Privacy Preservation: An In-sight Approach of Online Social Networks

Research Scholar Inderjit Kaur[1], Dr.Rajinder Singh[2]
[1]GuruKashi University Talwandi Sabo,Punjab,India, [2]Guru Kashi University Talwandi Sabo,Punjab,India

***Abstract*: In the growing era of technology, social networking has evolved to create world as a global village. The main focus has shifted from accessing huge amount of data to enhancing privacy of user's personal details. In social networks, privacy constraint is often mislead as security while privacy parameter includes privacy of users, web links and their attributes. There are multiple components of privacy parameter in a network that includes multiple further problems. In a social network, user privacy includes multiple sub-problems like user location privacy, user interest and user personal information privacy. In this paper, privacy breaches are discussed that leads to proposal of privacy preservation approaches by numerous researchers. This paper laid stress on privacy parameter in social networks which provides a basic introduction and commencing action in the social network's effective deployment. Method of anonymization is also discussed along with various privacy preserving models.**

***Keywords:* Social Networks, Privacy Preservation, Anonymization techniques, Security.**

I.INTRODUCTION

Social Networks are gaining popularity in recent times with the increase of use of mobile devices that are capable of communicating through internet access. A Social Network can be defined as a group of interconnected devices that are connected through relations that compose a network which allows interaction among the devices along with communicating relevant information.

Online social media applications allow one person to interact with other people on the internet. Generally, people A particular user can interact with others by (1) bookmarking web sites and browsing through marked sites of other users, (2) voting for articles and making comments about them, (3) adding friends, (4) making comments about the content of their profiles, (5) joining groups and discussions, (6) exchanging photos and video content as well as by (7) adding articles, editing and modifying the existing ones [1]. It can be inferred as a site where users interact with each other for entertainment purposes as well as knowledge sharing.

tend to surf social networking sites for searching old friends, exploring new friends, or discovering people with the same interests, hobbies or problems across multiple domains such as political, economic as well as geographic borders. Online social media applications are provided in various forms.

Generally, there are six different forms of social media [2]: collaborative projects, blogs, content communities, social networking sites, virtual game worlds and virtual communities. Figure 1 shows the taxonomy of social media services.

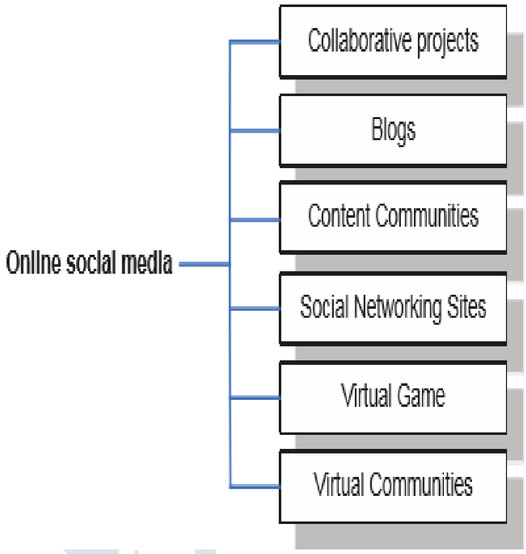

Figure 1: Taxonomy of Social Media

a. Collaborative projects are the ones that permit multiple users to collaborate for creating relevant contents. The prime motto behind such projects is the collaborative effort of the users that leads to increase the potential to a better outcome as compared with individual achieved result. Wikipedia is a well known project which can be edited by anyone with access to the site.

b. Blogs is a web based application which displays the updated content on the content entry date in chronological based order. These blogs may appear in variations, ranging from personal diaries of the author's life to innovation expression based on various issues to reviews or summaries of all relevant information in specific product or field. Each user has a privilege of creating a profile page containing user contact information, a brief biography, images and so on which is helpful in locating friends. Users can also hold a friend list which allows them to alert the most recent journal entries of the people on his/her friends list. Social interaction is performed by leaving comment to other users' entries.

c. Content communities enable the sharing of media contents between participants. Content communities comes in a multitude of different media types, including videos, photos, text, and PowerPoint presentations. In this form, users also have a "User Info" page, which contains a variety of data. Apart from sharing content, users perform social interaction by leaving comments and give rating to other users' contents. Each user also has a friend list to collect the most recent media entries from their friends.

d. The Social networking sites, the most popular form of social media are applications that enable participants to connect by creating personal information profiles, inviting friends and colleagues to have access to those profiles, and sending e-mails and instant messages between each other. These personal profiles can include any type of information, including photos, video, audio files, and blogs. Indeed, this form mixes several social media types into one package.

e. Virtual worlds are platforms that replicate a three-dimensional environment in which users can appear in the form of personalized avatars and interact with each other as they would in real life. Virtual world can have two forms;

i. Virtual game and

ii. Virtual social world.

In Virtual game, users are required to behave according to strict rules in the context of a massively multiplayer online role-playing game. The second group of virtual worlds, often referred to as Virtual Social Worlds, allows inhabitants to choose their behavior more freely and essentially live a virtual life similar to their real life. There are no rules restricting the range of possible interactions.

Social networks sites contain (collects) a tremendous amount of text, image, audio or video content which can be leveraged for a wide range of business purposes. Such content richness led to two basic data types that are analyzed in the context of social networks [10]:

- Links – Linkage-based and structural analysis is dealing with linkage behavior of a network to identify important nodes, communities, links, and so on.
- Content – Content-based analysis is focused on analysis of text, images, tags, and any additional added contents.

The flow of this research paper is as follows:

1. Section II provides the fundamentals of need for privacy.
2. Section III provides the difference between privacy and security.
3. Section IV explores types of privacy breaches.
4. Section V discusses overview of Anonymization techniques.
5. Section VI summarises the paper in concülsion.

II. NEED FOR PRIVACY

With the advent of new era of digitalization, the use of electronic gadgets and data has increased. It gives rise to analyse the trends of the users in the world. Parameters of social and economic are prime focus which leads to consider privacy as a major concern. Any personal details entered by the user is not to be disclosed because leaking of personal data is considered unethical and is against laws. This prevention of personal data from being unethically disclosed is termed as Privacy. In Social Networks, any detail entered by the user has to be shared with the permission of the user and right to reveal or hide certain details is the choice of the user.

It is often noted that privacy and security are considered as one but they are two sides of a coin. Both privacy and security are terms that are utilized interchangeably as per the context of usage. In section III, the difference between privacy and security is analyzed.

III. . HOW ARE PRIVACY AND SECURITY DIFFERENT

Privacy and security are two terms used interchangeably under different contexts. But both are related to each other and at the same time entirely separate issues.[3]

- The three fundamentals of security are Confidentiality, Integrity and Availability. In context of Census data, security can be termed as the facility for controlling person-specific access information, protect it from unauthorized disclosure, modification, loss or destruction of his information. Security can be accomplished through controls based on operational and technical knowhow.
- In contrast privacy is very specific. It can be termed as a right of an individual to keep his/her personal information from being disclosed. Privacy can be accomplished through policies and procedures. Person's personal information which may lead to his identification may not be disclosed under ethical grounds.

Privacy is considered as an important aspect of preserving information without information loss. The perspective of privacy differs based on the data in use and the way in which it is used. Many methods like attribute removal, anonymization, randomization, aggregation on numeric values are applied on data sets to provide privacy. These methods incur information loss in some situations too. Cryptographic techniques involve additional computational overhead. Secure sum computations require the feasibility of basic combinatorial circuit which computes the functions on data [3].

IV. TYPES OF PRIVACY BREACHES

There are mainly four types of privacy breaches in online social network data [4].

i. Vertex Re-Identification: The privacy breach where identity of an individual is revealed is known as vertex reidentification. This re-identification also reveals sensitive labels and all the relationships of the individual under attack with other individuals in the network.

ii. Edge or Relationship Re-Identification: Edge re-identification occurs when relationship between two individuals is revealed. Utilizing social network services (like sending an email or message) generates this kind of information.

iii. Sensitive Label or Attribute Re-Identification: This kind of re-identification leads to revelation of sensitive and confidential attributes, like Disease, Salary etc., of an individual.

iv. Content Discloser: This kind of breaches disclose the data associated with each vertex, e.g., emails sent and/or received by the individuals in a email network.

V.RESULTS

One of the simplest definitions of anonymity is the following. *Anonymity is the state of being not identifiable within a set of objects, the anonymity set[11].*

Anonymization technique plays the major role in the protection of sensitive information of an individual. This technique is used for detecting adversaries and provides privacy for sensitive information. Privacy can be provided by encrypting or removing personally identifiable information from the dataset. Anonymization can be applied in the fields of online shopping, online communication, telecommunication and social network carts [12].

A. K-Anonymity: It is a property which is controlled by certain anonymized data. Given individual particular field structured information; create an arrival of the information with logical ensures that the people who are the subjects of the information can't be re-recognized while the information remain basically valuable. An arrival of information is said to have the k-anonymity property if the data for every individual contained in the discharge can't be recognized from at any rate k-1 people.[5]

B. K-Anonymization Methods Suppression: In this technique, certain estimations of the attributes are supplanted by a mark '*'. All or a few estimations of a segment might be supplanted by '*'. Suppression comprises in averting delicate data by evaluating it. Suppression can be connected at the level of single cell, whole tuple, or whole segment, permits diminishing the measure of speculation to be forced to accomplish k-anonymity. [5]

1) Tuple (TS): Suppression is performed at column level; suppression operation evacuates entire tuple

2) Attribute (AS): Suppression is performed at segment level; suppression operation shrouds every one of the estimations of a segment.

3) Cell (CS): Suppression is performed at single cell level; at long last k-anonymized table may wipe out just certain cells of a given tuple/quality.

Privacy mechanisms for social networks mostly consider anonymization techniques for anonymizing network structure and user attributes. The anonymization techniques for network structure fall in four main categories:

(1) edge modification,

(2) randomization,

(3) network generalization, and

(4) differentially private mechanisms.

However, providing the anonymizited structure of social networks is often not sufficient for the purposes of the researchers. On the other hand, the assumption is that anonymized data will have utility only if it contains both structural properties and node attributes [6]

C. L-Diversity: In this type of gathering-based anonymization technique which is deployed for purpose of saving .This diversity is an extension of the base k-anonymity approach which decreases the granularity of the relevant information representation utilizing methods. For l-diversity, the anonymization conditions are satisfied if, for each group of records sharing a combination of key attributes, there are at least l -“well-represented” values for each confidential attribute [8]

D. T-Closeness: This approach is an advancement to the l-diversity approach where delicate fields' dispersion is taken into consideration along with the presence of assaults where touchy attributes might be construed based upon the circulation of qualities for l-diversity information. A new method named t- closeness was proposed in [9] to address these problems. This method requires the distribution of the sensitive attributes in an equivalent class to be close to the distribution of the attribute in the overall table, which in turn means that the distance between the two distributions should be no more than a specified threshold t.

E. Bottom-Up Generalization: On applying Map Reduce method on cloud to Bottom Up Generalization (BUG) for the purpose of information anonymization as well as gathering of inventive Map Reduce method employments are used to achieve the generalizations in acceptable trend.

F. Top-Down Specialization: By and large, TDS is an iterative procedure beginning from the topmost area esteems in the scientific classification trees of attributes. Each round of emphasis comprises of three stages [7]:

- Searching the best specialization
- Performing specialization
- Refreshing values of the search metric for the next round

Such a process is repeated until the point when k-anonymity is violated, to uncover the most extreme information utility. The decency of a specialization is estimated by an inquiry metric

VI. CONCLUSION

In this paper, a brief overview of Social Networks is discussed and the Privacy constraints in different social networks is analysed. With the growth of internet based devices usage, privacy of contact details and its interrelation with security measures is discussed. A spec of details regarding privacy breaches was examined and studied. In order to preserve privacy, various anonymization method were discussed as well as the main limelight was given on k-anonymity approach which included both generalization and suppression. The last part of the paper deals with approaches that are necessary in regulating disclosure content for securing the protection measures of information which is utilized in various modules of Social networking.

VII.REFERENCES

[1]. Grljević, O., Bošnjak, Z., Značaj, "analize sadržaja socijalnih medija", YU-INFO 2012, Kopaonik, Serbia, 2012.

[2]. Ninggal, M. I. H.," Privacy-Preserving Mechanism for Social Network Data Publishing", Deakin University, June, 2015

[3]. Shah, A., Gulati, R.,"Privacy Preserving Data Mining: Techniques, Classification and Implications - A Survey", International Journal of Computer Applications (0975 – 8887), Pp. 40-46, Vol. 137, No.12, March 2016.

[4]. Papri, M.,"A Brief Survey on Vertex and Label Anonymization Techniques of Online Social Network Data", Int. Journal of Engineering Research and Applications, ISSN : 2248-9622, pp.38-42, Vol. 5, Issue 6, (Part -4) June 2015.

[5]. Parkar, T.K., Chouhan, Y., Gajbhiye, S.,"A Survey on Anonymization Based Privacy Preserving Models", Journal of Network Communications and Emerging Technologies (JNCET), Vol. 8, Issue 3, March 2018.

[6]. Zheleva, E.,"Prediction, Evolution and Privacy in Social and Affiliation Networks", PhD dissertation, July, 2011.

[7]. Zhang, X., Liu, C., Nepal, S., Yang, C., Dou, W., Chen, J.,"Combining Top-Down and Bottom-Up: Scalable Sub-Tree anonymization over Big data using MapReduce on Cloud”.

[8]. Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M.,"L-diversity: Privacy beyond k-anonymity", ACM Transactions on Knowledge Discovery from Data (TKDD), Vol. 1, Issue 1, No. 3, 2007.

[9]. Li, N., Li, T., Venkatasubramanian, S.,"T-closeness: Privacy beyond K-anonymity and L-diversity", IEEE 23rd International Conference in Data Engineering, pp. 106–115, IEEE, 2007.

[10]. Aggarwal, C. C. (Ed.),"Social network data analytics", Springer, 2011.

[11]. Pfitzmann A., K¨ohntopp, M.,"Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology", In Proceedings of the Workshop on Design Issues in Anonymity and Unobservability, pages 1–9, 2000.

[12]. Gowthamy, R., Uma P.,"Identity Disclosure Protection In Dynamic Networks Using K^w – Structural Diversity Anonymity", International Journal on Integrating Technology in Education (IJITE), Vol. 5, No. 1, Pp. 27-35, March 2016.