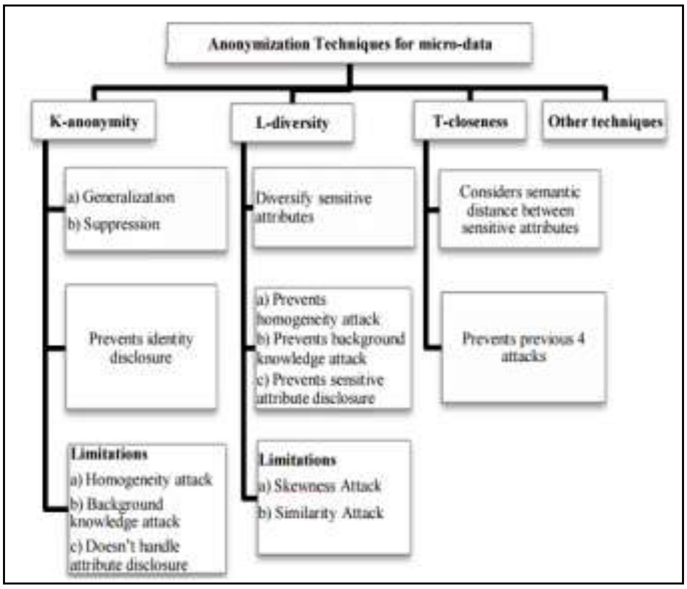
Anonymization Techniques for micro-data
K-anonymity
L-diversity
T-closeness
Other techniques
a) Generalization
b) Suppression
Diversify sensitive attributes
Considers semantic distance between sensitive attributes
Prevents identity disclosure
a) Prevents homogeneity attack
b) Prevents background knowledge attack
c) Prevents sensitive attribute disclosure
Prevents previous 4 attacks
Limitations
a) Homogeneity attack
b) Background knowledge attack
c) Doesn't handle attribute disclosure
Limitations
a) Skewness Attack
b) Similarity Attack

Figure 1: Privacy preserving Techniques for microdata

Semantic web technologies have been recently used for developing various policy and access control languages for domains different from OSNs. For example, in [15], authors compare various policy languages for distributed agent based systems that define authorization and obligation policies. In [16], OWL is used to express role-based access control policies. In [17], authors propose a semantic access control model that separates the authorization and access control management responsibilities to provide solutions for distributed and dynamic systems with heterogeneous security requirements. None of these previous work deals with the access control issues related to 9 online social networks. Among the existing work, [18] is the most similar to our proposal. Compared to [18], we provide a much richer OWL ontology for modeling various aspects of online social networks. In addition, we propose authorization, admin and filtering policies that depend on trust relationships among various users.

4. Survey on Privacy Preserving Techniques

Privacy preserving techniques are developed keeping following things into consideration:

1. Adversary's knowledge
2. Utility of the data after release

So, depending upon the knowledge that an adversary uses to attack the target node following techniques have been developed by various researchers using the notion of K-anonymity [19][20]. Wei et al. [21] considered the privacy disclosure in online social network data publishing. It has been assumed that adversaries have the knowledge of the degree of a target individual and the target's immediate neighbours. A practical solution to defend against background knowledge attacks has been proposed. Anonymized social networks obtained by proposed method can be used to answer aggregate network queries with high accuracy. Social network has been modeled as an undirected labeled graph. k-subgraph has been proposed to reduce the risk of privacy disclosure in social network data publication. Zou et al. [22] proposed k-automorphism based on the assumption that the adversary has knowledge about degree, subgraph and neighbor of the target node. Tripathy et al.[23] proposed an algorithm for graph isomorphism based on adjacency matrix. It says that a subgraph is indistinguishable from at least k-1 other subgraphs. Cheng et al. [24] used K-isomorphism to preserve privacy when adversary has subgraph knowledge. Wu et al. [25] proposed k-symmetry technique to protect privacy against re-identification using subgraph information. Lan et al.[26] developed an algorithm called KNAP against 1-neighborhood attack for publishing social networks data. Skarkala et al. [27] applied K-anonymity to weighted social networks. Liu et al. [28] proposed the concept of k-degree to prevent vertex re-identification through the information of vertex degree.

Technique	Advantages	Disadvantages
K- Anonymity	**High correlation among the tuples**	**More Number of dimensions would be violated**
l- Diversity	**Sensitive attribute would have at most same frequency**	**Homogeneity and background knowledge attack has lacked**
t- closeness	**Measure the distance between two probabilistic distribution that were indistinguishable from one another**	**Information gain was unclear**
K^m Anonymity	**Similar evaluated approach on k items**	**Loss of utility**

Table 1 : Privacy preserving Technique

Preserving privacy in social networks using k-anonymity protects against linking disclosure but still it may leak privacy under the cases of homogeneity and background knowledge attacks. Moreover, K-anonymity doesn't protect against attribute disclosure. So, L-diversity was developed by Machanavajjhala [29] in year 2007.

Panda et al. [30] used a new practical and efficient definition of privacy called l-diversity on preserving privacy in collaborative social network data and the effect on the utility of the data for social network analysis has been seen. It has been identified that l-diversity social network still may leak privacy as an adversary may have some prior knowledge about the sensitive attribute value of an individual before seeing the released table. After seeing the released table, the adversary may have a posterior knowledge. Information gain i,e., the difference between the posterior knowledge and the prior knowledge is the factor to leak privacy. So the concept of t-closeness has been suggested to be introduced. Li et al. [31] proposed to preserve relationship privacy between two users one of whom can be identified in the released social network data. l-diversity anonymization model has been defined to preserve users' relationship privacy. Two graph manipulation algorithms, MaxSub and MinSuper, have been proposed to achieve l-diversity anonymization.

Then, to preserve privacy in better way integrated approach of K-anonymity and L-diversity has been suggested by few authors as mentioned below.

Kavianpour et al. [32] proposed an integrated algorithm that takes the advantages of K-anonymity and l-diversity algorithm then evaluated the effectiveness of the combined strengths. Proposed algorithm has been able to increase the level of privacy for social network users by anonymizing and diversifying disclosed information. Tripathy et al. [33] proposed an algorithm which follows k-anonymity and l-diversity properties and can handle a variant of multisensitive attributes during anonymization process. Proposed algorithm is modified form of corresponding algorithms for micro data and it also depends upon some modified algorithms developed for anonymization against neighbourhood attacks. Drawback of proposed algorithm is that it still needs some improvements in order to reduce the complexity so that it can be applied to large social networks. Yuan et al. [34] defined a k-degree l-diversity anonymity model for the protection of structural information and sensitive labels of people. Many privacy models like k-anonymity to prevent node reidentification through structure information have been proposed but an attacker may still be able to obtain private information of a person i.e. the label-node relationship is not well protected by pure structure anonymization methods. An anonymization methodology has been proposed by adding noise nodes into the original graph with the consideration of introducing the least distortion to graph properties.

Other than above mentioned techniques for preserving privacy other techniques have also been proposed and developed as shown in table 1.

5. Direction of Research

Following are the few theorizing drawn from literature survey:

1) To conserve the functionality (usefulness) of anonymized data is an important aspect while applying techniques for privacy preservation. So, there is a need to develop methodologies that can quantitatively measure utility of data. There is need to evaluate various techniques in terms of tradeoff between privacy and utility.

2) Many algorithms like k-anonymity, L-diversity, integrated approach of k-anonymity & L-diversity have been developed for preserving privacy of social network user data but existing techniques leads to substantial information loss.

3) Anonymization techniques have been developed for one time released network data. But many applications require publishing data periodically so there is a need to develop techniques that can preserve privacy of dynamic releases.

4) Techniques are available for preserving privacy in case of distributed tabular data e.g. However, in case of social network distributed privacy preserving techniques are not well reported in literature except.

5) Existing privacy preserving approaches for social networks have been evaluated using either small datasets or synthetic datasets. There is need to conduct empirical experiments on large datasets.

6) There is no existing technique which can prevent homogeneity attacks, background knowledge attacks, attacks arising due to distance between sensitive values.

6. Conclusion

The availability of important and sensitive information has made social networking sites a potential target of attackers. Also the larger the user base, the more threat to the information hence, increasing the challenge of providing both, privacy and security to the online social networks. In this paper we have addressed the different issues of privacy and security techniques required to prevent social network data.

There are varieties of security mechanisms implemented by social networking sites for protecting the data and users. But there are also varieties of attackers trying to breach the defenses, so the social network users must be super aware of these threats and be very careful when using them.

References

[1] [1] Alexa 2013, The top 500 sites on the web, Available: http://www.alexa.com/topsites

[2] [2] B. Zhou, Jian Pei, Wo-Shun Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data," ACM SIGKDD Explorations Newsletter, Vol. 10, pp. 12-22, 2008.

[3] [3] D. Rosenblum, "What anyone can know: The privacy risks of social networking sites," In: IEEE Security & Privacy, Vol. 5, Issue 3, pp 40-49, 2007.

[4] [4] A. Narayanan, V. Shmatikov, "De-anonymizing social networks", In Proc of 30th IEEE Symposium on Security and Privacy, Berkeley, CA, pp 173-187, 2009.

[5] [5] J. M. Kleinberg, "Challenges in mining social network data: processes, privacy, and paradoxes," In Proc. of 13th ACM SIGKDD International conference on Knowledge discovery and data mining, ACM New York, NY, USA, pp 4-5, 2007.

[6] [6] L. Backstrom, Cynthia Dwork, Jon Kleinberg, "Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography," In Proc. of 16th International conference on World Wide Web, ACM, New York, NY, USA , pp 181-190, 2007.

[7] [7] Jaideep Srivastava, Muhammad A. Ahmad, Nishith Pathak, David Kuo-Wei Hsu, "Data mining based social network analysis from online behavior," SIAM conference on Data Mining, 2008.

[8] [8] A. F. Westin, Privacy and freedom vol. 97: London, 1967.

[9] [9] B. Carminati, E. Ferrari, and A. Perego. Security and Privacy in Social Networks, volume VII of Encyclopedia of Information Science and Technology, pages 3369–3376. IGI Publishing, 2 edition, 2008.

[10] [10] Sebastian Ryszard Kruk, Slawomir Grzonkowski, Adam Gzella, Tomasz Woroniecki, and Hee-Chul Choi. D-FOAF: Distributed identity management with access rights delegation. In Riichiro Mizoguchi, Zhongzhi Shi, and Fausto Giunchiglia, editors, ASWC, volume 4185 of Lecture Notes in Computer Science, pages 140–154. Springer, 2006.

[11] [11] Bader Ali, Wilfred Villegas, and Muthucumaru Maheswaran. A trust based approach for protecting user data in social networks. In 2007 Conference of the Center for Advanced Studies on Collaborative research (CASCON'07), pages 288–293, 2007.

[12] [12] Barbara Carminati, Elena Ferrari, and Andrea Perego. Enforcing access control in web-based social networks. ACM Transactions on Information Systems Security, 13(1), 2009.

[13] [13] Dan Brickley and Libby Miller. FOAF vocabulary specification 0.91. RDF Vocabulary Specification, November 2007. Available at http://xmlns.com/foaf/0.1.

[14] [14] Hee-Chul Choi, Sebastian Ryszard Kruk, Sławomir Grzonkowski, Katarzyna Stankiewicz, Brian Davids, and John G. Breslin. Trust models for community-aware identity management. In Identity, Reference, and the Web Workshop (IRW 2006), May 2006. Available at: http://www.ibiblio.org/hhalpin/irw2006/skruk.pdf.

[15] [15] Gianluca Tonti, Jeffrey M. Bradshaw, Renia Jeffers, Rebecca Montanari, Niranjan Suri, and Andrzej Uszok. Semantic web languages for policy representation and reasoning: A comparison of KAoS, rei, and ponder. In Dieter Fensel, Katia P. Sycara, and John Mylopoulos, editors, International Semantic Web Conference, volume 2870 of Lecture Notes in Computer Science, pages 419–437. Springer, 2003.

[16] [16] Timothy W. Finin, Anupam Joshi, Lalana Kagal, Jianwei Niu, Ravi S. Sandhu, William H.Winsborough, and Bhavani M. Thuraisingham. R OWL BAC: representing role based access control in OWL. In Indrakshi Ray and Ninghui Li, editors, SACMAT, pages 73–82. ACM, 2008.

[17] [17] Yague, Gallardo, and Mana. Semantic access control model: A formal specification. In ESORICS: European Symposium on Research in Computer Security. LNCS, Springer-Verlag, 2005.

[18] [18] N. Elahi, M.M.R. Chowdhury, and J. Noll. Semantic access control in web based communities.In Computing in the Global Information Technology, 2008. ICCGI'08. The Third International Multi-Conference on, pages 131–136, 2008.

[19] [19] 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.

[20] [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] [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] [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] [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] [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] [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] [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] [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] [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] [29] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, "l-diversity: Privacy beyond k-anonymity," In: ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 1, 2007.

[30] [30] 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.

[31] [31] Na Li, Nan Zhang, Sajal K. Das, “Relationship Privacy Preservation in Publishing Online Social Networks”, In Proc. of IEEE International Conference on Privacy, Security, Risk, and Trust, Boston, MA, pp 443-450, 2011.

[32] [32] Sanaz Kavianpour, Zuraini Ismail, and Amirhossein Mohtaseb, “Preserving Identity Of Users In Social Network Sites By Integrating Anonymization And Diversification Algorithms”, In: International Journal of Digital Information and Wireless Communications (IJDIWC), Hongkong, Vol. 1, Issue 1, pp 32-40, 2011.

[33] [33] B. K. Tripathy, Anirban Mitra, “An Algorithm to achieve k-anonymity and l-diversity anonymisation in Social Networks”, In Proc. of Fourth International Conference on Computational Aspects of Social Networks (CASoN), Sao Carlos, pp 126 – 131, 2012.

[34] [34] Mingxuan Yuan, Lei Chen, Philip S. Yu, Ting Yu, “Protecting Sensitive Labels in Social Network Data Anonymization”, In: IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 3, pp 633-647, 2013.