ISSN: 0975-766X
CODEN: IJPTFI
Available Online through *Research Article*
www.ijptonline.com

IMBRICATING SLICING AND SHUFFLING IN DATABASES

Ishwarya M.V[1*], Dr.K.Rameshkumar[2]
Research Scholar, HITS, Assistant Professor, CSE Dept, Sri SaiRam Engineering College, West Tambaram, Tamil Nadu, India.
Professor, IT Dept, HITS, Tamil Nadu, India.

Received on: 03-02-2017 *Accepted on: 12-03-2017*

Abstract

The major threat in database is the leakage of information access pattern of queries. The hacker may find the access pattern of queries; he might try to steal the data. So shuffling was introduced in which the data records are shuffled thus enhancing the security of the database. If the hacker comes to know the access pattern of the database he can access the data, Shuffling prevents leakage of access pattern. As new patients arrive, the hospital new records are created and inserted in the database. Here third party involved is research center and care is taken to ensure that there is no loss of data.

Keywords: Slicing, Bucketisation, Generalisation, Imbricating Technique, Shuffling, Black Record, White Record.

1. Related Work

Data mining and knowledge discoveries are the two recent research areas which enquire the extraction of large quantity of data of unknown patterns. Data mining is where non-trivial and useful knowledge are drawn out from large databases .Sequential data mining techniques are successful in various areas such as science, engineering and medicine. But the need for parallel and distributed data mining has been in existence over the past years. Data mining research is concerned with obtaining information from various areas such as bio informatics, customer relationship management etc. The information obtained can be in the form of patterns or clusters. Consider the example where association rules in a super market which gives the relationship among the items bought together. Here customers could be clustered in form of segments.

Data mining techniques has its use in various security applications for identifying behavior, link analysis which deals with multi -part databases.

Table 1. Literature Survey.

LITERATURE SURVEY

TITLE	AUTHOR	JOURNALS	FINDINGS	SCOPE FOR THE FUTURE WORK
- Database Access Pattern Protection Without Full Shuffles. - A Review of Privacy Preserving Data Publishing Technique. - An Overview of sectional Shuffle for Database Access Pattern Protection Using Reverse Encryption Algorithm. - A Review Paper on sectional Shuffle for Database Access Pattern Protection Using Reverse Encryption Algorithm.	- Xuhua Ding, Yanjiang Yang, and Robert H Deng . - Amar Paul Singh, Mr.Dhanshri Parihar, Priti V Bhagat and Rohit Singhal. - Priti V Bhagat and Rohit Singhal.	- International Journal of Data Engineering (IJDE - International Journal of Emerging Research in Management &Technology. - An Overview of sectional Shuffle for Database Access Pattern Protection Using Reverse Encryption Algorithm. - International Journal of Application or Innovation in Engineering & Management (IJAIEM).	- A novel scheme in the same model with provable security, which only shuffles a portion of the database. - A systematic review of several anonymisation techniques such as generalization and bucketization, have been designed for privacy preserving micro data publishing. - A novel scheme in the same model with provable security, which only shuffles a portion of the database. - A new encryption algorithm, which we call Reverse Encryption Algorithm (REA). Our new encryption algorithm (REA) is simple and is fast enough for most applications.	- With a secure storage storing thousands of items, our scheme can protect the access pattern privacy of databases of billions of entries, at a lower cost than those using ORAM-based poly-logarithm algorithms. - It focus on effective method that can be used for providing high data utility and high security in high-dimensional data. - By the use of two interval and sectional-shuffle, our scheme avoids full-database shuffle and reduces the amortized server computation complexity and improves the performance. - Reverse Encryption Algorithm (REA)algorithm is simple and fast enough for most applications.algorithm REA can reduce the cost time of the encryption/decryption operations and improve the performance.

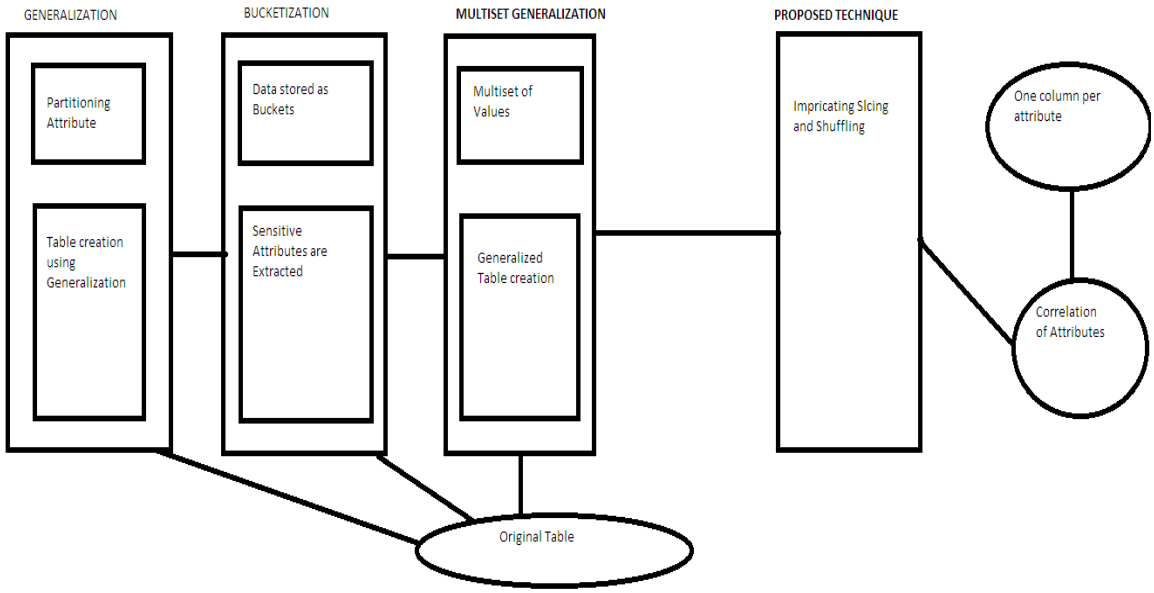
2. Proposed Architecture

When the membership disclosure is carried out the additional benefit is that the identity disclosure and attribute disclosure is also maintained as well as prevented. Sensitive information has to be maintained but when it is regarding or maintained for a particular individual it should be secretly maintained. And when this information is taken into consideration it should be clear that whether it is the necessary information about that particular person or a thing .When the details is regarding the person it should be clear and safe. So these conditions are taken into account slicing is the best technique that furnishes these above mentioned facilities and satisfies the needful to the best[4].

We have specified about privacy-publishing and preserving of data using full functional dependencies i.e. FFD. First of all functional dependency is the one which is used to sum up the left and right side dependencies to a certain level. This Fully Functional Dependencies will get into act only if the left side dependencies cannot be deduced further[1] . This FFD is mainly used to preserve the privacy of a record and which cannot be modified by a third party without any prior authentication. This is mainly important because when all people get access to all data we may get leaded to a situation where we won't get any privacy for our data and that will lead to plagiarism and damage of our own data. This include two phases which is used for deviating our idea from privacy and utility. Secondly it provides complete set of control over our data using full functional dependencies which overcomes the disadvantages of partial functional dependency. This also provides a full complete and an efficient algorithm that could help to find out the anonymous micro data that could result in low loss in information. This method also provides us with guidance for the path between privacy and data utility because we should not be able to access all the data in terms of utility and there should be a privacy among the files which we are using.

This involve partition of data both horizontally and vertically, this can be even explained in three forms i.e. Attribute partitioning, tuple partitioning and column generalization. This technique is mainly useful for total portioning too which permute the sensitive attribute values randomly. But this has main disadvantage, because while we are using slicing we will generally face data utility .So the proposed model of ours will help you in better privacy threats. So we recommend overlap slicing than to slicing which could lead to loss in data utility. This involves a better attribute called chi value which could give you the information about highly correlated attributes as in [9]. This value even lead us to chi matrix which is used to find the correlated attributes in the form of columns. These are mainly used to deliver us out ultimate goal which is data privatization. But in this method we have more efficient way of disclosure of attributes and membership attributes disclosure. But the overall methodology is that before anonymizing the data ,person or a record has to analyses the data characteristics in data anonymization. This could rationalize the design to get a better data .

Figure 1. Proposed Architecture



The figure 1 shown above gives you the proposed Architecture of Imbricating Slicing and Shuffling for privacy preservation in databases. We also include the experimental results in this conclusion, which is being retrieved from the database. In this the preprocessing steps must be applied on the table before the experiments on workload is done on the data. After computation, the pre-processor data the sensitive attribute and quasi identifiers are examined. After this step of identifying the pre-processor the modified technologies such as Mondrian diversity functions are employed for the overlapped sliced table. In order to measure the performance level of overlapping

slicing technique against several privacy threats such as identity, membership and attribute disclosure the accuracy of the method can be identified/measured. The accuracy of this machine can be determined by matching the fake tuple and buckets to the original data. The experiments demonstrate that overlapping slicing preserves better data utility, privacy in disclosures. While considering the performance of the methods discussed here ,it is said that the generalization and bucketisation are having lesser accuracy rate when compared to the slicing method. While considering the slicing the overlapping slicing has higher performance while considering the slicing as in terms of anonymization . Here the overlapping technique is having higher performance, which includes omitting the duplication of data.

3. Future Enhancements

Two popular anonymization technique that we are using are generalization and bucketisation which can be applied on the quasi identifiers which will replace the QI value with the less value as a process of converting the QI value of a data with almost equal values. This will result in more record with same set of quasi identifiers value.As in [8] two anonymization technique also uses two privacy preserving paradigm such as k-anonymity and l-diversity .There are three types of encoding scheme, local recording, regional recording and as well as global recording.

In this the global recording it has a practice of replacing multiple values with an average values. Consider where 23, 25, 34, 56, 25, 25 are data where 25 is being repeated for three times. So in order to avoid the repetition the global recording will take the average values of all the values and will replace at the places where the repetition will happen. This kind of anonymized data consists of sets of buckets, which will be sent by the data in permuted sensitive attributed values. In particular, bucketisation is designed to have and to handle multi – dimensional data values. But the slicing on the other hand proposes data utility more comparable to the traditional techniques which we are using and we used before[7]. Anonymization technique is a powerful method for privacy preservation of publishing the data. This paper presents a new technique on anonymization that includes slicing by overlapping with a privacy preservation and privacy model. While considering the performance of the methods discussed here, it is said that the generalization and bucketisation are having lesser accuracy rate when compared to the

slicing method. While considering the slicing the overlapping slicing has higher performance while considering the slicing as in terms of anonymization. Here the overlapping technique are having higher performance which include omitting the duplication of data. It is nothing but a new breed of non-uniform memory access (NUMA) systems, which has multi-socket servers of multicores. For future work we have provided heuristic approach in the first phase of the partition step. This involves comparison of the data set anonymized using different set of requirements. In the second step, we measure the utility loss rather measuring the utility gain but when we implement the privacy function using the FFD we could end up in utility loss. So our main motto in our future enhancement is that is to eliminate the utility loss by implementing full privacy too[6]. This is mainly important because we need both the gain and also the privacy which could develop our data consistencies. From this we can say that we can construct an initial partition for bottom-up approach which has been leaded by frequency distributers. Likewise, we can also use ideas that can be applied to top – down approach too. Finally, we are in our development plan where we could enhance our privacy loss in worst-case scenario and measure the total utility loss that we will get from our system models. Thus our main aim in this paper is to minimize the privacy loss for each individual systems and to decrease the utility loss for all the pieces of useful and helpful knowledge. The work in overlapping slicing has lead us to work even more on it in terms of future enhancements. First what we have proposed is that we have an attribute which will have exactly only one column. As an extension we proposed a technique called overlapping slicing which will allow us to duplicate an attribute in more than one column. This is mainly usually in attribute correlation. Say for example if one could duplicate an attribute called occupation in a database where the age, dept., place and occupation are given[8]. This will be helpful to a greater extent when the people referring to the place cannot view the unrelated information such as age and department. This will greatly be helpful in the trade – off between the privacy and data utility. Secondly we are researching to enhance the membership disclosure which we have seen in full functional requirements in a more appropriated and protected manner. Our research shows that the random grouping, which we have used in the previous model, is not useful and less effective. So we have dedicating ourselves to design a more efficient tuple grouping algorithms which we have already discussed.

4. Conclusion

Functional dependency is used to sum up the left and right side dependencies to a certain level. This Fully Functional Dependencies will get into act only if the left side dependencies cannot be reduced further . This FFD is mainly used to preserve the privacy of a record and which cannot be modified by a third party without any prior authentication. This is mainly important because when all people get access to all data we maybe leaded to a situation where we will not get any privacy for our data and that will lead to plagiarism and damage of our own data. This includes two phases which is used for deviating our idea from privacy and utility. Secondly it provides complete set of control over our data using full functional dependencies which overcomes the disadvantages of partial functional dependency. This also provides a full complete and an efficient algorithm that could help to find out the anonymous micro data that could result in low loss in information. This method also provides us with guidance for the path between privacy and data utility because we should not be able to access all the data in terms of utility and there should be a privacy among the files which we are using.

5. Acknowledgment

I thank my Supervisor Dr. K. Ramesh Kumar, Professor, HITS, PADUR for his Guidance and support for the Research work.

6. References

1. Dasseni, V. S. Verykios, A. K. Elmagarmid, and Elisa Bertino. Hiding Association Rules by Using Confidence and Support. In Proceedings of the 4th International Information Hiding Workshop (IHW). Pittsburg, PA. April 2001. pp.369-383.
2. Piatetsky-Shapiro. Knowledge Discovery in Personal Data vs. Privacy: A mini-symposium. In IEEE Expert, v.10, n.2. April 1995.
3. Schadow, S. J. Grannis and C. J. McDonald. Privacy-Preserving Distributed Queries for a Clinical Case Research Network. In Proceedings of the IEEE ICDM Workshop on Privacy, Security and Data Mining, Maebashi City, Japan, December 2002.
4. Gilburd, B. Schuster, A. Wolff, R. Privacy-preserving data mining on data grids in the presence of malicious participants. High performance Distributed Computing. 2004. Proceedings. 13th IEEE International Symposium on. 4-6 June 2004.

5. Polat and W. Du. Privacy-Preserving Collaborative Filtering Using Randomized Perturbation Techniques. In Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03). Melbourne, Florida, USA. November 2003.
6. B. D. Cabrera, L. Lewis, and R. K. Mehra. Detection and Classification of Intrusions and Faults using Sequences of System Calls. In SIGMOD Record, v.30, n.4. December 2001.
7. Vaidya and C. Clifton. Privacy Preserving Nave Bayes Classifier for Vertically Partitioned Data. In Proceedings of the 2004 SIAM Conference on Data Mining. Lake Buena Vista, Floria, USA. April 2004.
8. Vaidya and C. Clifton. Privacy-Preserving K-Means Clustering over Vertically Partitioned Data. In Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, DC, USA. August 2003.
9. Vaidya and C. Clifton. Privacy-Preserving Outlier Detection. In Proceedings of the Fourth IEEE International Conference on Data Mining (ICDM 2004). Brighton, UK. November 2004.
10. Jaideep Vaidya and Chris Clifton. Privacy preserving association rule mining in vertically partitioned data. Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. Edmonton, Alberta, Canada. 2002. Jiawei Han and Micheline Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann. April 2000.