IJRRECS/April 2014/Volume-2/Issue-4/2152-2157 **ISSN 2321-5461**

INTERNATIONAL JOURNAL OF REVIEWS ON RECENT ELECTRONICS AND COMPUTER SCIENCE

AN EXPOSURE TOWARDS ENHANCED DATA UTILITY INTENDED FOR SECURING MEMBERSHIP EXPOSURE

Nakkala Hema Sekhar Reddy[1], V. Jagadesh[2]

[1]M.Tech Student, Dept of CSE, Bandari Srinivas Institute of Technology, Chevella, R.R Dist., A.P, India

[2]Assistant Professor, Dept of CSE, Bandari Srinivas Institute of Technology, Chevella, R.R Dist., A.P, India

ABSTRACT:

For the past few years, the fast improvement of Internet technology and the privacy safeguarding of data has become one of the most important research topics and turn out to be a serious unease in publication of personal data. Several techniques are proposed for protecting individual privacy and sensitive information in order to avoid the obstacles. Encryption is another commonly employed technique for confidentiality protection but may not be directly applicable to some privacy preserving data publishing problems. Novel method of data anonymization known as slicing was introduced to get better the existing state of art. It protects confidentiality because it breaks the associations between unconnected attributes. By involving the sensitive attribute, slicing conserves improved utility of data compared to generalization and is more efficient than bucketization within workloads.

Keywords: Encryption, Slicing, Data anonymization, Generalization.

1. INTRODUCTION:

Besides limiting the usefulness of unnecessary inferences that may possibly be resulting from the data release, preservation of the data intended for a required data analysis. The essential thought of slicing is to break the connection cross columns to preserve the association within each column. Slicing which does not have horizontal partitioning is the first marginal publication

which can be viewed as a special case consequently; correlations among attributes in different columns are lost in marginal publication [4]. The data preserves better utility than generalization and bucketization and reduces the dimensionality. Slicing has some connections to marginal publication; both of them release correlations among a subset of attributes. The accessibility of high quality data and helpful information sharing depends on data mining and the main obstacle to the advancement of technology is the lack of trust in data mining [8]. By horizontal partitioning attribute correlations between different columns are preserved. By means of the region they are into, the multidimensional recoding which is also called regional recoding partitions the space of domain into the regions of non-intersect in addition to data points within the identical region are represented. By horizontal partitioning attribute correlations between different columns are preserved [1]. Many techniques are proposed for protecting individual privacy and sensitive information in order to avoid the obstacles. The current privacy protection practice primarily depends up on the policies and guidelines to confine the types of publishable information besides agreements on the applying as well as storage of sensitive information [11]. By the confidence bounding which is the first contribution, the privacy concern related to the input of data mining methods is addressed but the output of data mining methods could also cause confidentiality threats [3]. For confidentiality protection encryption is another commonly employed technique. For acting on the corresponding patient even though an encrypted record communicates to a real life patient, the encryption hides the semantics required. Even though it can be used to assume perceptive properties about record holders the output is an aggregate pattern, it is not intended to identify a record holder [14]. Depending on mining the accessibility of high quality data and eventual information sharing depends. With the intention that the published data remains almost functional while confidentiality of individual is preserved which is intended for publishing information in an additional hostile atmosphere a task of the utmost importance is to develop methods as well as tools [9]. Encryption intends to prevent an unofficial party from accessing the data, but enable an authorized party to have full access to the data. Data publishing in privacy preserving is the authorized party who may also play

the role of the adversary with the goal of inferring sensitive information from the data received [7]. As a consequence, encryption may not be directly applicable to some privacy preserving data publishing problems.

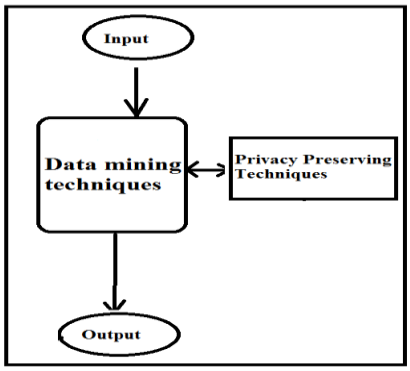

Fig 1: An overview of Privacy preserving data publishing

2. METHODOLOGY:

Novel method of data anonymization known as slicing was introduced to get better the existing state of art. It protects confidentiality because it breaks the associations between unconnected attributes, which are infrequent. By slicing the data set is partitioned both vertically and horizontally [2]. In data distribution privacy preserving has become one of the most significant research topics in the field of data safety and it has turn out to be a severe unease in publication of personal data in recent years. Due to the increased level of security after revolutionary attacks the public has an increased sense of privacy invasion [15]. In the data collection phase the data publisher collects data from record owners and releases the collected data to a data miner or to the public in the data publishing phase, called the data recipient, who will then conduct data mining on the published data. In analysis of data anonymizing classification is a primary difficulty. A classifier is required to access large collection of data [12]. The techniques of privacy-preserving are regularly essential to diminish the likelihood of identifying sensitive information concerning individuals, when a data set is unrestricted to other parties intended for data examination. By releasing person-specific data a threat may pose to an individual's privacy [5]. The initial involvement is to proficiently identify a k-anonymous solution which preserves the classification structure. By featuring the information level in a manner of top-down by considering both information and privacy, the search is done and this refinement algorithmic framework is highly efficient and natural for handling different types of attributes. For achieving both a privacy goal and a classification goal this approach exploits by the noise and

redundant structures in the data [10]. Generalization and bucketization are two popular anonymization techniques. Various schemes of encoding have been introduced for generalization such as: regional recoding, universal recoding and local recoding. Global recoding contains the assets of abundant occurrences of the similar value are constantly replaced through the similar comprehensive value [6]. Major problems concerning generalization are: due to the curse of dimensionality it fails on high-dimensional data; due to the uniform-distribution assumption it causes too much information loss. By means of arbitrarily permuting the sensitive attribute values in each bucket, with the responsive attribute bucketization initially partitions tuples within the table into buckets as well as subsequently divides the quasi identifiers. Besides limiting the usefulness of unnecessary inferences that may possibly be resulting from the data release, preservation of the data intended for a required data analysis [13]. For anonymizing information of high-dimension particularly bucketization has been intended. Similar to marginal publication overlapping vertical partitioning is left as our future work. In addition to permitting unusual incidents of the identical value to be generalized differently the local recoding does not have the above constraints. Different columns within each bucket, the values within every column are at random permutated. The partitioning of vertical is completed by means of attributes grouping into columns based on the associations between the attributes. By grouping tuples into buckets horizontal partitioning is done. To break the linking between each column is randomly permutated. To provide their personal information to the data publisher in the trusted model, the data publisher is trustworthy and record owners are willing however, to the data recipient the trust is not transitive [14]. Every data publishing scenario shown in fig1 has its own assumptions and requirements of the data publisher, the data recipients, and the data publishing purpose.

3. RESULTS:

Slicing conserves improved utility of data compared to generalization and is more efficient than bucketization within workloads by means of involving the sensitive attribute. As slicing breaks the associations between unconnected attributes, protects confidentiality. The limitation of

generalization and bucketization are overcome and preserves better utility by slicing while protecting against privacy threats. The relations between columns values of a bucket are randomly generated and this may lose data utility.

4. CONCLUSION:

In some data publishing scenarios, it is important that each published record corresponds to an existing individual in real life. The accessibility of high quality data and helpful information sharing depends on data mining and the main obstacle to the advancement of technology is the lack of trust in data mining. The techniques of privacy-preserving are regularly essential to diminish the likelihood of identifying sensitive information concerning individuals, when a data set is unrestricted to other parties intended for data examination. A novel technique of data anonymization known as slicing was introduced to preserve the association within each column the essential thought of slicing is to break the connection cross columns. Overlapping vertical partitioning which is left as our future work is similar to Marginal publication.

REFERENCES:

[1] D. J. Martin, D. Kifer, A. Machanavajjhala, J. Gehrke, and J. Y. Halpern. Worst-case background knowledge for privacy-preserving data publishing. In ICDE, pages 126–135, 2007.

[2] Tiancheng Li, Ninghui Li, Jian Zhang, Ian Molloy, "Slicing: A New Approach to Privacy Preserving Data Publishing" IEEE 2012 Transactions on Knowledge and Data Engineering, volume:24,Issue:3.

[3] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating Noise to Sensitivity in Private Data Analysis," Proc. Theory of Cryptography Conf. (TCC), pp. 265-284, 2006.

[4] A. Blum, C. Dwork, F. McSherry, and K. Nissim, "Practical Privacy: The SULQ Framework," Proc. ACM Symp. Principles of Database Systems (PODS), pp. 128-138, 2005.

[5] C. Dwork, "Differential Privacy," Proc. Int'l Colloquium Automata, Languages and Programming (ICALP), pp. 1-12, 2006.

[6] A. Inan, M. Kantarcioglu, and E. Bertino, "Using Anonymized Data for Classification," Proc. IEEE 25th Int'l Conf. Data Eng. (ICDE), pp. 429-440, 2009.

[7] V. Rastogi, D. Suciu, and S. Hong. The boundary between privacy and utility in data publishing. In VLDB, pages 531–542, 2007.

[8] J.H. Friedman, J.L. Bentley, and R.A. Finkel, "An Algorithm for Finding Best Matches in Logarithmic Expected Time," ACM Trans. Math. Software, vol. 3, no. 3, pp. 209-226, 1977.

[9] Y. He and J. Naughton, "Anonymization of Set-Valued Data via Top-Down, Local Generalization," Proc. Int'l Conf. Very Large Data Bases (VLDB), pp. 934-945, 2009.

[10] B.C.M. Fung, K. Wang, and P.S. Yu, "Top-Down Specialization for Information and Privacy Preservation," Proc. Int'l Conf. Data Eng. (ICDE), pp. 205-216, 2005.

[11] R. C.-W. Wong, J. Li, A. W.-C. Fu, and K. Wang. (α, k)-anonymity: an enhanced k-anonymity model for privacy preserving data publishing. In KDD, pages 754–759, 2006.

[12] B.-C. Chen, K. LeFevre, and R. Ramakrishnan, "Privacy Skyline: Privacy with Multidimensional Adversarial Knowledge," Proc. Int'l Conf. Very Large Data Bases (VLDB), pp. 770-781, 2007.

[13] G. Ghinita, Y. Tao, and P. Kalnis, "On the Anonymization of Sparse High-Dimensional Data," Proc. IEEE 24th Int'l Conf. Data Eng. (ICDE), pp. 715-724, 2008.

[14] C. Dwork, "Differential Privacy: A Survey of Results," Proc. Fifth Int'l Conf. Theory and Applications of Models of Computation (TAMC), pp. 1-19, 2008.

[15] J. Brickell and V. Shmatikov, "The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 70-78, 2008.