International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 13 Issue 3 –MARCH 2015.

PRIVACY PRESERVING ACCESS CONTROL MECHANISM USING ACCURACY CONSTRAINED FRAMEWORK FOR RELATIONAL DATA

*S.PRASANTH KUMAR, *A.V. ANANDA MOUROUGANE, *G. LAAZER, #P. SATHIYANARAYANAN.

**Ug scholar, cse, manakula vinayagar institute of technology*

#Assistant professor, cse, manakula vinayagar institute of technology

***ABSTRACT*--Various grouping organizations which aggregate and assign in-collect important special data for an assortment of clashing need which in addition to numerical and common aspect research. In these position the data dealer is usually allow with difficulty. On the one side it is more wanted to preserve the anonymity and important message of existence. On another side it is also adequacy of the data for research. In this dissertation a considerable of study of this problem. We covered the initial on the approach of anonymity that are explained with good manner to original identification or with good value to the delicate quality. A considerable valuation that shows to pass the data to the available of high quality data that respects different meaningful concept of privacy that is it is possible to do this efficiently for large data sets. Cloud computing is changing the way that organizations supervise the data because of physique and minimum cost and pervasive process. Privacy preserving has commenced as an essential entanglement with the presence of active cloud computing. In this concept we describes the numerous characteristic anonymization privacy preserving method recycled in the cloud computing**

I. INTRODUCTION

Personal information is collected, stored, analyzed, and distributed in the course of everyday life. In the medical domain, the US Department of Health and Human Services has announced a major initiative toward digitizing the patient records maintained by hospitals, pharmacies, etc. [2]. In the United States, three independent credit reporting agencies maintain databases of personal finance information that are widely used in credit evaluation [3,4] Enterprises supported their business by procuring information technology infrastructure and developing their software on top of that infrastructure. Supermarkets and other retailers maintain and analyze large databases of customer purchase information, collected by way of various affinity and discount programs. For example, when a customer makes a purchase using its “Club Card,” the Safeway supermarket chain records data about the transaction, including “the amount and content of your purchases and the time and place these purchases are made” [5]. On the surface, this appears harmless, yet there is the potential for abuse. For example, in a Los Angeles court case, Robert Rivera sued Vons grocery store (owned by Safeway) after a slip-and-fall incident. During negotiations, Mr. Rivera’s attorney claimed that Vons had accessed his client’s shopping records, and planned to introduce at trial information regarding Rivera’s frequent purchases of alcohol, implying that he was drunk at the time of the accident .

II. APPROACH

where Explicit Identifier is a set of attributes, such as name and social security number (SSN), containing information that explicitly identifies record owners; Quasi Identifier (QID) is a set of attributes that could potentially identify record owners; Sensitive Attributes consists of sensitive person-specific information such as disease, salary, and disability status; and Non-Sensitive Attributes contains all attributes that do not fall into the previous three categories [Burnett et al. 2003]. The four sets of attributes are disjoint. Most works assume that each record in the table represents a distinct record owner.

III. ABOUT CLOUD COMPUTING

Cloud computing is a computing term or metaphor that evolved in the late 2000s, based on utility and consumption of computing resources. Cloud computing involves deploying groups of remote servers and software networks that allow centralized data storage and online access to computer services or resources. Clouds can be classified as public, private or hybrid.

IV. INCOGNITO ALGORITHM

We noticed a number of convincing parallels between Samarati and Sweeney's generalization framework and ideas used in managing multi-dimensional data [21, 44] and mining association rules. By bringing these techniques to bear on the anonymization problem, we developed a core algorithm (as well as several variations) that is often substantially more efficient than previous algorithms.

For example, because the Patients data in Table 2.1 is 2-anonymous with respect to hS0i, then it must also be 2-anonymous with respect to hS1i, a generalization of S0. The second property is reminiscent of operations along dimension hierarchies in OLAP processing. We also noticed a close connection with the a priori observation, a dynamic programming approach that formed the basis for a number of algorithms for mining frequent item sets. This observation is easily applied to full-domain anonymization by way of the subset property.

POSSIBLE SOLUTIONS FOR THE PRIVACY PROBLEMS BASED ON DIFFERENT METHODS

Cloud Computing is presently one of the hottest topics in information technology. Since the outsourcing of all the essential data is available with a third party, there is always having a concern of cloud service providers trust worthiness. Due to data privacy, it is essential for users to encrypt their sensitive data before storing them into the cloud. Yet, there exist some shortcomings in the situation of traditional encryption. When a secret key owner wants to look for some data that are stored in the cloud storage, he may be needed to download all encrypted data from the cloud server, and then decrypts and searches them. If the encrypted data are huge or the client is a mobile user, then it will be very inefficient and is not convenient. Otherwise he must send his key to the cloud server which performs the decryption and search procedures. It causes a serious trouble that the cloud server obtains the secret key So many models were existed to ensure the integrity of data file.

A) Statistical database

A related and long-standing field of research considers the problem of disclosure control for Statistical databases. In this case, a central organization maintains a database of (potentially sensitive) information, over which it answers aggregate queries. (These queries typically consist of an aggregate function, such as SUM or MAX, and perhaps a selection predicate.) It is well known that, given a sufficient number of queries, an adversary can reconstruct the contents of the underlying database. There have been a variety of approaches proposed for addressing this problem, and in this section, we briefly describe two: auditing and output perturbation.

B) Query-view security

In another related line of research, Miklau and Suciu began a series of papers on the problem query-view security [64]. Given a public view V of a (relational) database, the goal is to determine whether it reveals any information about a private query Q of the same database. In this work, views and queries are defined by conjunctive queries. No assumptions are made about the computational capacity of the adversary. The standard of privacy initially proposed by Miklau and Suciu ("perfect privacy") is quite strict, and requires that the public views provide no additional information with respect to the private query. That is, the posterior probability of a particular answer to the private query should not be altered by an adversary seeing the public views. (This formalism is closely-related to Shannon's definition of perfect secrecy). The formulation of perfect privacy makes no assumptions about the computational capabilities of the adversary. In this initial work, Miklau and Suciu show that the problem of checking perfect privacy in general, for conjunctive queries.

C) Database authorization, access control & security

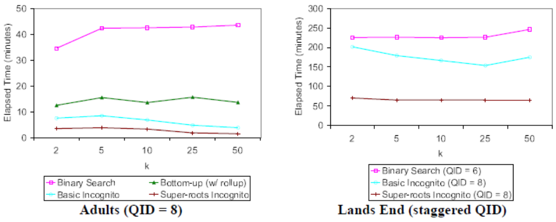
While not the primary focus of this thesis, authorization and access control for data management systems are also important and active research areas. In this section, we briefly describe three distinct settings: access control models for traditional database servers, security and confidentiality for outsourced databases, and authorization for published data. Traditionally, database authorization has focused on the setting where data is stored on a trusted server, and the system must control the outflow of information according to some access control policy. This is the setting addressed by the discretionary access control mechanism of SQL,as well as role-based and mandatory access control systems. More recently, several fine-grained data-centric access control mechanisms have been developed.

V. FUTURE RESEARCH DIRECTIONS

Information sharing has become part of the routine activity of many individuals, companies, organizations, and government agencies. Privacy-preserving data publishing is a promising approach to information sharing, while preserving individual privacy and protecting sensitive information. In this survey, we reviewed the recent developments in the field. The general objective is to transform the original data into some anonymous form to prevent from inferring its record owners' sensitive information. We presented our views on the difference between privacy-preserving data publishing and privacy-preserving data mining, and gave a list of desirable properties of a privacy-preserving data publishing method. We reviewed and compared existing methods in terms of privacy models, anonymization operations, information metrics, and
anonymization algorithms.

VI. EXPERIMENTAL RESULTS

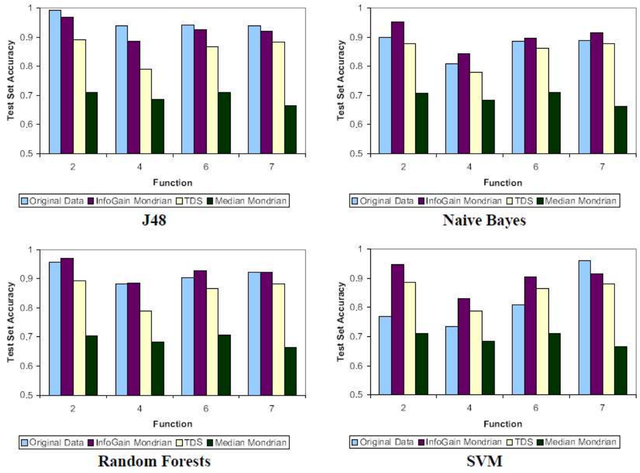
The worst-case time complexity of each of the algorithms considered in this chapter, including Incognito, is exponential in the size of the quasi-identifier. However, we found that in practice the rollup and a priori optimizations go a long way in improving performance. Figure 2.6 shows the execution time of Incognito and previous algorithms on the experimental databases for varied quasi-identifier size (k = 2, 10). We began with the first three quasi-identifier attributes from each schema and added additional attributes in the order they appear in these lists. We found that Incognito substantially outperformed Binary Search on both databases, despite the fact that Incognito generates all k-anonymous full-domain generalizations, while Binary Search finds only one.



Incognito performance evaluations for varied K

Experimental evaluation of data quality

We conducted an experimental evaluation with two main goals. The first goal was to provide insight about experimental quality evaluation methodology. We outline an experimental protocol for evaluating an anonymization algorithm with respect to a workload of classification and regression tasks. A comparison with the results of simpler general-purpose quality measures indicates the importance of evaluating data quality with respect to the target workload when it is known. The second goal is to evaluate the extensions to Mondrian for incorporating workload. We pay particular attention to the impact of incorporating one or more target classification / regression models and the effects of multidimensional recoding. We also evaluate the effectiveness of our algorithms with respect to selections and projections.



Classification-based model evaluation using synthetic data

VII. CONCLUSION:

Cloud is achieved demand and improvement in now days. The most recent techniques that we surveyed in this concept. Cloud can attempt so may challenges for balanced scheduling with recent technologies and methodologies. This thesis has not, by and large, attempted to address the higher-level policy issues surrounding data privacy. The computer security community has long drawn a distinction between the ideas of policy and mechanism, and a similar distinction can be made here. Unfortunately, as in security, where it is not always clear what it means for a system to be secure, it is sometimes difficult to precisely define what it means to protect privacy. those reasonable expectations of privacy and policies designed in accordance with these expectations are dictated by the application at hand. For this reason, it seems unlikely that there will ever be a single catch-all framework for reasoning about all types of privacy and disclosure. the future research direction appears to lie in defining policies high-level statements of “privacy” that are appropriate philosophically, legally, and technically to specific application scenarios, and developing mechanisms that rigorously enforce these policies. Many technical mechanisms have been developed over the years, including authorization, encryption, aggregation, generalization, output perturbation, and more. In the future, we expect that these mechanisms will form the building blocks for enforcing emerging classes of policies.

REFERENCES

[1] Cloud Computing: Special theme, European research consortium for Informatics and mathematics (ERCIM), ISSN 0926-4981.

[2] Reuters. US pushes digital medical records, July 22 2004.

[3] Equifax. http://equifax.com.

[4] Experian. http://www.experian.com.

[5] Safeway Inc. Privacy policy, May 26 2007. http://www.safeway.com/privacy page.asp.

[6] L. Sweeney. K-anonymity: A model for protecting privacy. International Journal on Uncertainty, Fuzziness, and Knowledge-based Systems, 10(5):557–570, 2002.

[7] U.S. Department of Health and Human Services Office for Civil Rights. HIPAA administrative simplification regulation text, February 16 2006.

[8] Enterprise Privacy Group. (2008). Privacy by Design: An Overview of privacy Enhancing Technologies. Retrieved from www.ico.gov.uk/upload/documents/pdb_report_html/pbd_pets_paper.pdf

[9] K. LeFevre, D.DeWitt, and R. Ramakrishnan. Incognito: Efficient full-domain k-anonymity. In Proceedings of the ACM SIGMOD International Conference on Management of Data, 2005.

[10] K. LeFevre and D. DeWitt. Scalable anonymization algorithms for large data sets. Universityof Wisconsin Computer Sciences Technical Report 1590, 2007.

[11] K. LeFevre, D. DeWitt, and R. Ramakrishnan. Mondrian multidimensional k-anonymity. In Proceedings of the 22nd International Conference on Data Engineering (ICDE), 2006.

[12] G. Aggarwal, M. Bawa, P. Ganesan, H. Garcia-Molina, K. Kenthapadi, R. Motwani, U. Srivastava, D. Thomas, and Y. Xu. Two can keep a secret: A distributed architecture for secure database services. In Proceedings of the 2nd Conference on Innovative Data Systems Research(CIDR), 2005.

[13] G. Aggarwal, T. Feder, K. Kenthapadi, R. Panigrahy, D. Thomas, and A. Zhu. Achievinganonymity via clustering in a metric space. In Proceedings of the 25th ACM SIGACTSIGMOD-SIGART Symposium on Principles of Database Systems (PODS), 2006.

[14] D. Bell and L. LaPadula. Secure computer systems: Unified exposition and multics interpretation. Technical Report ESD-TR-75-306, MITRE Corp., Bedford, Mass., 1976.

[15] A. Blum, C. Dwork, F. McSherry, and K. Nissim. Practical privacy: the SuLQ framework.In Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS), 2005.

[16] S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. SIGMOD Record, 26, 1997.

[17] B. Chen, L. Chen, Y. Lin, and R. Ramakrishnan. Prediction cubes. In Proceedings of the 31st International Conference on Very Large Databases (VLDB), 2005.

[18] A. Deutsch and Y. Papakonstantinou. Privacy in database publishing. In Proceedings of the 10th International Conference on Database Theory (ICDT), January 2005.

[19] J. Domingo-Ferrer and J.M. Mateo-Sanz. Practical data-oriented microaggregation for statistical disclosure control. IEEE Transactions on Knowledge and Data Engineering, 4(1), 2002.

[20] A. Evfimievski, J. Gehrke, and R. Srikant. Limiting privacy breaches in privacy-preserving data mining. In Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS), 2003.

[21] J. Gehrke, V. Ganti, R. Ramakrishnan, and W. Loh. BOAT: Optimistic decision tree construction. In Proceedings of the ACM SIGMOD International Conference on Management of Data, 1999.

[22] D. Kifer and J. Gehrke. Injecting utility into anonymized datasets. In Proceedings of the ACM SIGMOD International Conference on Management of Data, 2006.

[23] J. Kleinberg, C. Papadimitriou, and P. Raghavan. Auditing boolean attributes. In Proceedings of the 19th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS), 2000.