*Impact Factor (SJIF): 5.301*

**International Journal of Advance Research in Engineering, Science & Technology**

*e-ISSN: 2393-9877, p-ISSN: 2394-2444*

***Volume 6, Issue 3, March-2019***

# PRIVACY PRESERVING ANALYSIS OF MULTIDIMENSIONAL DATA

MS.SHEJAL GAIKWAD, MS.SHAISTA ZAMEER, MS.ARCHANA GHANWAT,
MR. AVINASH PALAVE
Department of Computer Engineering.
K.J's. Educational Institute, Trinity College of Engineering and Research, Pune, Maharashtra, India

**ABSTRACT**
*Data management context for big data research is represented by the issue of effectively and efficiently supporting analytics over big data a collection of models, algorithms and techniques oriented to extract useful knowledge from Cloud-based big data repositories for decision making and prediction purposes, e.g. by means of multidimensional data analysis paradigms. Big Social Data: here, social data coming from different social networks, such as Twitter, Facebook, LinkedIn, etc. Big Legacy Data: this big data source stores legacy data coming from legacy applications, such as government data, work data, Big Event Data: here, event data, such as calendar data, organization data, contact data, and so forth, are located; Big Pro le Data: this big data source stores pro le data coming from users activities, such as smart-phone data, network data, The privacy of big data is provided.*

*Keywords- Big data: privacy, DRIPROM, networks, big data, summary, dataset etc.*

## 1. INTRODUCTION

In the future we expect to improve this research in three aspects: (1) providing a more specialized review of frameworks and programming languages for Big Data analysis to extend the information about the features, benefits and limitations of each framework and language to provide a set of guidelines, and proofs of concept to expose the functionality of each programming language or framework; (2) including a review of works published in workshops, technical reports and white papers, and (3) considering papers from other scientific databases.

## 2. Literature Survey

Name - QoE-Driven Big Data Management in Pervasive Edge Com-puting Environment.

Author - I. Oakley and A. Bianchi

Description - In the age of big data, services in the pervasive edge environment are expected to offer end-users better Quality-of-Experience (QoE) than that in a normal edge environment. However, the combined impact of the storage, delivery, and sensors used in various types of edge devices in this environment is producing volumes of high-dimensional big data that are increasingly pervasive and redundant.

Name - Edge-based mining of frequent subgraphs from graph streams

Author - Alfredo Cuzzocreaa , Zhao Hanb, Fan Jiangb, Carson K. Leungb,∗ , Hao Zhangb

Description - In the current era of Big data, high volumes of valuable data can be generated at a high velocity from high-varieties of data sources in various real-life applications ranging from sensor networks to social networks, from bio-informatics to chemical informatics.

## 3. Proposed System

The system will collect the large amount of data and storing it into the distributed database and on that data system will generating the pedigree/summary.

The results of the proposed techniques are quit promising and contributing the sig-ni cant improvement over the existing techniques. However, some directions for future work are given below to further improve the overall performance of the sys-tem: Since the main objective of database distribution design is to improve system performance and throughput. Concurrency control mechanism has to be incorporated in the cost model. The problem here is how to incorporate the information related to concurrent queries within the cost model. Security issues should be included in future improvements since distributed databases are frequently used for electronic commerce and encryption of the data is applied during the transactions of electronic commerce. The problem is how to calculate cost a query when the data associated with the query is encrypted. Static data allocation cost model can be extended for more complex queries.

## 4. Algorithms to be used

**1. Sentence level Algorithm:**

Step 1: Enter the user query.

(E.g. What is java.)

Step 2: Apply Steaming and stopping (e.g., Remove stop words. Here in this example, ‘what’ and ‘is’ words will be removed)

Step 3: Remaining words will be search in another document file in database.

Step 4: result will be displayed as sentences where the keyword is present.

**2. LDA: Topic modeling** is a type of statistical modeling for discovering the abstract “topics” that occur in a collection of documents. Latent Dirichlet Allocation(LDA) is an example of topic model and is used to classify text in a document to a particular topic. It builds a topic per document model and words per topic model, modeled as Dirichlet distributions.

## 5. System Architecture

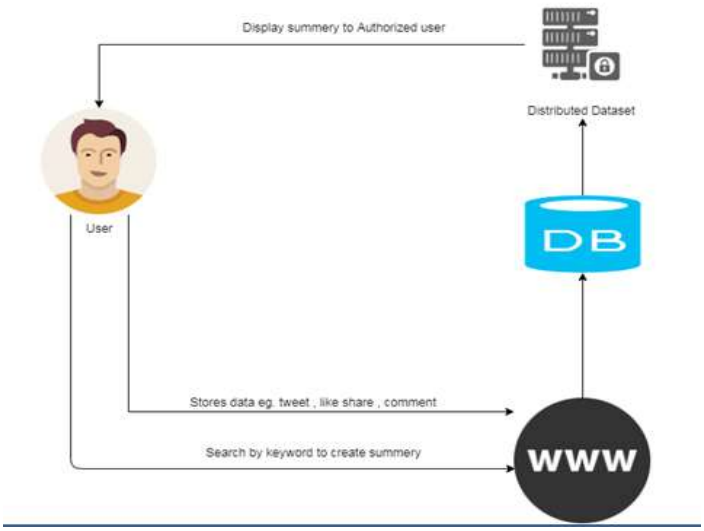


**Fig 1: System architecture**

## 6. Conclusion

We will collect the large amount of data and storing it into the distributed database and on that data we will generate the pedigree/summary.

## 7. REFERENCES

1. S. Sood, A. Sarje, and K. Singh, Cryptanalysis of password authentication schemes: Current status and key issues, in Methods and Models in Computer Science, 2009. ICM2CS 2009. Proceeding of International Conference on, Dec 2009, pp. 17.

2. A. Cuzzocrea, Privacy and security of big data: Current challenges and fu-ture research perspectives, in Proceedings of the First International Workshop on Privacy and Secuirty of Big Data, PSBD@CIKM 2014, Shanghai, China, November 7, 2014, 2014, pp. 4547. [Online].

3. A. Cuzzocrea and E. Bertino, A comprehensive theoretical framework for pri-vacy preserving distributed OLAP, in On the Move to Meaningful Internet Sys-tems: OTM 2014 Workshops - Confederated International Workshops: OTM Academy, OTM Industry Case Studies Program, C and TC, EI2N, INBAST, ISDE, META4eS, MSC and OnToContent 2014, Amantea, Italy, October 27-31, 2014. Proceedings, 2014, pp. 117136. [Online].

4. A. Cuzzocrea, Privacy-preserving big data management: The case of OLAP, in Big Data - Algorithms, Analytics, and Applications., 2015, pp. 301326. [Online]. Available: http://www.crcnetbase.com/doi/abs/

5. D. Wu, B. Yang, H. Wang, C. Wang, and R. Wang, Privacy preserving multimedia big data aggregation in large-scale wireless sensor networks, TOMC-CAP, vol. 12, no. 4s, pp. 60:160:19, 2016.

6. J. Mai, Big data privacy: The data action of personal information, Inf. Soc., vol. 32, no. 3, pp. 192199, 2016.

7. C. Lettner, M. Pichler, W. Kirchmayr, F. Kokert, and M. Habringer, "Rdfreduce: Customized aggregations with provenance for RDF data based on an industrial use case," in The 15th International Conference on Information Integration and Web-based Applications & Services,

8. F. Giunghiglia and M. Reyad, "Provenance in open data entity centric aggregation," in Provenance and Annotation of Data and Processes - 5th International Provenance and Annotation Workshop,

9. T. Debatty, P. Michiardi, O. Thonnard, and W. Mees, "Scalable graph building from text data," in Proceedings of the 3rd International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, Big Mine 2014, New York City, USA, August 24, 2014, 2014,

10. H. Wang, N. Li, J. Li, and H. Gao, "Parallel algorithms for flexible pattern matching on big graphs,"

11. W. Fan and C. Hu, "Big graph analyses: From queries to dependencies and association rules," Data Science and Engineering, vol. 2, no. 1, pp. 36–55, 2017.

12. N. Kushwaha and M. Pant, "Link based BPSO for feature selection in big data text clustering," Future Generation Comp.