iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** XI **Month of publication:** November 2014

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Secure Private Data Release in Vertical Partitioned Data

Kanchan Kauthale[#1], Lakshmi Madhuri[#2]
[1,2]Department of Computer Engineering, Dr D.Y.Patil School of Engineering,
Savitribai Phule Pune University, India

Abstract— When mining the useful information some privacy preserving data publishing addresses the problem of exposing the sensitive data. In this paper, the source database table is divided into some parties which hold different attributes for the same set of individuals where we address the problem of private data publishing. For this we extend the differential privacy model and proposed the algorithm for vertically partitioned data which guarantees that the other party can't derive extra information from the answered query. This method will provide the security for the private data which is released from the scattered framework.
Keywords— Differential privacy, secure data integration, classification analysis, Security

I. INTRODUCTION

In today's world there is a rapid development in the organizations due to this the customer satisfaction is one of the important aspect. Therefore to satisfy the customer's need, we have to understands and analyze their requirements. For this, organizations provide different strategies to fulfill the customer's need. One of the most important organizations like bank plays an important role in day today's life. Bank provides many services to a customer. Hence, to provide these services to the customer within a given allotted time, banks are searching and trying to implement new strategies in there organization.

To provide the better services like loan, insurance, credit card etc. The bank owns several organizations for instance, finance industry which is owned by the bank. Banking system will provide the customer's data to the organization which will interact with customers, like customer care services that will call the customer to inform them about the services and policies like loan, insurance, credit card that the bank can provide. For this bank will partitioned the customer's data so that only the basic information regarding the customer will be forwarded to the organization like the customer name and the contact number. The private data such as account details will be isolated from the organization inorder to secure it.

We generalize the problem, for example suppose a bank X owns a Loan company Y. Both of them have different sets of attributes for the same set of individual which are identified by the customer's ID, like bank X owns database x having fields like ID, Job, Account_Balance where company Y owns database b having fields like ID, gender, income. For better decision about the credit card or the loan sanctions, both the entities have to merge their data. Suppose there is one more company Z like credit card who also wants to merge its data with X and Y, so that the company X, Y, Z all receives the final combined data. Like Party X and Party Y will join their database x, y and we will discover the sensitive data to the other entity. At the same time the new information that is result from the integration of the databases should not be misused by other organization. In this case the party X and Y do not contain any specific information regarding the customer still the integration of the data make the chances of recognizing the customer profile. This will be a threat to the customer's private data.

To overcome this problem some partitioned techniques are used like vertical partitioning and the horizontal partitioning. In horizontal partitioning, the rows are partitioned into multiple tables with same columns in table. Whereas, in case of vertical partitioning, the columns are divided into multiple tables with same rows in table. In case of horizontal partitioning we reduce the load of the database by distributing it to the other locations, due to this the performance and the availability of the database will be increased. However in case of vertical partitioning different attributes for each data are placed on different locations or the partitions. Vertical partitioning will increases the security for the private data and also reduce the concurrent access of the data. So to provide the security in the system like bank we are using the vertical partitioning techniques.

In this vertical partitioning we will divide the database into two parties suppose party X and the party Y. Party X will contain all the frequently used details of the customers i.e the profile information like name , contact number, whereas party Y will contain the sensitive information about customer i.e account details etc. So in our scenario we have to provide the security for the partitioned data which contain the private information. Bank will provide only the profile information to customer care company so that it will call to the customer for various services but the company never knows the account details of the customer.

To provide the security for this private data we are using some techniques to hide the sensitive data from the other party. In this paper we pick up the differential privacy model [1].

Newly proposed some techniques for differentially privacy model for the vertically partitioned data which guarantee that only the necessary information is provided to the other party and the sensitive data is protected from the other party. In this paper we providing security for the release scattered private data for the vertically partitioned data between two entities. For this we hold the single-party algorithm for differentially privacy that has been proposed by Mohammed et.al [2] and extend this to multi party setting. This proposed algorithm satisfies parties execute the algorithm but may not analyze the additional data for other party.

II. RELATED WORK

Privacy of the data is one of the effective topic in the database, security for the recent year. [3]
In existing system there are different scenarios used like
Interactive mechanism in this data requester fetch the queries through some private mechanism and administrator reply the answer queries in response. Limitation of this is that it can only answers the liner number of queries. Otherwise the attacker will try to construct the original data

A. In case of non-interactive mechanism the owner of the data first analyzes the data and then publishes. Once the data is published the data can't be change by the owner. This mechanism is known as privacy preserving data publishing (PPDP) [3].

Some limitations of the non-interactive mechanism is that sometime this mechanism does not answers the queries appropriately because the data receiver can't construct a query for data mining in short period of time. For this an attacker can develop unlimited queries so this model cannot achieve privacy model as that of interactive.

B. In case of distributed mechanism the data may be gain by one party or the many party but the owners of the data want to gain the same task as the unique one party without sharing their data with other parties.[3]

Some anonymization algorithms are proposed like

1) Optimal Anonymization Algorithms: In this algorithm we are finding some optimal anonymization for given data. But there is one limitation is that in this algorithm we can't find the optimal solution for the huge data sets. To overcome this new algorithm is introduced [3]

2) Minimal Anonymization Algorithms: This algorithm finds the minimal solution but it is not given an appropriate solution when more than 3 attributes are taken. But some of them fulfill the goal of the classification analysis. [3]

Differential privacy [1] model proposed the alternative for the partitioned-based privacy model for PPDP. Many of the research related to the differential privacy [1] focuses on the goal of reducing the added noise which is received during the data mining results on interactive setting [4].In case of the non-interactive mechanism it only hold the single-party mechanism. Therefore the proposed techniques do not satisfy the requirements of the privacy model. [5][6]. Privacy preserving distributed data mining (PPDDM) [7] is one of the proposed approach in which many data owners calculate their inputs without sharing data with other parties. The function used in this scenario is like a clustering, classification etc. Many organizations like bank wants the finance data for analysis some customer's record. For this different techniques have been proposed for data mining including association rule [8], classification [9], clustering [10]. But all these proposed techniques do not provide any privacy guarantees on calculated output. Dwork et al. and Narayan [11] [12] proposed the differentially private queries for horizontal and vertical partitioned data. In this case the non-interactive approach is more flexible than the interactive approach because in non-interactive the data receiver can analyze and exploration their data. Clifton and Jiang [13] have suggested the techniques for distributed k-anonymity framework which securely divide the database into two parts and also fulfill the k-anonymization requirements.

III. PROPOSED APPROACH

To overcome the problems of existing system we proposed new techniques for the differential privacy model [1] .In our scenario we proposed some algorithms for the distributed and the non-interactive mechanism.
Fig.1 shows the system architecture of our proposed approach. In this scenario we first find the source database in our case it is bank database. Then we analyze the attributes which are sensitive in the database. After analysis we then find the dependency between each attribute with other one. Once we finalize that then we make a vertical partitioning of the source data in such way that the one part of the partitioning contain the common attributes of the customers profile whereas the other part of the vertically partitioning data contain the account attributes or the sensitive data. To partitioning in a precise manner we are using our proposed algorithm of exponential partitioning mechanism in this algorithm the integrated data which is generated from the vertically partitioned data satisfies the requirements of the differential privacy. Also our algorithm guarantees that it satisfies the security in the secure multiparty computation literature [13]. After partitioning by using algorithm then we are using the privacy for release scattered private data algorithm in this algorithm we provide security for the data which is release from the distributed framework.

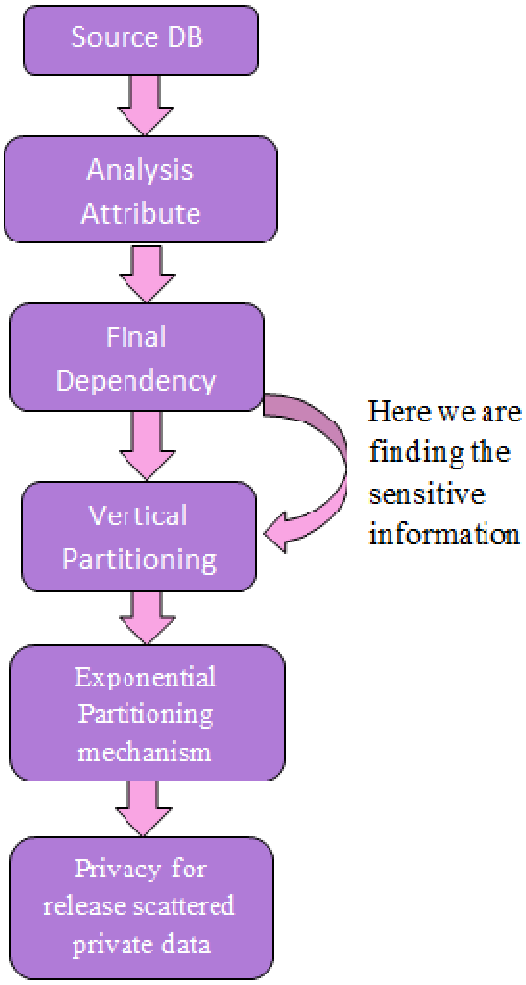


Fig.1.Sysyem Architecture

In our proposed approach we are having mainly three modules as follows:

- Vertically partitioning on given database
- Exponential partitioning mechanism
- Privacy for the release scattered data

In the first module of the vertically partitioning we partitioning our source data that is the bank data then we have two entities the first entity will hold the profile data which not that much sensitive and the other entity hold the sensitive data. In second module of the exponential partitioning mechanism algorithm we integrated the data which is generated from the vertically partitioned data that satisfies the requirements of the differential privacy. In last module of the privacy for the release scattered data we are providing security for the data which is release from the distributed framework.

IV. CONCLUSIONS

We proposed that the vertically partitioned data by using our proposed algorithm of the exponential partition mechanism ensures that the algorithm satisfy the differential privacy model and also secures the data in the distributed framework. Our second algorithm guarantees that the private data which is released from the scattered framework are secured. Also algorithm provides better data utility than the single-party algorithm [2] and the distributed k-anonymity algorithm [14].

V. FUTURE WORK

The algorithm proposed for the vertically partitioning data can be extended to multiparty by modifying the algorithms. To find out the optimal vertical partitioning we can use some more extended mechanism. Also to improve the query response time we can propose some new algorithms. In this paper we are using the SMC [13] model which provide security for mutual data

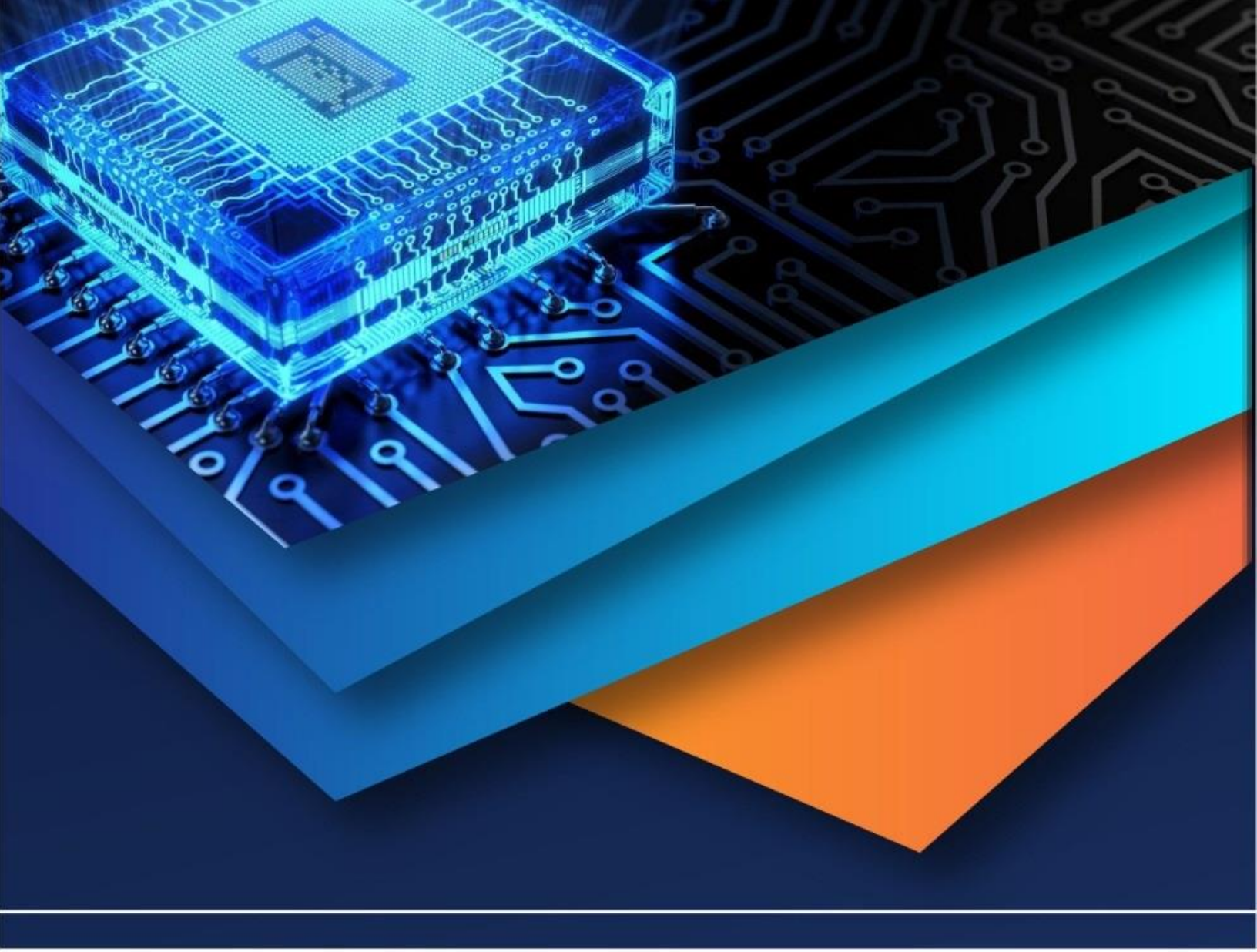
sharing but when the other party is trying to read more information from the answered query then for this we can extend this algorithm to secure the virulent parties.

VI. ACKNOWLEDGMENT

I take this opportunity to thank all in individuals for their guidance, help and timely support .It gives me great pleasure and immense satisfaction to present this paper. Which result of unwavering, Support expert guidance and focused direction of my guide Prof. Lakshmi Madhuri to whom I express my deep sense of gratitude and humble thanks, for valuable guidance throughout the work.

REFERENCES

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating Noise to Sensitivity in Private Data Analysis," Proc. Theory of Cryptography Conf. (TCC '06), 2006.
[2] N. Mohammed, R. Chen, B.C.M. Fung, and P.S. Yu, "Differentially Private Data Release for Data Mining," Proc. ACM Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '11), 2011.
[3] B.C.M. Fung, K. Wang, R. Chen, and P.S. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Developments," ACM Computing Surveys, vol. 42, no. 4, pp. 1-53, June 2010.
[4] I. Dinur and K. Nissim, "Revealing Information while Preserving Privacy," Proc. ACM Symp. Principles of Database Systems (PODS '03), 2003.
[5] B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar, "Privacy Accuracy, and Consistency Too: A Holistic Solution to Contingency Table Release," Proc. ACM Symp. Principles of Database Systems (PODS '07), 2007.
[6] C. Dwork, "A Firm Foundation for Private Data Analysis," Comm. ACM, vol. 54, no. 1, pp. 86-95, 2011
[7] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M.Y. Zhu, "Tools for Privacy Preserving Distributed Data Mining," ACM SIGKDD Explorations Newsletter, vol. 4, no. 2, pp. 28-34, Dec. 2002
[8] J. Vaidya and C. Clifton, "Privacy Preserving Association Rule Mining in Vertically Partitioned Data," Proc. ACM Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '02), 2002.
[9] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining," J. Cryptology, vol. 15, no. 3, pp. 177-206, 2002
[10] J. Vaidya and C. Clifton, "Privacy-Preserving k-Means Clustering over Vertically Partitioned Data," Proc. ACM Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '03), 2003.
[11] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our Data Ourselves: Privacy via Distributed Noise Generation," Proc. 25th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT '06), 2006.
[12] A. Narayan and A. Haeberlen, "DJoin: Differentially Private Join Queries over Distributed Databases," Proc. 10th USENIX Conf. Operating Systems Design and Implementation (OSDI '12), 2012.
[13] W. Jiang and C. Clifton, "A Secure Distributed Framework for Achieving k-Anonymity," Very Large Data Bases J., vol. 15, no. 4, pp. 316-333, Nov. 2006.
[14] N. Mohammed, B.C.M. Fung, and M. Debbabi, "Anonymity Meets Game Theory: Secure Data Integration with Malicious Participants," Very Large Data Bases J., vol. 20, no. 4, pp. 567-588, Aug. 2011.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY