Privacy Preservation of Sensitive Information in Big Data

Manu Gupta[1] , Swati Bhardwaj[2], Girdhar Gopal[3]
[1]Assistant Professor, Department of Computer Science,
Sanatan Dharma College, Ambala Cantt
[2]Assistant Professor, Department of Computer Science,
Sanatan Dharma College, Ambala Cantt
[3]Assistant Professor, Department of Computer Science,
Sanatan Dharma College, Ambala Cantt

ABSTRACT

Data can be in different forms text, images, audio etc. This large amount of data which is indifferent forms called big data[2]. Big data can handle different type of enormous data and this large amount of data is difficult to handle by traditional methods in a limited amount of time. Big data uses different methods to process the large data and then gave results of analysis[1]. As Big Data is now being extensively used by all sectors and contains personal information of individuals. Privacy and security of personal information become a serious concern because of malicious use of information of users. . Privacy preserving techniques can be used to protect the sensitive information of individuals [12]. Data anonymization is the most beneficial technique when we require the privacy preservation of user information Several techniques are used for preserving privacy of sensitive information of users in big data such as k-anonymity, l-diversity, t-closeness and differential privacy. All above approaches for privacy preservation has some merits and demerits. Preserving privacy is essential because, beneficiary can use sensitive information of individual's and integrating it with public information to reveal a person specifically.

KEYWORDS:-Big Data, Anonymization

INTRODUCTION

Big data indicates data in enormous amount is difficult to handle by traditional systems. Big Data describes a huge volume of data - both structured and unstructured [1]. A huge amount of data is processed by all of us for example; via Social networking, transportation, Healthcare etc. Data is generated at a high speed and this data can be of various formats generated from different sources. It became very difficult to process and store huge data by traditional systems.

SOURCES OF BIG DATA

Big data can be of different size e.g. in peta bytes, exa bytes or in zettabytes also. Different Sources of bigdata are Weather forecasting, Business, Social-Media, Electronic Commerce, Weblogs, Medical Records etc.

Big Data can be of different types such as image, text, audio or video. Big Data can be of following category:

CHARACTERISTICS OF BIG DATA

Volume: System/users generating Terabytes, Petabytes, Zettabyte's and even Exabyte's of data and we need large storage for this huge data.

Velocity: Velocity plays a main role as compared to the others; there is no point in investing so much to end up waiting for the data. So, the main aspect of big data is to provide data on demand and at a faster pace.

Variety: The data comes from different sources that are structured data, semi- structured data or unstructured form of data in the form of audio, video and data- blogs etc.

LITERATURE REVIEW

W. Fang et al. [1], introduced different privacy preservation technique and how to preserve privacy by using these techniques. Major issue in big data is to preserve the privacy of user information. Some important aspects of privacy preserving, such as access control, encryption, anonymization, data auditing, and differential privacy, are discussed A. Khanan et al. [2], discussed the major security and privacy issues of big data. As big data is used both in private and public sector, managing data and securing infrastructure, visualizing the data all these are the issues that can be faced in big data system. M.A. Srinuvasu et al. [3], provided the overview of big data, its size, nature, 12Vs of big data and some technologies to handle it..S.Sathyamoorthy et al. [4], technologies are developing and people are using more technologies. With increase in technologies data is also increasing day by day and it is going to be difficult to handle the sensitive data..Jain et al. [5], discussed firstly the privacy specifications in big data. It also introduced some privacy preservation techniques that can be used to preserve privacy of personal information of users.. Goswami et al. [6], provided a study on big data, challenges, privacy and security and also distinguishes the privacy and security requirements in big data. The author explained different privacy methods of anonymization Ram Mohan Rao et al. [8], introduced different threats that can be harmful to user privacy. Some techniques for preserving privacy of user information are also described with their limitations. Charles et al. [9], in this paper a systematic study to security in the Big Data ecosystem is done and results are also discussed. A. Pawar et al. [18], data anonymizing and differential privacy approaches are discussed with all issues and countermeasures for described privacy preserving techniques. The author represents a comparative analysis of privacy preserving techniques. N. Maheshwarkar et al. [14], explained that big data have many techniques to solve privacy preserving issues. One of them is anonymization, used to preserve privacy of sensitive data. K-anonymity is the approach of anonymization that is used for preventing identity disclosure M S Simi et al. [17], proposed three best algorithms along with their adaptability and effectiveness. Recognizing the association between the values of k, degree of anonymization, selecting a quasi-identifier and emphasize on execution time is all worked in algorithms.

OBJECTIVES OF RESEARCH

- Study and Compare the various privacy preserving techniques for preserving privacy of sensitive information in big data.
- To empirically analyze the two efficient privacy preservation techniques using various evaluation criteria.

SECURITY AND PRIVACY

As Big Data is now being extensively used by all sectors and contains personal information of individuals. Privacy and security of personal information become a serious concern because of malicious use of information of users.

Privacy: Utilization and governance of personal data of individual's are under privacy concern of big data. Privacy ensures to make new strategies to protect personal information of user's and also make sure that information is being gathered, shared and used in correct means. Privacy is to make own information available to authorized people rather than all the people. Privacy can be disturbed when personal information can be identified through third party.

Security: Security is basically protecting data using different techniques, process from illegal usage, reviewing, discovering, splitting, alteration, recounting and extinction [6].Although security is necessary for securing data but consigning privacy is inadequate.

VARIOUS TECHNIQUES

There are several techniques to preserve privacy of personal information. We have studied some of the techniques for privacy preservation of sensitive information in big data. In this report, twenty four works has been analyzed from 2014-2019. Different techniques are discussed in those works such as k-anonymity, l-diversity, t-closeness and differential privacy in big data.

CONCLUSION

We have analyzed that in today's world everyone is producing vast amount of personal data. Data is producing in different forms and in a large amount. Privacy of individual's personal data from attackers is needed so that any third party cannot use the personal information for their own benefits. Several techniques are used for preserving privacy of sensitive information of users in big data such as k-anonymity, l-diversity, t-closeness and differential privacy. All above approaches for privacy preservation has some merits and demerits. Preserving privacy is essential because, beneficiary can use sensitive information of individual's and integrating it with public information to reveal a person specifically.

REFERENCES

[1] W. Fang, X. Z. Wen, Y. Zheng and M. Zhou, "A Survey of Big Data Security and Privacy Preserving," IETE, pp. 1-17, 2016.

[2] A.Khanan, S. Abdullah, A. H. H. M. Mohamed, A. Mehmood and K. Z. Ariffin, "Big Data Security and Privacy Concerns: A Review," Springer, pp. 55-61, 2019.

[3] M.Srinuvasu, A. Koushik and E. S. , "Big Data: Challenges and Solutions," JCSE, vol. 05, no. 10, pp. 250-255, 2017.

[4] S.Sathyamoorthy, "Data Mining and Information Security in Big Data," International Journal of Scientific Research in Computer Science and Engineering, vol. 5, no. 3, pp. 86- 91, 2017.

[5] P. Jain, M. Gyanchandani and N. Khare, "Big data privacy: a technological perspecctive and review," Springer, pp. 1-25, 2016.

[6] D. P. Goswami and M. S. Madan, "A Survey on Big Data & Privacy Preserving Publishing Techniques," vol. III, pp. 395-408, 2017.

[7] L.Hongling, "Research On Solutions To Privacy Security Problems Based On Big Data," in IEEE, Guangzhou, 2019.

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

[9] P. Charles, I.Carol and S.Mahalakshmi, "Big Data Security an Overview," IRJET, vol. 05, no. 02, pp. 130-134, 2018.

[10] T. Karle and P. D. Vora, "Privacy Preservation In Big Data Using Anonymization Technique," in IEEE, Pune, 2017.

[11] J.Vinothkumar and V. Santhi, "A Study on Privacy Preserving Methodologies in Big Data," ijst, vol. 9, pp. 1-16, 2016.

[12] J. Vasa and P. Modi, "Review of Different Privacy Preserving Techniques in PPDP," IJETT, vol. 59, no. 5, pp. 223-227, May 2018.

[13] B.Sreevidya, M.Rajesh and T.Sasikala, "Performance Analysis of Various Anonymization Techniques For Privacy Preservation of Sensitive Data," springer, pp. 687-693, 2019.

[14] N.Maheshwarkar, K. Pathak and V. Chourey, "Privacy Issues for K-anonymity Model," IJERA, vol. 1, no. 4, pp. 1857-1861.

[15] N. Victor and D. Lopez, "Privacy models for big data: a survey," IJBDI, vol. 3, pp. 61-75, 2016.

[16] Athiramol.S and Sarju.S, "A Survey on Data Anonymization for Big Data Security," Journal for Research, vol. 03, no. 01, pp. 88-91, 2017.

[17] Ms.Simi, M. Sankara Nayaki and D. Elayidom, "An Extensive Study on Data Anonymization Algorithms Based On K-Anonymity," in IOP, 2017.