ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

Accuracy-Constrained Privacy-Preserving Access Control Mechanism for Health Care Systems

Ankitha.N.Kulkarni[1], Ashwini.S[2], Vidhyashree.S[3], Chandrashekar.K.T[4]

B.E, Dept. of Information Science & Engg, B.M.S Institute of Technology & Mgmt, Bangalore, Karnataka, India[1,2,3,4]

Abstract: Data privacy issues are increasingly becoming important for many applications. Protective individual privacy is a crucial downside. However, sensitive data will still be ill-used by approved users to compromise the privacy of shoppers. Traditionally, research in the database community in the area of data security can be broadly classified into access control research and data privacy research. Access Control Mechanisms (ACM) is used to ensure that only authorized information is available to users. Privacy Protection Mechanism (PPM) uses suppression and generalization of relational data to anonymize and satisfy privacy needs. Recent research studied the problem of publishing data in databases without revealing the sensitive information, moving to the privacy preserving paradigms of k-anonymity and L-diversity. K-anonymity protects against the identity of an individual's record. L-diversity, in addition to this, safeguards against the association of an individual with specific sensitive information. The aim of this paper is to provide better security and minimum level of precision to the obtained data, for that in this paper an accuracy constrained privacy preserving access control mechanism is implemented with additional constraint on each selection predicate called imprecision bounds. The accuracy constraints are satisfied for multiple roles. We propose heuristics for anonymization algorithms to show empirically that the proposed approach satisfies imprecision bounds for more permissions and has lower total imprecision than the current state of the art.

Keywords: K-Anonymity, L-Diversity, Suppression, Generalization, Privacy

I. INTRODUCTION

As organizations increase their adoption of database systems as the key data management technology for day-to-day operations and decision making, the security of data managed by these systems becomes crucial. Damage and misuse of data affect not only a single user or application, but may have disastrous consequences on the entire organization. The recent rapid proliferation of Web based applications and information systems have further increased the risk exposure of databases and, thus, data protection is today more crucial than ever. It is important to appreciate that data needs to be protected from external threats and also from insider threats.

Organizations, such as hospitals, need to release microdata (e.g., medical records) for research and other public benefit purposes. However, sensitive personal information (e.g., medical condition of a specific person) may be revealed in this process. Identifying attributes such as name or social security number are not disclosed to protect privacy. Quasi-identifiers are sets of attributes (e.g., _ZIP, Gender, DateOfBirth_) which can be joined with information obtained from diverse sources (e.g., public voting registration data) in order to reveal the identity of individual records. *k*-anonymity is commonly achieved either by generalization (e.g., show only the area code instead of the exact phone number) or suppression (i.e., hide some values of the quasi-identifier), both of which inevitably lead to information loss. Still, the data should remain as accurate as possible in order to be useful in practice. Hence a trade-off between privacy and information loss emerges.

Recently, the concept of *l*-diversity was introduced to address the limitations of *k*-anonymity. The latter may disclose sensitive information when there are many identical Sensitive Attribute (*SA*)values within an equivalence class1 (e.g., all persons suffer from the same disease).L-diversity prevents uniformity and background knowledge attacks by ensuring that at least *SA* values are *well represented* in each equivalence class.

II. EXISTING SYSTEM

The concept of privacy-preservation for sensitive data can require the enforcement of privacy policies or the protection against identity disclosure by satisfying some privacy requirements. Investigate privacy-preservation from the anonymity aspect. The sensitive information, even after the removal of identifying attributes, is still susceptible to linking attacks by the authorized users.

The disadvantage of the existing system is the fact that it can minimize the imprecision aggregate for all queries, the imprecision added to each permission/query in the anonymized micro data is not known and it does not satisfy the accuracy constraints for individual permissions in a policy/workload.

III. PROPOSED SYSTEM

A number of extensions to the basic model have been proposed with the goal of enriching the expressive power of the authorization languages in order to address a large variety of application requirements. A first extension deals

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

with negative authorizations. The System R authorization model, as the models of most DBMSs, uses the closed world policy.

To address this threat, proposed the *k*-anonymity model: For every record in a released table there should be at the least $k - 1$ other records identical to it along a set of quasi-identifying attributes.

Records with identical quasi-identifier values constitute an *equivalence class*. *k*-anonymity is commonly achieved either by generalization (e.g., show only the area code instead of the exact phone number) or suppression(i.e., hide some values of the quasi-identifier), both of which inevitably lead to information loss. Still, the data should remain as accurate as possible in order to be useful in practice. Hence a trade-off between privacy and information loss emerges.

Recently, the concept of _-diversity was introduced to address the limitations of *k*-anonymity. The latter may disclose sensitive information when there are many identical Sensitive Attribute (*SA*) values within an equivalence class1 (e.g., all persons suffer from the same disease)._-diversity prevents uniformity and background knowledge attacks by ensuring that at least _ *SA* values are *well represented* in each equivalence class (e.g., the probability to associate a tuple with an *SA* value is bounded by 1/*l*).
In this paper we suggest that any *k*-anonymization algorithm can be adapted to achieve _-diversity. However, the following example demonstrates that such an approach may yield excessive information loss. So far, research efforts focused on the privacy-constrained anonymization problem, which minimizes information loss for a given value of *k* or _; we call this the *direct* anonymization problem. However, the resulting information loss may be high, rendering the published data useless for specific applications. In practice, the data recipient may require certain bounds on the amount of information loss. For instance, it is well known that the occurrence of certain diseases is highly correlated to age (e.g., Alzheimer's can only occur in elderly patients).

To ensure that anonymized hospital records make practical sense, a medical researcher may require that no anonymized group should span a range on attribute *Age* larger than 10 years. Motivated by such scenarios, we introduce the accuracy-constrained or dual anonymization problem. Let *E* be the maximum acceptable amount of information loss. The accuracy-constrained anonymization problem finds the maximum degree of privacy (i.e., *k* or _) that can be achieved such that information loss does not exceed *E*. Subsequently, the data publisher can assess whether the attainable privacy under this constraint is satisfactory, and can decide whether it makes sense to publish the data at all. To the best of our knowledge, the dual problem has not been addressed previously, despite its important practical applications.

IV. SYSTEM ARCHITECTURE

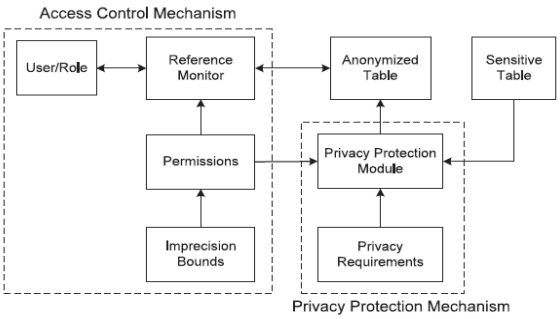

Fig 1: System architecture

The privacy protection mechanism ensures that the privacy and accuracy goals are met before the sensitive data is available to the access control mechanism. The permissions in the access control policy are based on selection predicates on the QI attributes. The policy administrator defines the permissions along with the imprecision bound for each permission/query, user-to-role assignments, and role-to-permission assignments. The specification of the imprecision bound ensures that the authorized data has the desired level of accuracy. The imprecision bound information is not shared with the users because knowing the imprecision bound can result in violating the privacy requirement. The privacy protection mechanism is required to meet the privacy requirement along with the imprecision bound for each required permission.

V. SEQUENCE DIAGRAM

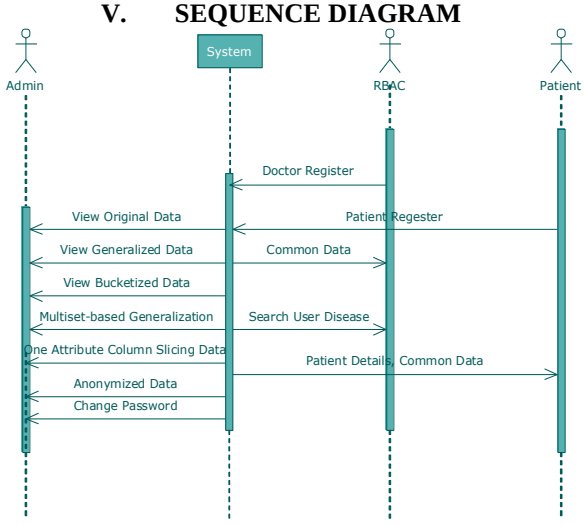

Fig 2: Sequence Diagram

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. The above figure suggests the sequence diagram for health care systems. A patient will first register in the system and enter the required details. Within an organization, roles are created for various job functions. The permissions to perform certain operations are assigned to specific roles for example doctors; nurses etc come under RBAC or Role Based Access Control who also register in the system. The admin will have the authority to anonymize data view

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

bucketized data and other operations. If the patient wants to view the common data then permissions are to be taken from the administrator.

VI. CONCLUSION

An accuracy-constrained privacy-preserving access control framework for relational data has been proposed. The framework is a combination of 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. We formulate this interaction as the problem of k-anonymous Partitioning with Imprecision Bounds (k-PIB). We give hardness results for the k-PIB problem and present heuristics for partitioning the data to the satisfy the privacy constraints and the imprecision bounds. In the current work, static access control and relational data model has been assumed. For future work, we plan to extend the proposed privacy-preserving access control to incremental data and cell level access control.

ACKNOWLEDGEMENT

I would like to thank my co-authors for providing immense support and help in preparing this document.

REFERENCES

[1] 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.
[2] B. Fung, K. Wang, R. Chen, and P. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Developments," ACM Computing Surveys, vol. 42, no. 4, article 14, 2010.
[3] A. Frank and A. Asuncion, "UCI Machine Learning Repository," 2010.
[4] B. Steven, A. Trent, G. Katie, G. Ronald, B.S. Matthew, andM.Sobek, "Integrated Public Use Microdata Series: Version 5.0 [Machine-Readable Database],," https://usa.ipums.org/usa/, 2010.
[5] R. Agrawal, P. Bird, T. Grandison, J. Kiernan, S. Logan, and W. Rjaibi, "Extending Relational Database Systems to Automatically Enforce Privacy Policies," Proc. 21st Int'l Conf. Data Eng., pp. 1013-1022, 2005.
[6] S. Chaudhuri, R. Kaushik, and R. Ramamurthy, "Database Access Control & Privacy: Is There a Common Ground?" Proc. Fifth Biennial Conf. Innovative Data Systems Research (CIDR), pp. 96-103, 2011.
[7] N. Li, W. Qardaji, and D. Su, "Provably Private Data Anonymization: Or, k-Anonymity Meets Differential Privacy," Arxiv preprint arXiv:1101.2604, 2011.
[8] 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.