GJCRT : Global Journal of Computer Science Research & Technology
ISSN (online): 2349- 0608

Digital Dreams: Privacy Preserving And Cyber-Security Access Control Mechanism For Sensitive Information Using Hypercube

Jaspreet Singh[1] *Deepak Kumar*[2]
Assistant Professor, GSSDGS Khalsa College Patiala, Punjabi University, Patiala, India[1]
Research Scholar, Department of Computer Science, Punjabi University, Patiala, India[2]
jaspreetissaj@gmail.com[1], *dmehta1001@gmail.com*[2]

ABSTRACT

Data privacy refers to the concern of how to preserve the information in data cells. Privacy preserving deals with hiding the sensitive information without sacrifice the usability of data. Every organization today have become well aware of the privacy intrusions of their sensitive data and not want an unauthorized user to access all the information stored in data cube. The major area of concern is that non-sensitive data even may deliver sensitive information, including personal information, facts or patterns. Several techniques of privacy preserving data mining have been proposed in literature. In this paper, we have building a multi-viewing cube of data for analyzing, and applying various access control mechanism for privacy preserving of factual data, and to secure against the unauthorized access to sensitive information

KEYWORDS: **-** Data-mining, Privacy preserving, Sensitive data, Data-cube.

INTRODUCTION

Data mining is the process used to analyze large amount of data and gather useful information from them. It extracts the information from large heterogeneous databases in many different dimensions and finally summarizes it into categories and relations of data [1]. The primary goal of privacy preserving is to hide the sensitive data before it gets published. Organizations collect and analyze data to improve their services. Access Control Mechanismsis used to ensure that only authorized information is available to users. However, sensitive information, even after the removal of identifying attributes, is still susceptible to linking attacks by the unauthorized users.The privacy is achieved at the cost of accuracy and imprecision is introduced in the authorized information under an access control policy.

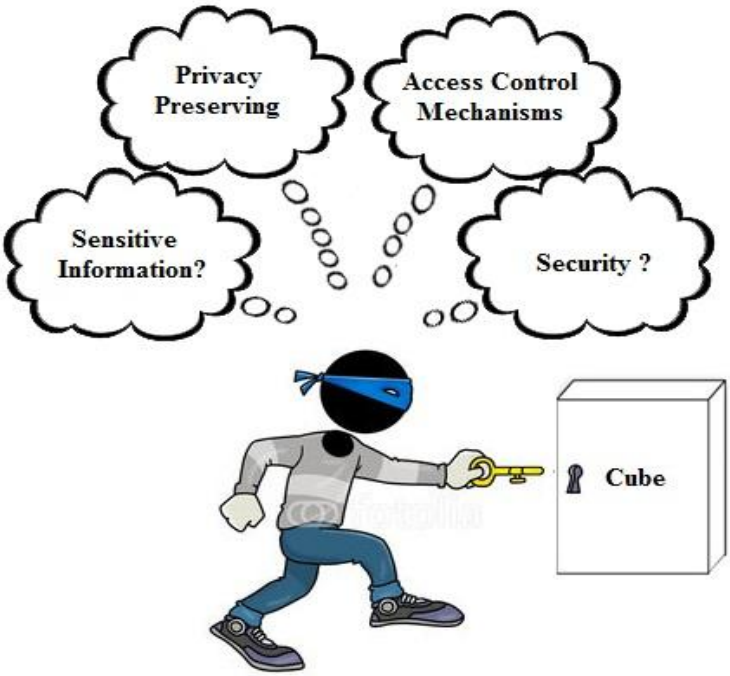


Figure 1: - Privacy Preservations and leaks of Sensitive data.

OBJECTIVES OF THIS RESEARCH STUDY

This proposed study has been focused on building a multidimensional cube and securing privacy preserving mechanism of sensitive data against unauthorized access control. Specifically, the following research objectives have been defined for the proposed study:

1. To perform demographic analysis of a dataset and identifying various dimensions, levels of dimension and building a multi-dimensional cube based on basic elements required for the structure of multidimensional cube.

2. To avail sensitive information and provide protection and preservation to sensitive information against unauthorized access.

In this paper, we discuss different approaches and techniques in the field of Privacy Preserving Data Mining. The paper is organized as. First, we give the basic concept of data mining and privacy.Second, we describe Privacy Preserving data mining with its framework.Then, define different techniques of privacy preservation. And finally,we describe conclusions.

PRIVACY PRESERVING DATA MINING

Privacy preserving has originated as an important concernwith reference to the success of the data mining. Privacypreserving data mining deals with protecting theprivacy of individual data or sensitive knowledge withoutsacrificing the usefulness of the data [2].

LITERATURE SURVEY

The review of literature based on various research articles and papers upon multi-dimensional modeling under privacy preserving mechanisms. The researcher arranges all the based studies into chronological order for granularity level refinements of research area

PRIVACY PRESERVING DATA MINING TECHNIQUES

In this, we focus on the different techniques which are developed like cryptographic techniques, data perturbation, blocking based, etc.

(a) Cryptographic Technique- *Cryptography* is a technique which sensitive data can be encrypted available in the databases. It is better technique to preserve the data.Cryptographic technique introduced because it provides security and safety of sensitive attributes. The cryptographic technique includes various encryption and decryption algorithms in order to secure the data so that no individual or no computer program can recognize that data [3]. The encryption is mostly usage in this technique and its result is more secure when data sent to the other party. But, this technique is no longer reliable when there are large numbers of parties sharing the data [4]. Also, these algorithms are unusable where millions of data are generated frequently. It can break the individual's privacy and can lead to attacks.

(b) Data Perturbation - Data Perturbation is a technique for modifying datausing random process. This technique sensitive data ischangesusing by adding, subtractingor any other mathematical formula [5] [6]. Data perturbation technique can handledifferent-different data types: character type, Boolean typeand integer type. In discrete data it isrequired to preprocess the original data set [5]. The preprocessingof data is classified into attribute coding and obtaining setscoded data set. The method of average data is used here. Discrete formula introduced [7] is: A (max) - A (min)/n =length. A is continuous attribute, n is number of discrete, andlength is the length of the discrete interval. The techniquedoes not reconstruct the original data values, it onlyreconstructs the distribution

(c)Blocking based technique -In blocking based technique, a sensitive classification rule which is used for hidingsensitive information from others. In this technique, there are twosteps which are used for preserving privacy. First is toidentify transactions of sensitive rule and second is to replacethe known data values to the unknown data values [8] [9]. In thistechnique, there is scanning of database andidentifying the transactions sensitive rule. Andthen for each transaction, this algorithm replaces the sensitive data with unknown data values. In thishide the actual value, theyreplace '1' by '0' or '0' by '1' or with any unknown values in a specific transaction [8]. The replacement of these values does not depend on any specific rule. The objective of this technique is to preserve the sensitive data from unauthorized access. The algorithm replaces unknown values in the place of attribute for every transaction which supports that sensitive rule. These steps will continue till all the sensitive attributes are hidden by the unknown values.

IMPLEMENTATION

This phase describes the implementation of proposed work, how the work has been done and various problems occur during the implementation phase. In this proposed study, we build a cube and secure against the unauthorized access to sensitive information. The diagram defines the complete overview of the study which is going to be accomplished.

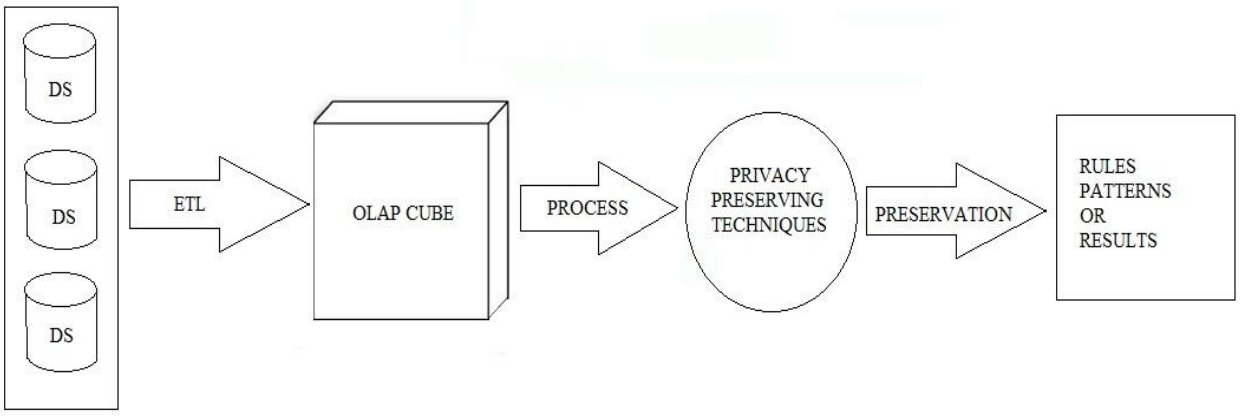


Fig. 2: -Context diagram of privacy preserving data mining

In this framework for privacy preserving Data Mining is shown. Data from different data are collected and are preprocessed using ETL tools. This transformed and clean data from Level 1 is stored in the data cube. Data in data cube is used for mining. In level 2 privacy preservation

techniques are used to protect data from unauthorized access. Sensitive data of an individual can be prevented from being misused. Basically, this paper is focused on applying various privacy preserving mechanisms to protect sensitive data from the unauthorized access control because a single data item in a data cube is not likely to be accessed alone, but a number of data are often aggregated to give summarized information and the trends of the database [10].

APPLYING ANONYMIZATION TECHNIQUE

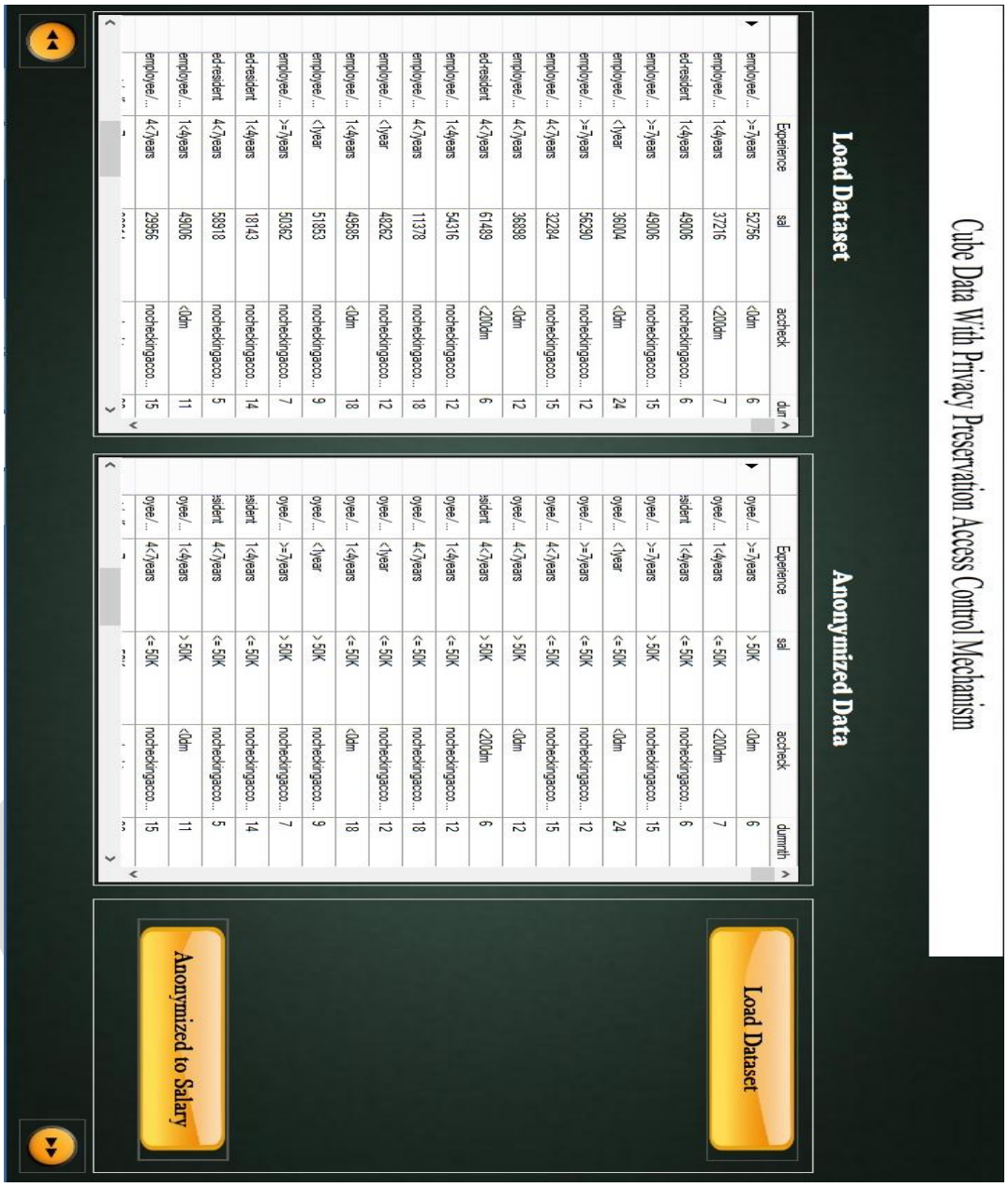
Anonymization technique use suppression and generalization of records to satisfy privacy requirements with minimal distortion of micro data. The privacy is achieved at the cost of accuracy and imprecision is introduced into the authorized information under an access control policy. The concept of imprecision is bound for permission to define a threshold on the amount of imprecision that can be tolerated. The anonymity technique is used by an access control mechanism to ensure both security and privacy of the sensitive information. Data access control is based on the notion of privileges- the authorization to perform a particular operation. Privileges are required to gain access to information in the data warehouse. However, it is well known that data access control is insufficient in controlling information leakage because information is not released directly by manipulating legitimate queries about aggregated information.

TESTING AND RESULTS

Screenshot 1: - Obtaining Sensitive information from dataset.

In this point of practical implementation, identified the sensitive information from the whole dataset of employees of an organisation.

Here for the privacy preserving point of view, anonymization is applied on the basis of the range value of the threshold to the perspective view of salary details of the employees of the organization.



Screenshot 2: - Applying anonymization based on threshold range.

Screenshot 3: - Multiview perspective view of sensitive information.

Finally, multi-perspective view of sensitive information by anonymizing data from different levels as min-salary, max-salary and use the min and max section from the perspective view of residential address.

CONCLUSION

This present study provides a better solution to protecting sensitive data in OLAP data cubes from unauthorized accesses. In this study a privacy-preserving access control framework for cube data has been proposed. This framework is a combination of privacy protection mechanisms and access control. The access control mechanism allows only authorized query predicates on sensitive data. The privacy preserving module anonymizes the data to meet privacy requirements and imprecision constraints on predicates set by the access control mechanism.The result is secure that both privacy preserving and unauthorized accesses are eliminated.

REFERENCES

[1] Sairam et al "Performance Analysis of Clustering Algorithms in Detecting outliers",International Journal of Computer Science and Information Technologies, Vol. 2 (1) , Jan-Feb 2011, 486- 488.

[2] P.Deivanai, J. JesuVedhaNayahi and V.Kavitha," A Hybrid Data Anonymization integrated with Suppression for Preserving Privacy in mining multi party data" in *proceedings ofInternational Conference on Recent Trends in InformationTechnology*, IEEE 2011.

[3] J. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data", in The Eighth ACM SIGKDD International conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, CA, July 2002, IEEE 2002.

[4] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy", in proceedings of Int'l Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 2002.

[5] J. Liu, J. Luo and J. Z. Huang, "Rating: Privacy Preservation for Multiple Attributes with Different Sensitivity requirements", in *proceedings of 11th IEEE International Conference on DataMining Workshops,* IEEE 2011.

[6] H. Kargupta and S. Datta, Q. Wang and K. Sivakumar, "On the Privacy Preserving Properties of Random Data Perturbation Techniques", in *proceedings of the Third IEEE InternationalConference on Data Mining,* IEEE 2003.

[7] S. Lohiya and L. Ragha, "Privacy Preserving in Data Mining Using Hybrid Approach", in *proceedings of 2012 FourthInternational Conference on Computational Intelligence andCommunication Networks,* IEEE 2012.

[8] S. Lohiya and L. Ragha, "Privacy Preserving in Data Mining Using Hybrid Approach", in *proceedings of 2012 FourthInternational Conference on Computational Intelligence and Communication Networks,* IEEE 2012.

[9] A. Parmar, U. P. Rao, D. R. Patel, "Blocking based approach for classification Rule hiding to Preserve the Privacy in Database" , in *proceedings of International Symposium on Computer Scienceand Society,* IEEE 2011.

[10] D. Kim, E. Lee, M. Kim, Y. Lee, An efficient processing of range-min/max queries over dataCube source, Information Sciences 112 (1998) 223–237.

[11] T. Priebe, G. Pernul, Towards OLAP security design—survey and research issues, in:Proceedings of the third ACM International Workshop on Data Warehousing and OLAP,2000.

[12] Sung, S. Y., Liu, Y., Xiong, H., & Ng, P. A. (2005). Privacy preservation for data cubes. *Knowledge and Information Systems*, *9*(1), 38-61.