International Journal of Latest Trends in Engineering and Technology
Vol.(10)Issue(1), pp.375-378
DOI: http://dx.doi.org/10.21172/1.101.68
e-ISSN:2278-621X

ANALYSIS OF EXISTING TECHNIQUES TO COMBAT PRIVACY & SECURITY ISSUES IN SOCIAL NETWORKS

Sanjeev Dhawan[1], Kulvinder Singh[2], Kartik Sharma[3]

Abstract- Social networking sites have become a vital part of our day to day lives and are treated as the accustomed means for interpersonal communication and the spread of information. Consequently, social networks have turned into repositories of more than half a billion registered users' possibly sensitive and private data. The data comes as a result of the information posted by users themselves and their interactions with each other. Leakage of such personal information may result in malicious attacks from the actual world and cyberspace e.g. phishing, spamming, stalking, character assassination etc. Due to the aforementioned reason and the sensitive nature of the data stored in social networks, privacy and security become the core concern. The issue has gotten ample recognition in the last decade, which has resulted in numerous proposed techniques intended to enhance the privacy and security of social networks. This paper is focused on examining and scrutinizing such techniques proposed till date to combat the security and privacy issues of the social networks. It contains three partitions, the first one being the introduction about the issue at hand, the second one being the existing techniques to contest the issues specified in the first part and the third one being the overall conclusion of the analysis. The second portion is informally partitioned into two types of techniques, anonymization techniques and Sybil defense techniques.
Keywords – OSN; Social Networks; Security; Privacy-preserving techniques; anonymization techniques; Sybil attacks;

1. INTRODUCTION

The concept of social networking has been around for ages. With the advent of the Internet, it has transformed more or less into an instrument for connecting people and letting them communicate with the ease and extravagance which was considered impossible a while ago. Privacy and Security have been the most neglected aspects of social networks for quite a long time. Even though the risks are well known, some social networks still intentionally make/keep the privacy and access control mechanisms weak in order to make the process of joining and sharing relatively easy as well as to cut costs. Achieving anonymity and recognition/denial of access to Sybil nodes have been in the spotlight for quite some time. Anonymity refers to the imperative facet in which person-centric data once released cannot be used to pinpoint the accurate identities of the people who are the subjects of that released data without losing the worth of that data. There have been various successful and unsuccessful attempts at algorithms for achieving anonymity in social networks.

Sybil nodes are the ones which create several bogus identities for themselves in order to gain access to the confidential data or to hinder the performance of the network. Social Networks are decentralized distributed systems and hence are more susceptible to Sybil attacks. In this paper, our focus has been on the analysis of techniques which intend to enhance the privacy and security of social networks with an emphasis on anonymization techniques and Sybil attack prevention techniques.

2. PROPOSED TECHNIQUES TO ENHANCE THE PRIVACY & SECURITY ASPECT OF ONLINE SOCIAL NETWORKS

2.1 k-Anonymity

Latanya Sweeney's k-Anonymity[1, 2], being a formal protection model, is one of the first substantial privacy-preserving techniques. Her research focused on sharing a version of data without compromising the identities of the subject individuals and the integrity/usefulness of that data. k-Anonymity was aimed as a technique to halt the re-identification of the subjects with scientific guarantees and hence thwart identity disclosure.The k-Anonymity model made sure that eachindividual's informationincluded in the data release cannot be used to distinguish that individual from at least other k-1 individuals in the data. She recognized that multiple queries could become a hazard to individual's privacy and could leave room for inference depending on the size of the database.

[1] Faculty of Computer Science & Engineering, University Institute of Engineering & Technology, Kurukshetra University, Kurukshetra, India
[2] Faculty of Computer Science & Engineering, University Institute of Engineering & Technology, Kurukshetra University, Kurukshetra, India
[3] Research Scholar of Computer Engineering, University Institute of Engineering & Technology, Kurukshetra University, Kurukshetra, India

Null-map, k-map, and wrong-map were her previous attempts at privacy preservation which used the concept of information mapping to no, 'k' and incorrect entities. This model uses the concepts of generalization and suppression in order to achieve anonymity.The k-Anonymity model is still vulnerable to various kinds of attacks including unsorted matching attack, complimentary release attack, temporal attack, homogeneity attack and background knowledge attack. It also fails at handling attribute disclosure.

2.2 l-Diversity

l-Diversity is a refined extension of k-Anonymity which aims at removing the exposures of the k-anonymity model. The basic concept of l-diversity is to ensure more diversity in the data to thwart homogeneity attacks, background knowledge attacks, and sensitive attribute disclosure. Daniel Kiferet al.[3] based their model on the foundation of Bayes-Optimal Privacy, which is considered as the epitome of privacy notions, and privacy principles namely, positive disclosure and negative disclosure. Bayes-Optimal Privacy's practical setbacks were studied as well and taken into consideration. In their model, lack of diversity could be represented by the following equation in case of sensitive attributes:

$\forall s' \neq s, \quad n_{(q^*,s')} \ll n_{(q^*,s)}$

In the above equation, let's consider a sensitive attribute denoted by A which has the same value of 's' in case of all the tuples of A. $n_{(q^*,s)}$being the quantity of tuples t* in the anonymized table T* and $n_{(q^*,s')}$being the quantity of tuples t* in the anonymized table T* where t*[S] = s'.

l-Diversity requires the data to be diverse which can be satisfied using the above equation if the q*block has at least $l \geq 2$

diverse sensitive values. However, the adversaries may still attack successfully using the background knowledge if the below-mentioned equation is true.

$$\exists s', \quad \frac{f(s'|q)}{f(s'|q^*)} \approx 0$$

In the above equation, $f(s'|q)$ is the sensitive attribute's conditional probability inured on the notion that non-sensitive attributes of some q' may be generalized to q and $f(s'|q^*)$ is the sensitive attribute's conditional probability inured on the notion that non-sensitive attributes of some q' may be generalized to q*.

l-Diversity prevents the homogeneity attack, background knowledge attack as well as sensitive attribute disclosure but like any other model, it has vulnerabilities as well. Skewness attack and similarity attack are the two major weaknesses of l-Diversity. It is also considered unnecessary and difficult to achieve in some cases. Probabilistic l-Diversity, entropy l-Diversity, and recursive l-Diversity are the modified versions of l-Diversity.

2.2 t-Closeness

t-Closeness was another refinement to both the k-Anonymity and l-Diversity. Ninghui Li et al.[4] analyzed both the abovementioned infamous techniques and studied their drawbacks. t-Closeness uses 't' as a threshold for calculating and assessing the degree of closeness by considering the semantic distance between sensitive attributes with respect to the threshold 't'.t-Closeness covers for most of the limitations of k-Anonymity,andl-Diversity but even a combination of all three can leave room for exposure.

2.3 Zhou and Pei's approach against neighborhood attacks

Zhou and Pei's[5] research focused on averting neighborhood attacks by providing a certain level of anonymity to be achieved by preserving the privacy of individuals denoted as vertices in any social network. They introduced a practical approach to satisfy the k-Anonymity requirement of the social network in an effort to achieve network anonymity. Their approach comprised of two steps with the first one being the extraction of neighborhoods of all vertices in the network and the second one being the greedy organization of vertices in groups and anonymization of the neighborhoods in the same group. This method used the depth-first-search algorithm for anonymizing the network.

This approach successfully handled 1-neighborhood attacks but was not tested on neighborhoods beyond degree 1. The solution offered by their approach was a partial one since their algorithm could not deal with situations where the adversaries have background knowledge of the vertices.

2.4 Edge Anonymity

Lijie Zhang and Weining Zhang[6] proposed a probabilistic concept of edge anonymity named graph confidence. Their concept focused on recognizing privacy breaches occurred while attackers identify/isolate target individuals in a graph partition. Their concept also considered the special anonymity case where vertex degree is used to partition a graph. They presented three experiential algorithms to obtain τ-confident anonymized graphs. These algorithms were applied on three real-world social networks by them with positive results by yielding edge-anonymous graphs without losing the usefulness of the data.

2.5 Fresh Approach against neighborhood attacks

B.K. Tripathy and G.K. Panda[7] proposed a fresh approach to avoid and deal with the neighborhood attacks in social networks. They explored the key difficulties in anonymizing social networks and the possibilities and working of neighborhood attacks as well. They proposed a modification to a pre-existing algorithm[5]which relied on k-Anonymity[1]. The original algorithm used DFS technique for graph isomorphism. The modification was to replace that technique with the adjacency matrix technique in order to increase its efficiency in terms of security and time complexity.

Their proposed algorithm protects the users and anonymizes them not only when the adversaries have information about the user's immediate neighbors but also when they have information about neighbors within a finite number of hops as well.

2.6 KNAP - Personalized Anonymity

Lihui Lan et al.[8] took a different approach and considered the concept of personalizing anonymity for different types of entities. Existing technologies have their focus on common approaches for all types of entities but in the real world, different entities have different privacy requirement. Their approach classified entities into two groups namely sensitive and non-sensitive entities. They took this approach against 1-neighborhood attacks. Their proposed algorithm, 'KNAP – K Neighborhood Anonymous Publication', gave positive results along with maximum information retention and minimum cost.

2.7 SybilGuard

Social Networks are very vulnerable to Sybil attacks where attackers obtain multiple fake identities and misuse them to gain access or information which they are not entitled to. Haifeng Yu et al. proposed a decentralized protocol named 'SybilGuard' in order to curb the damaging influences of such Sybil attacks. Their concept is based on the notion that Sybil nodes may be able to get new identities, but they can be recognized andcounteracted by scrutinizing the trust relationships between nodes aka edges.[9]

SybilGuard works by creating a distinctive borderline between the honest nodes and the Sybil nodes. From a set of 'n' sized honest nodes, each node first obtains $\sqrt{n}$ independent samples. SybilGuard uses a sole instance of a very long arbitrary course and is bound by the threshold of

$$g = o\left(\frac{\sqrt{n}}{\log n}\right)$$

in order to offer constraints on Sybil identities. It can support up to $O(\sqrt{n}\log n)$ Sybil nodes per attack edge.

SybilGuard was subjected to mock experiments with a one million node sample size, resulting in positive results. However, the lack of real-life social network as a subject for the experiments and its dependency on number of attack edges leave room for uncertainty.

2.8 SybilLimit

SybilLimit operates on the same approach as SybilGuard but promises considerably enhanced and near-optimal results[10]. SybilLimit works by using multiple instances of short arbitrary walks since short arbitrary walks are expected to remain in the honest node regions. SybilLimit's improvement over SybilGuard may be credited to various modifications including a fresh benchmarking method for the valuation of τ, a fresh balance condition for dealing with the verifier's escaping tails, utilization of intersection of edges rather than the intersection of nodes etc.

However, SybilLimit depends on the correct estimation of the length of short arbitrary walks without any actual accurate way to quantify it.

2.9 SybilInfer

SybilInfer is another worth-mentioning attempt at detection of Sybil nodes in peer-to-peer social networks. It is a decentralized algorithm proposed with the objective of labeling and classifying nodes as genuine/honest nodes and Sybil nodes with the aid of machine learning techniques. It is based on a probabilistic model of genuine nodes and an inference engine which supplies us with the most probable home regions to the Sybil nodes[11]. SybilInfer does not make any assumptions with respect to the number of conniving entities.

SybilInfer proved itself to be a secure algorithm with minimum leeway for manipulation by the malicious/Sybil nodes. However, it requires the complete knowledge of the network topology.

3. CONCLUSION

Social Networks have become a vital part of our technological environment as well as our day-to-day life. Social Networks also expose us to numerousthreats which often solicit attacks from the cyberspace and the real world. In this paper, we've discussed and analyzed a variety of proposed privacy-preserving techniques and security enhancement techniques. According to ourassessment, a network would feature the best anonymity when the data is k-Anonymous, l-Diverse and t-Close as well. The coupling of these three is the finest solution for preserving privacy but that still isn't adequate and leaves a few loopholes depending on the type of data in question. Out of all the above approaches for handling Sybil attacks, each one brings an advancement to the table with time and has their own pros and cons. None of the above-mentioned approaches are entirely foolproof and hence more research focus is crucial in this area of research.

4. REFERENCES

[1] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," International Journal in Uncertainty, Fuzziness and Knowledge-based Systems, vol. 10, no. 5, pp. 557-570, 2002.

[2] P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," Technical report, SRI International, 1998.

[3] A. Machanavajjhala, J. Gehrke, D. Kifer and M. Venkitasubramaniam, "l-Diversity: Privacy beyond k-Anonymity," in Proceedings of the 22nd International Conference on Data Engineering, Atlanta, USA, 2006.

[4] N. Li, T. Li and S. Venkatasubramanian, "t-Closeness: Privacy beyond k-Anonymity and l-Diversity," in 23rd International Conference on Data Engineering, Istanbul, Turkey, 2007.

[5] B. Zhou and J. Pei, "Preserving Privacy in Social Networks Against Neighborhood Attacks," in 24th International Conference on Data Engineering, Cancun, Mexico, 2008.

[6] L. Zhang and W. Zhang, "Edge Anonymity in Social Network Graphs," in International Conference on Computational Science and Engineering, Vancouver, Canada, 2009.

[7] B. Tripathy and G. Panda, "A New Approach to Manage Security Against Neigborhood Attacks in Social Networks," in International Conference on Advances in Social Networks Analysis and Mining, Denmark, 2010.

[8] L. Lan, H. Jin and Y. Lu, "Personalized Anonymity in Social Networks Data Publication," in IEEE International Conference on Computer Science and Automation Engineering (CSAE), Shanghai, China, 2011.

[9] H. Yu, M. Kaminsky, P. B. Gibbons and A. Flaxman, "SybilGuard: Defending against Sybil Attacks via Social Networks," ACM SIGCOMM Computer Communication Review, vol. 36, no. 4, pp. 267-278, 2008.

[10] H. Yu, P. B. Gibbons, M. Kaminsky and F. Xiao, "Sybillimit: A near-optimal social network defense against sybil attacks," in IEEE Symposium on Security and Privacy, Oakland, California, 2008.

[11] P. Mittal and G. Danezis, "SybilInfer: Detecting Sybil Nodes using Social Networks," in 16th Annual Network & Distributed System Security Symposium, San Diego CA, 2009.

[12] G. Wondracek, T. Holz, E. Kirda and C. Kruegel, "A Practical Attack to De-anonymize Social Network Users," in 2010 IEEE Symposium on Security and Privacy, Berkeley/Oakland, CA, USA, 2010.

[13] N. Anjaiah and C. Ravi, "Protecting the Sensitive Information on Online Social Networks," International Journal of Research(IJR), vol. 1, no. 9, pp. 1163-1168, October 2014.