A Review of Challenges and Methods on Privacy Preservation of Social Networking Data

Himani Saini[1](Research Scholar), Gopal Singh[2](Asst. Professor)

[1,2] Department of Computer Science and Application, Maharshi

Dayanand University, Rohtak(Haryana), India

[1]himanisaini836@gmail.com, [2]gsbhoria@gmail.com

Abstract **Enormous amount of data is being generated by various ecommerce or social Networking sites. To provide suggestions and recommendation to the customers or users, there is a need to processed and analyzed this voluminous data. Sharing of these data is helpful for application users. While sharing the private data, privacy preservation become one of the major concern for everyone. Due to the easy availability of numerous digital technologies, tons of data is generated every minute and preserving or securing such huge amount of data is not an easy task. this paper examines various privacy threats and privacy preservation techniques such as anonymization, randomization and cryptography.**

Keywordsprivacy preservation, privacy threats, data anonymization, Randomization, cryptography.

I. Introduction

Nowadays, the Use of various social networking sites become a trend. the term **social networking sites** means **the online platforms on which people can create their public profile and interconnect with the other people(family, friends and relatives etc) on the website**[1]**.** The most common or trending social networking sites are Facebook, Whatsapp, Instagram, Twitter, Snapchat, LinkedIn etc. Due to the exquisite use of these sites, the social media data generated by the users is exploding [2].These social sites have gained enormous popularity in a very short period of time, so people tend to use diverse sites for various purposes like -To communicate with their family and friends, to share their daily experiences with each other, express their views on some commercial products or social events etc.

In Today's digital era, where there is plenteous availability of digital tools on internet , people are not much interested to use offline sources to store all their private information. People are onlinestoring their personal content such as date of births, contacts, pictures, files, zip code, bank related information, some official documents and bookmarks[3]. Even people can easily interact with these social networking sites through tweets, posts and tags.

Fig.1 Various Social Networking Icons

The volume and variety of these user-generated social media data is procreating exponentially. There can be many reasons behind this rapid growth but the main cause can be the affordable availability of internet connectivity, storage and various computer technologies. Among all these social sites ,facebook is one of the most widely used social networking site and at the top with an active users of 2.13 billion worldwide. whatsappan another messaging app with an additional calling feature is at the second level with an active users of 1.5 billion. Facebook messenger is just behind it with 1.3 billion users and last but not the least Instagram has its 800 million active users all over the world[4].

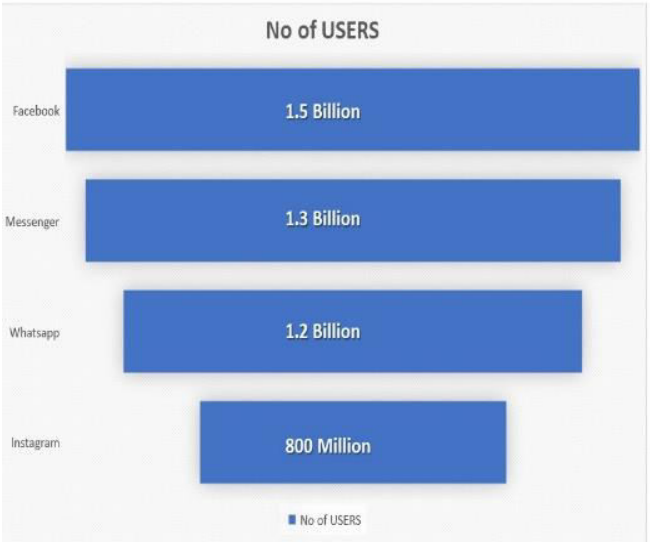

Fig.2 Numbers of the Active Users in Some Trending Social Networking Sites

II. Privacy Preservation

As we all know that the social media data is increasing sharply day by day. The privacy of this data is one of the major growing concern in recent time. Before going in detail, let's make you clear about the term "privacy preservation". The term privacy preservation mainly **refers to the safekeeping of personal data or information specifically from being leaked, damage , loss or destroyed.** the personal information not only means an important document. It can be anything like images, audio or video recordings, applications, databases or can be a combination of them.

Sometimes the user himself contributes to a data leak either intentionally or unintentionally. This is the best example to understand when we allow all mobile applications to seek access to our personal data like images, contacts, documents etc. without reading the rules and privacy status.

III. Challenges in Preserving Privacy: Privacy threats

To avoid such circumstances which is described above where user itself contributing to the data leakage, there is need to educate the users about such privacy threats. There are several privacy threats which can be riskier but we will discuss about some of the major privacy threats which are [5] :-

- vigilance
- Data Revelation
- favouritism or biasness

A. *Vigilance*

It mainly refers to the action of observing something carefully. Another term used for this is surveillance. All these social networking sites does provide suggestions like choose friends, people you may know and much more, just by looking on the actions they performed and the type of data they share.

B. *DataRevelation*

Data revelation can be define as the process of disclosingof all the sensitive information of the customer to the third party. Example disclosing of the patient personal data(eg-age, name, dob, gender, disease etc.) from the hospital is a serious kind of privacy threat. It can be used to violate customers privacy.

C. Favouritism

The act of inequality where a person or group is treated unfairly. The related term to this are discrimination, bias, injustice etc.To combat with such situations there is a strong need to spread awareness among people because one of the main reason of this is lack of awareness among people.

IV. Privacy preservation methods

Now when people are using the social networking sites continuously, they always have a fear that their personal information will fall into wrong hands or can be misused by anyone. To protect the loss or damage of the personal data we need to preserve it by various means.

Fig.3 Privacy Preservation

There are various privacy preserving techniques for securing the data and few of them are listed below:

- Data Anonymization
- Randomization

- Cryptography

A. DataAnonymization

It is defined as the process by which personal data is modified in such a way that the original data cannot beidentified either directly or indirectly by anyone. It is also called asde-identification [6]. Data anonymization helps in transferring the data with in a closed boundary between two or more parties. It reduces the risk of unwanted disclosure of information which is one among the several privacy threats.Anonymization can be obtain by several methods like Generalisation, swapping, data removal, suppression etc. Two of the main privacy preserving approaches are k anonymity or l diversity.

1) *K-Anonymization:*Some of the Anonymized data possessed this property. The concept was first introduced by Latanya Sweeney and PierangelaSamarati in 1998.A release of data is said to have the k anonymity property if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appear in the release [7]. Example: While trying to identify a person from a table the only input information given is person's birth date and gender, then there will be k people meeting the requirement [8].

2) *L-Diversity:*l-diversity is another typeof group based anonymization which basically aim is to preserve privacy in datasets. "An equivalence class is said to have l-diversity if there are at least l well represented values for sensitive attributes. A table is said to have l-diversity if every equivalence class of the table has l-diversity [9]. l-diversity model is just a modification of the k anonymity model which handle some of the gaps that is present in k anonymity model.

B. Randomization

Randomization is one of the easy, efficient and less expensive technique for privacy preservation [10]. Randomization is the process in which real data is altered by taking away some sensitive information and adding noise before sharing them. It is said to be an easy approach because it doesn't require knowledge of other records in the data. It can be applied at the time of collection of data or at pre-processing stage.

C. Cryptography

Cryptography has been defined as the technique of converting the plain text into cypher text. It is mainly used for storing or transferring the data in such a way so that only the receiver can read that useful information, no one else can misused the data. It's a complex method to use on large datasets, because encryption on large datasets is difficult to implement and it also reduces the utility of data. It is used to maintain the data confidentiality, integrity, authentication etc [11].

V. Research gaps and status

This section covers the gaps present in all these privacy preserving methods.

A. Limitations of k anonymity

The k-anonymity model does not protect against attacks based on background knowledge.it reveals individual sensitive information. The loss of data utility, when applied on high dimensional data [12].

B. Limitationsofl-diversity

L diversity method is more prone to the skewness attack and similarity attacks, because of the semantic relationship between the sensitive attributes it is inappropriate to neglect the attribute exposure [13].

C. Limitations of Randomization

It is not possible to apply randomization method on large datasetsbecause of time complexity and data utility. preserving privacy at the cost of data utility is not acceptable that's why it may not be an appropriate method of privacy preservation [14].

D. Limitations of cryptography method

Cryptography is very complex method to apply on large datasets and it is not suitable for unstructured data. It's not a good method when it comes to preservation of data attributes[14].

VI. Conclusion and Future trends

In this paper, we have discussed about various privacy preservation techniques. These methods are helpful in achieving privacy or securing the personal data upto some extent. All the above discussed methods have their own advantages and disadvantages that's why new method with higher approximation and better results is need to be

developed. Machine learning and various other soft computing techniques can be used to preserve privacy of user's data. There is still a lot of future work can be performed on this, which can provide more optimum results to the users.

REFERENCES

[1] "techopedia," [Online]. Available: https://www.techopedia.com/definition/4956/social-networking-site-sns. [Accessed 29 may 2020].

[2] J. Zhang, J. Sun, R. Zhang, Y. Zhang and X. Hu, "Privacy-Preserving Social Media Outsourcing," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, Honolulu,HI,USA, 2018.

[3] M. Siddula, L. LI and Y. LI, "An Empirical Study on the Privacy Preservation of Online Social Networks," *IEEE,* vol. 6, pp. 19912-19922, 2018.

[4] "Whatsapp Revenue And Usage Statistics 2020," 24 april 2018. [Online]. Available: https://www.businessofapps.com/data/whatsapp-statistics/. [Accessed 29 may 2020].

[5] P. M. Rao, S. Krishna and P. Kumar, "Privacy Preservation Techniques in big data analytics: a survey," *Springer open,* pp. 1-12, 2018.

[6] S. and K. , " A Research on Bigdata Privacy Preservation Methods," *IJRTE,* vol. 8, no. 1S4, pp. 175-177, 2019.

[7] w. contibutors, "k-anonymity," WIKIPEDIA,The Free Encyclopedia, 2014. [Online]. Available: https://en.wikipedia.org/w/index.php?title=K-anonymity&oldid=959297811. [Accessed 30 may 2020].

[8] S. Shelke and B. Bhagat, "Techniques for Privacy Preservation in Data Mining," *International Journal of Enginnering Research & Technology,* vol. 4, no. 10, pp. 490-493, 2015.

[9] WilliamKF, "L-diversity," WIKIPEDIA,The Free Encyclopedia, 2014. [Online]. Available: https://en.wikipedia.org/w/index.php?title=L-diversity&oldid=910039769. [Accessed 30 may 2020].

[10] P. Nivetha and s. Thamarai, "A Survey on Privacy Preserving Data Mining Techniques," *International Journal of Computer Science and MobileComputing,* vol. 2, no. 10, pp. 166-170, 2013.

[11] The Economic Times, "Definition of Cryptography," Economic times, 2020. [Online]. Available: https://economictimes.indiatimes.com/definition/cryptography. [Accessed 30 may 2020].

[12] N. Mogre, G. Agarwal and P. Patil, "A Review On Anonymization Technique For Data Publishing," *International Journal of Engineering Research & Technology,* vol. 1, no. 10, pp. 1-5, 2012.

[13] K. Rajendran, M. Jayabalan and M. E. Rana, "A Study on K-anonymity, l-diversity and t-closeness Techniques focusing Medical Data," *International Journal of Computer Science and Network Security,* vol. 17, no. 12, pp. 172-177, 2017.

[14] P. M. Rao, S. Krishna and A. S. Kumar, "Privacy Preservation techniques in big data analytics: a survey," *springer nature,* pp. 1-12, 2018.

[15] "Social Media logo Collection," Freepik, 2010. [Online]. Available: https://www.freepik.com/free-vector/social-media-logotypeset_3765835.htm#page=1&query=social%20media%20icon&position=43. [Accessed 2 Jun 2020].

[16] "9 Biggest threats to Privacy- is the Right to Privacy dead?," blokt, 2020. [Online]. Available: https://blokt.com/guides/biggest-privacy-threats. [Accessed 2 Jun 2020].