MJRET
Open Access
Multidisciplinary Journal of Research in Engineering and Technology
Journal homepage: www.mjret.in
ISSN:2348 - 6953

REVIEW ON DIFFERENTIAL PRIVACY OF REAL TIME DATA

Ms.Neeta Pratap Patil, Prof. Nitin Hambir
Computer Dept.
Dr. D. Y. Patil School Of Engineering(Affiliated to Savitribai Phule Pune University)
Pune, India
patilneeta.333@gmail.com, nitin.hambir@dypic.in

Abstract : *Now a days performing data mining of real time data has great value. Generally it has great value in traffic congestion applications. Data values are highly correlated so providing differential privacy is necessary. So while preserving the privacy, the purpose of statistical database is to allow third parties or users to study the properties of traffic congestion applications or contents in the databases but individuals contribution is keep safe from harm. Trusted server carry the database or private information of individual users. User make query to the trusted server for accessing some data. Trusted server sends the requested data to user in the form of aggregate function form. This aggregate function is formed as answer to the query plus some random noise. So to provide strong privacy guarantee all appreciable data is examine methodically.*

Keywords*: Differential privacy, trusted server, statistical database,differentially private histogram*

1. INTRODUCTION

When real time data is get shared, it has great value. Because many times such data is private data and needs to maintain privacy of this data. So real time data is mostly shared in the form of aggregate function. When real time data is shared in the aggregate function form security of that data is maintained properly. So to provide strong privacy guarantee all appreciable data is examine methodically. Over the past few years, privacy of data is preserved by maintaining some terms. Privacy is provided to databases, the general principles of science and other kind of data. This privacy is provided by means to increase

the correctness of queries from statistical databases while decreasing opportunity of identifying its records. In many data mining applications, real time data is provided in the form of aggregate function to provide the security to private data. Applications where such privacy can be provided are Disease Surveillance, monitoring of road traffic, medical records, voter registration information, email usage. The goal of this differential privacy are analysis of confidential data by preserving the privacy of data. Consider following examples:

Disease Surveillance: Providers of health care collects data from individual visitors. After that all this collected data is shared with third party, who maintain record of this data properly and securely. This third party has all the collected data. So third party may share this data to other parties. Other parties are may be trusted or not.

Monitring of traffic: Providers of GPS service collect data from individual users. This data is related to location, speed of vehicle and mobility. Again here providers of GPS service share data with third party which is trusted or not.

In recent years lot of research is done on privacy preserving of whatever data get published. Different terms or concepts are there to preserve the privacy of statistical databases. Among those differential privacy has lots of importance. As different applications like disease surveillance, bank database or other maintain lots of data in their databases. This data needs to collect and then stored. All stored data is maintained properly so that guarantee about privacy is provided. So preserving privacy of secure data is big question in front of different organizations and government. Attacker can do any type of attack to access the data from database. While making access of data available to user, strong privacy guarantee needs to provide. Trusted party done not have attackers background knowledge.

Different sites such as social networks, online retailers and search engines like Yahoo, Google make required data available to users to understand clearly social and economical benefits. When new data is available then these websites continuously updates their data over time.

2. RELATED WORK

The goal of differential privacy is to preserve the privacy of confidential data. Even though differential privacy provides the strong guarantee about data privacy, then also recent work in this [16] has shown that it is possible to analyze sensitive data. Through random perturbation differential privacy is achieved, here analyst first issues the query and then he receives a noisy answer. Small changes in the database like adding or removing a person's private data, affect the answer of query. This is get considered in query sensitivity measure.

2.1 Differentially Private Histogram:

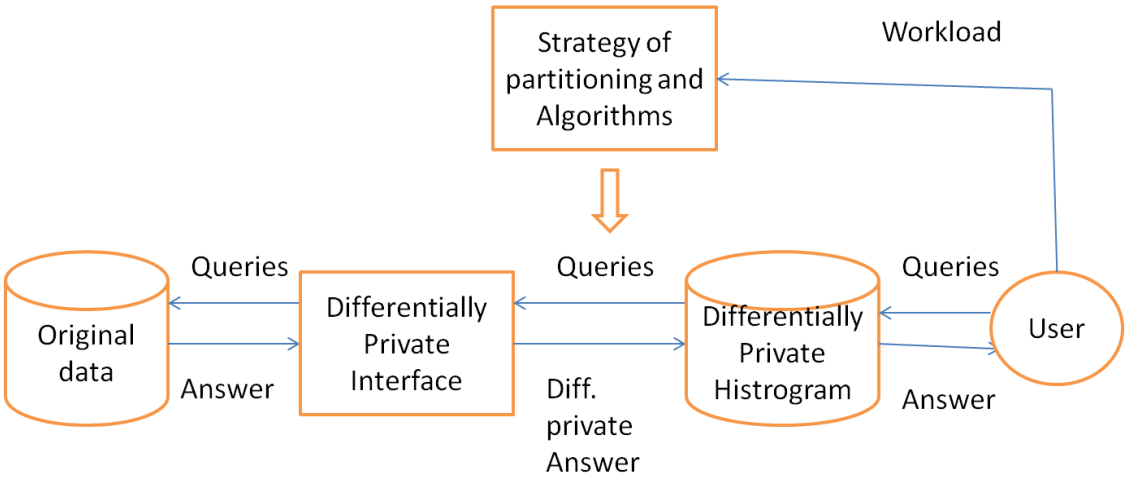

Figure 2.1 Differentially Private Histogram

Histogram means it is the partition of database points. Each database points belongs to different partitions. Private access to raw database is given by using the interface Privacy INtegrated Queries platform (PINQ) [10]. Differentially private histogram of raw database is produced by partitioning strategy implementing, which provides queries to the differentially private interface. Histogram then contains the raw database and synthesized dataset. Histogram bear all type of weight of Count queries and other OLAP queries.

In this paper[10] two approaches are studied:

- A baseline strategy
- A kd-tree partitioning

Kd tree partitioning has original ideas. Foremost in position is the distinctive measure in process of partitioning , in this approach errors are decreased. In second approach, they implement concept with the help of two-step algorithm which produces kd-tree partitions, where histogram generates from cell partitioning.

2.2 €-differential privacy

For queries related to individuals entries in the frequency matrix, Dwork et al.'s method [16] provides reasonable accuracy. Dwork et al.'s method, inject only small amount of noise in the query. But for aggregate queries, where large number of entries are envolved Dwork et al.'s method is not able to provide required results.

In paper [5], Xiaokui Xiao introduce a data publishing technique Privelet, which provides the differential privacy of data and for range count queries it provides accurate results.

Xiaokui Xiao[5] specified that several solutions are there to provide privacy guarantee, one of them is differential privacy. To give strong guarantee about privacy concept of differential privacy is used. Those data publishing methods provide differential privacy offer less usefulness. Consider count queries are answered through output dataset, then here noise is equivalent to total number of tuples in the dataset. In this [5],to provide accurate answer to range queries data publishing techniques are studied. Here Privelet data publishing method is presented to give guarantee about differential privacy.

2.3 Private and Continual Statistics

In this paper [9], it is specified that as data is continuously updated in the database, the websites provides upgraded data over time while preserving the privacy. Consider different websites, such as online retailers, different search engines like Yahoo, Google and many social networks like facebook, flicker etc. all this provides data in the form of aggregate function. When user make request for any data on search engine like Google, that search engine provide requested data in the form of aggregate function and provides guarantee of data on the server.

2.4 Fast Fourier Transform

Papadimitriou et al.[12] studied to achieve balance between time series compressed property and deviation caused by an outside influence. Based on Fast Fourier Transform and Discrete Wavelet Transform two algorithms are developed so that time series frequencies are disorder.

The drawback of this system is addition of noise in data does not give guaranty about differential privacy. It causes attacker or third party which is un-trusted, can get knowledge about stored data on trusted server.

3. CONCLUSION

In this work, we view the problem of database privacy and also get different new ideas about how to provide differential privacy to individuals private data. In many organizations the main issues is maintaining the security of confidential data. As large amount of data is collected on a server, it is necessary to give guarantee about data confidentiality. This review specifies data publishing technique Privelet, which provides the differential privacy of data and for range count queries it provides accurate results.

ACKNOWLEGEMENT

I wish to express my sincere thanks to the guide Prof. Nitin Hambir and Head of Department, Prof. Soumitra Das , as well as our principal Dr.S.S.Sonavne , also Grateful thanks to our

PG Coordinator Prof.P.M.Agarkar and last but not least, the departmental staff members for their support.

REFERENCES

[1] A. Blum, K. Ligett, and A. Roth, "A learning theory approach to non-interactive database privacy," in Proc. 40th Annu. ACM STOC, New York, NY, USA, 2008.

[2] C. Dwork, F. Mcsherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in Proc. 3rd TCC, New York, NY, USA, 2006, pp. 265–284.

[3] M. Hay, V. Rastogi, G. Miklau, and D. Suciu, "Boosting the accuracy of differentially private histograms through consistency,"PVLDB, vol. 3, no. 1–2, pp. 1021–1032, Sept. 2010.

[4] X. Xiao, G. Bender, M. Hay, and J. Gehrke, "iReduct: Differential privacy with reduced relative errors," in Proc. ACM SIGMOD, Athens, Greece, 2011, pp. 229–240.

[5] X. Xiao, G. Wang, and J. Gehrke, "Differential privacy via wavelet transforms," IEEE Trans. Knowl. Data Eng., vol. 23, no. 8, pp. 1200–1214, Aug. 2011..

[6] A. Doucet, N. de Freitas, N. Gordon, and A. Smith, Sequential Monte Carlo Methods in Practice. New York, NY, USA: Springer, 2001.

[7] S. Papadimitriou, F. Li, G. Kollios, and P. S. Yu, "Time series compressibility and privacy," in Proc. 33rd Int. conf. VLDB, 2007, pp. 459–470.

[8] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, "State estimation in discrete-time linear dynamic systems," in Estimation with Applications to Tracking and Navigation. New York, NY, USA: Wiley, 2002, pp. 199–266.

[9] T.-H. H. Chan, E. Shi, and D. Song, "Private and continual release of statistics," in Proc. 37th ICALP (2), Bordeaux, France, 2010, pp. 405–417, LNCS 6199.

[10] R. Chen, N. Mohammed, B. C. M. Fung, B. C. Desai, and L. Xiong, "Publishing set-valued data via differential privacy," PVLDB, vol. 4, no. 11, pp. 1087–1098, Aug. 2011.

[11] P. Brockwell and R. Davis, Introduction to Time Series and Forecasting. New York, NY, USA: Springer, 2002.

[12] S. Papadimitriou, F. Li, G. Kollios, and P. S. Yu, "Time series compressibility and privacy," in Proc. 33rd Int. conf. VLDB, 2007, pp. 459–470.

[13] R. E. Kalman, "A new approach to linear filtering and prediction problems," Trans. ASME J. Basic Eng., vol. 82, pp. 35–45, Mar. 1960.

[14] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, "State estimation in discrete-time linear dynamic systems," in Estimation with Applications to Tracking and Navigation. New York, NY, USA: Wiley,2002, pp. 199–266.

[15] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in TCC, 2006, pp. 265–284.

[16] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In TCC, 2006.