Volume 5, Issue 6, June 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Review on Accuracy Constrained Privacy-Preserving Access Control Mechanism for Relational Data

Kaushlendra Tiwari, Awadhesh Kumar Sharma
Department of Computer Science & Engineering,
MMMUT, Gorakhpur, Uttar, India

Abstract*- There are some situations when sensitive information is shared to authorized person for some authentication, verification or authorization purpose. A set of control mechanism protects sensitive information from unauthorized users. When sensitive information is shared, the authorized user can still compromise the privacy of a person such as identity disclosure. Privacy Protection Mechanism (PPM) can satisfy privacy requirements such as k-anonymity and l-diversity with its supression and generalization of relational data. While satisfying the privacy requirement, k-anonymity or l-diversity, the access control policies define selection predicates available to rolls. An additional constraint that needs to be satisfied by the PPM is the imprecision bound for each selection predicate. However, privacy is achieved at the cost of precision of authorized information. Thus access control mechanism protects the sensitive information from unauthorized user , but Privacy Protection Mechanism (PEM) protects the privacy of users from both unauthorized and authorized user. The literature survey might provide techniques for workload-aware anonymization for selection predicates, as the problem of satisfying the accuracy constraints for multiple roles has not been studied before. The purpose of the present project is to propose heuristics for anonymization algorithms and to show the viability of the proposed approach for empirically satisfying the imprecision bounds for more permission.

Key words: Access control, Privacy, k-anonymity, Precision, Imprecision, l-diversity.

I. INTRODUCTION

As organizations increase their reliance on, possibly distributed, information systems for daily business, they become more vulnerable to security breaches even as they gain productivity and efficiency advantages. Though a number of techniques, such as encryption and electronic signatures, are currently available to protect data when transmitted across sites, a truly comprehensive approach for data protection must also include mechanisms for enforcing access control policies based on data contents, subject qualifications and characteristics, and other relevant contextual information, such time. It is well understood today that the semantics of data must be taken into account in order to specify effective access control policies [11]. Also, techniques for data integrity and availability specifically tailored to database systems must be adopted. In this respect, over the years the database security community has developed a number of different techniques and approaches to assure data confidentiality, integrity, and availability. However, despite such advances, the database security area faces several new challenges [3]. Factors such as the evolution of security concerns, the "disintermediation" of access to data, new computing paradigms and applications, such as grid-based computing and on demand business, have introduced both new security requirements and new contexts in which to apply and possibly extend current approaches [3].

Privacy-preservation concept for sensitive data can require the enforcement of privacy policies or the protection against identity disclosure by satisfying some privacy constraints [11]. The sensitive information, even after the removal of identifying attributes, is still susceptible to linking attacks by the authorized users [2]. This problem has been studied extensively in the area of micro data publishing [13] and privacy definitions, e.g., k-anonymity [2], l- diversity [8] [2], and variance diversity [4]. The anonymity techniques can be used with an access control mechanism to ensure both security and privacy of the sensitive information. The privacy is achieved at the cost of accuracy and imprecision is introduced in the authorized information under an access control policy [1].

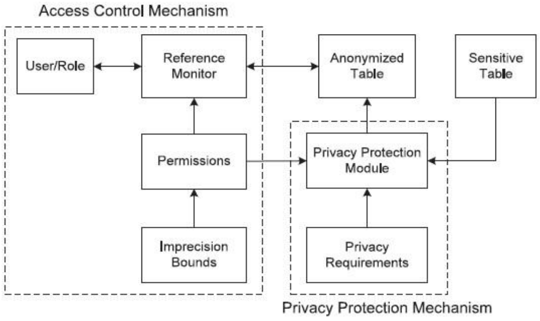

Fig. 1 Accuracy-constrained privacy-preserving access control mechanism

An integrated framework of achieving both privacy and security is proposed though the integration of Access Control Mechanism with Privacy Preservation Technique to prevent the authorized user from misusing the sensitive information [1].

The enforcement of privacy policies or the protection against identity disclosure satisfying some privacy requirements are the pre-requisites for privacy-preservation of sensitive data. Even after removal of identifying attributes, the sensitive information is susceptible to liking attacks by the authorized users. So the present investigation is proposed to study the area of micro data publishing [13] and privacy definitions such as k-anonymity [2], l-diversity [8] and variance diversity [4].

II. LITERATURE REVIEW

The collection of digital information by governments, corporations, and individuals has created tremendous opportunities for knowledge and information-based decision making. Driven by mutual benefits, or by regulations that require certain data to be published, there is a demand for the exchange and publication of data among various parties. Data in its original form, however, typically contains sensitive information about individuals, and publishing such data will violate individual privacy. The current practice in data publishing relies mainly on policies and guidelines as to what types of data can be published and on agreements on the use of published data. This approach alone may lead to excessive data distortion or insufficient protection. Privacy-preserving data publishing (PPDP) provides methods and tools for publishing useful information while preserving data privacy. Recently, PPDP has received considerable attention in research communities, and many approaches have been proposed for different data publishing scenarios. In this survey, we will systematically summarize and evaluate different approaches to PPDP, study the challenges in practical data publishing, clarify the differences and requirements that distinguish PPDP from other related problems, and propose future research directions [13] [11].

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 reviewed and compared existing methods in terms of privacy models, anonymization operations, information metrics, and anonymization algorithms. Most of these approaches assumed a single release from a single publisher, and thus only protected the data up to the first release or the first recipient. We also reviewed several works on more challenging publishing scenarios, privacy protection and privacy preserving mechanisms [3].

Privacy protection is a complex social issue, which involves policy-making, technology, psychology, and politics. Privacy protection research in computer science can provide only technical solutions to the problem. Successful application of privacy preserving technology will rely on the cooperation of policy makers in governments and decision makers in companies and organizations. Unfortunately, while the deployment of privacy-threatening technology, such as social networks, grows quickly, the implementation of privacy-preserving technology in real-life applications is very limited. As the gap becomes larger, we foresee that the number of incidents and the scope of privacy breach will increase in the near future. Below, we discuss a few potential research directions in privacy preservation, together with some desirable properties that could facilitate the general public, decision makers, and systems engineers to adopt privacy-preserving technology. Most previous privacy-preserving techniques were proposed for data publishers, but individual record owners should also have the right and responsibility to protect their own private information. There is an urgent need for personalized privacy-preserving tools, such as privacy-preserving web browsers and minimal information disclosure protocols for e-commerce activities. It is important that the privacy-preserving notions and tools developed are intuitive for novice users. Xiao and Tao's work on "personalized privacy preservation" provides a good start, but little work has been conducted on this direction since. Privacy protection in Emerging technologies, like location based services [Atzori et al. 2007]; bioinformatics, and mashup web applications, enhance our quality of life. These new technologies allow corporations and individuals to have access to previously unavailable information and knowledge; however, such benefits also bring up many new privacy issues. Nowadays, once a new technology has been adopted by a small community, it can become very popular in a short period of time. A typical example is the social network application called Facebook. Since its deployment in 2004, it has acquired 70 million active users. Due to the massive number of users, the harm could be extensive if the new technology is misused. One research direction is to customize existing privacy-preserving models for emerging technologies. The issue of privacy protection is often considered after the deployment of a new technology. Typical examples are the deployments of mobile devices with location-based services [Abulet al.2008; Atzori et al. 2007; Hengartner 2007]; sensor networks, and social networks. The privacy issue should be considered as a primary requirement in the engineering process for developing new technology. This involves formal specification of privacy requirements and formal verification tools to prove the correctness of a privacy-preserving system. Finally, we emphasize that privacy-preserving technology solves only one side of the problem. It is equally important to identify and overcome the nontechnical difficulties faced by decision makers when they deploy a privacy-preserving technology. Their typical concerns include the degradation of data/service quality, loss of valuable information, increased costs, and increased complexity. We believe that cross-disciplinary research is the key to remove these obstacles, and urge computer scientists in the privacy protection field to conduct cross-disciplinary research with social scientists in sociology, psychology, and public policy studies. Having a better understanding of the privacy problem from different perspectives can help realize successful applications of privacy-preserving technology [3].

Various papers were referred for the present research regarding access control mechanism, privacy preserving, k-anonymity, l-diversity, and for workload aware anonymity concepts. In this, The Arbac99 Model for Administration of Roles was proposed by R. Sandhu et al. in 1999 [14]. P. Samarati raised concern about Protecting Respondents' Identities in Microdata Release in 2001 and discussed various aspects regarding this. Protecting individual privacy is an important problem in microdata distribution and publishing. Anonymization algorithms typically aim to satisfy certain privacy definitions with minimal impact on the quality of the resulting data [13]. In 2011, Proposed NIST Standard for Role-Based Access Control was discussed by D. Ferraiolo, R. Sandhu et al. [12]. Security Concepts,Approaches, and Challenges was also discussed by them [11][12]. The definition of differential privacy was used [9] whereby random noise is added to original query which results to satisfy privacy constraints. However, the accuracy constraints for permissions were not considered. But the present study defines the privacy requirement in terms of k-anonymity. The concept of L-Diversity: Privacy Beyond k-anonymity was discussed by A. Machanavajjhala in 2007. Publishing data about individuals without revealing sensitive information about them is an important problem. In recent years, a new definition of privacy called k-anonymity has gained popularity. In a k-anonymized dataset, each record is indistinguishable from at least k−1 other records with respect to certain "identifying" attributes[8]. We have observed that since building an index over a data set leads to a natural partitioning of the data set, k-anonymity can be introduced by enforcing a minimum occupancy threshold on partitions. Moreover, we can take advantage of these indexes as a means for producing dynamic k-anonymous data sets. Experiments indicate that the spatial indexing approach to k-anonymization is scalable for very large data sets, is faster than previously proposed algorithms. Experiments with random queries over these anonymized data sets also show that the accuracy of the queries on the R-tree anonymized data have higher accuracy than the same queries on data anonymized by a previously proposed algorithm. The philosophy behind k-anonymization [2] [10] is that preventing any attacker from isolating any individual from k − 1 others is sufficient for privacy, and the compaction procedure maintains that philosophy. Hence one can interpret our results on compaction as demonstrating that given current definitions of k-anonymity, compaction is important to producing high quality anonymization. However, if one is convinced that compaction reveals too much information, then our results indicate that the definitions of anonymity need to be augmented to prevent disclosures that result from "overcompaction," because compaction respects whatever definition of anonymity the R-tree building procedure is given as input. Deciding whether or not compaction is desirable is an interesting challenge area for future work.

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 and the advantages of proposed system are formulate the accuracy and privacy constraints. Concept of accuracy-constrained privacy-preserving access control for relational data was studied and the solution of the k-PIB problem was approximated and empirical evaluation was conducted.[8].

T. Iwuchukwu and J. Naughton discussed about K-Anonymization as Spatial Indexing. It was a study toward scalable and incremental anonymization [7]. Further , in 2008, Workload-Aware Anonymization Techniques for Large-Scale Datasets was discussed by K. LeFevre, D. DeWitt, and R. Ramakrishna. Most of the previous literature has measured quality through simple one-size-fits-all measures, K. LeFevre et al. argue that quality is best judged with respect to the workload for Further, the quality of the data with respect to a particular workload is not necessarily correlated with simple general-purpose measures that have been proposed in the previous literature. [6]. A Framework for Efficient Data Anonymization Under Privacy and Accuracy Constraints was provided by G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis in the year of 2009 . The problem of accuracy-constrained anonymization for a given bound of acceptable information loss for each equivalence class was proposed [5].

Further research on Database Access Control & Privacy was done by S. Chaudhuri, R. Kaushik, and R. Ramamurthy in 2011 named as "Database Access Control & Privacy: Is There a Common Ground?" [3]. In this work , Access control with privacy mechanisms provided, with the sketch of an architecture for a hybrid system that enhances the authorization policy with the abstraction of noisy views that encapsulate previously proposed privacy mechanisms. Accessing data through a set of views is natural for users of database systems and thus the noisy views abstraction represents a natural progression of the concept of authorization views. It was also stated how noisy views based on differentially private algorithms could be implemented. A key advantage of the proposed hybrid system is its flexibility. It can support queries that refer to both the base tables and the differentially private views thus resulting in a system that is more powerful than using access control techniques or differential privacy techniques in isolation. While combining authorizations and differentially private views in this manner seems ad-hoc, it is shown to be a principled way to integrate differential privacy primitives with privacy guarantees [3]. Further research on Private Data Anonymization: Or, k-Anonymity Meets Differential was discussed by N. Li et al. , they defined the privacy requirement in terms of k-anonymity that after sampling, k-anonymity offers similar privacy guarantees as those of differential privacy [2]. The proposed accuracy-constrained privacy preserving access control framework allows the access control administrator to specify imprecision constraints that the privacy protection mechanism is required to meet along with the privacy requirements. The proposed accuracy-constrained privacy preserving access control framework allows the access control administrator to specify imprecision constraints that the privacy protection mechanism is required to meet along with the privacy requirements [1].

III. CONCLUSION

Latest research presents an accuracy-constrained privacy-preserving access control framework for relational data. The planned additive approach of access management and privacy protection mechanisms in our system provides a lot of security and information is retrieved during a custom-made approach which will build users to access during as lot of versatile approach. Any access management mechanism concentrates on anomaly users to avoid privacy breach .The Access Control Mechanism (ACM) allows solely licensed user predicates on sensitive information and Privacy Protection Mechanism (PPM) anonymizes the information to satisfy privacy necessities and inexactness constraints on predicates set by the access management mechanism. The present framework for access and privacy protection is a combination of both, access control and privacy protection mechanisms. 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. This interaction is formulated as the problem of k-anonymous Partitioning with Imprecision Bounds (k-PIB). Hardness results are given for the k-PIB problem and the heuristics for partitioning the data are presented to satisfy the privacy constraints and the imprecision bounds. In the current work, static access control and relational data model has been assumed. The proposed privacy-preserving access is extended to control incremental data and cell level access control.

REFERENCES

[1] ZahidPervaiz, Walid G. Aref, ArifGhafoor, and NagabhushanaPrabhu "Accuracy-Constrained Privacy-Preserving Access Control Mechanism for Relational Data" IEEE Trans. On Knowledge and Data Engineering, Vol. 26, No. 4, April 2014.

[2] N. Li, W. Qardaji, and D. Su, "Provably Private Data Anonymization: Or, k-Anonymity Meets Differential Privacy," Arxiv preprint arXiv: 1101. 2604, 2011.

[3] "Database Access Control & Privacy: Is There a Common Ground?" Proc. Fifth Biennial Conf. Innovative Data Systems Research (CIDR), pp. 96-103, 2011.

[4] X. Xiao, G. Bender, M. Hay, and J. Gehrke, "Ireduct: Differential Privacy with Reduced Relative Errors," Proc. ACM SIGMOD Int'l Conf. Management of Data, 2011

[5] G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis, "A Framework for Efficient Data Anonymization Under Privacy and Accuracy Constraints," ACM Trans. Database Systems, vol. 34, no. 2, article 9, 2009.

[6] K. LeFevre, D. DeWitt, and R. Ramakrishna, "Workload-Aware Anonymization Techniques for Large-Scale Datasets," ACM Trans. Database Systems, vol. 33, no. 3, pp. 1-47, 2008.

[7] T. Iwuchukwu and J. Naughton, "K-Anonymization as Spatial Indexing: Toward Scalable and Incremental Anonymization," Proc. 33rd Int'l Conf. Very Large Data Bases, pp. 746-757, 2007.

[8] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-Diversity: Privacy Beyond k-anonymity," ACM Trans. Knowledge Discovery from Data, vol. 1, no. 1, article 3, 2007. .

[9] C. Dwork, "Differential Privacy," Proc. 33rd Int'l Colloquium Automata, Languages and Programming, pp. 1-12, 2006.

[10] K. LeFevre, D. DeWitt, and R. Ramakrishna, "Mondrian Multidimensional K-Anonymity," Proc. 22nd]Int'l Conf. Data Eng., pp. 25- 25, 2006.

[11] E. Bertino and R. Sandhu, "Database Security Concepts,Approaches, and Challenges," IEEE Trans. Dependable and Secure Computing, vol. 2, no. 1, pp. 2-19, Jan.-Mar. 2005.

[12] D. Ferraiolo, R. Sandhu, S. Gavrila, D. Kuhn, and R. Chandramouli, "Proposed NIST Standard for Role-Based Access Control," ACM Trans. Information and System Security, vol. 4, no. 3, pp. 224- 274, 2001.

[13] P. Samarati, "Protecting Respondents' Identities in Microdata Release," IEEE Trans. Knowledge and Data Eng., vol. 13, no. 6,pp. 1010-1027, Nov. 2001.

[14] R. Sandhu and Q. Munawer, "The Arbac99 Model for Administration of Roles," Proc. 15th Ann. Computer Security Applications Conf., pp. 229-238, 1999.