UNIVERSITY OF SOUTHERN QUEENSLAND

ACCESS CONTROL MANAGEMENT AND PRIVACY-PRESERVING

A Dissertation submitted by

Md Enamul Kabir
B.Sc. Honours, M.Sc.

For the award of
Doctor of Philosophy

2010

Certification of dissertation

I hereby declare that the work presented in this dissertation is my own and is, to the best of my knowledge and belief, original except as acknowledged in the text. It has not previously been submitted either in whole or in part for a degree at this or any other university.

Signed: Date:

(Md Enamul Kabir)

Signed: Date:

(Hua Wang)

(Principal Supervisor)

Acknowledgements

First of all, I would like to thank almighty Allah, for His guidance and strength.

I would like to express my sincere gratitude and appreciation to my supervisor Associate Professor Hua Wang for his continuous inspiration, encouragement, patience, and individual feedback throughout the course of my Ph.D. study. I feel very grateful and blessed to have worked under his supervision. Special thanks to him for granting me the ARC scholarship to pursue my Ph.D. study. I would also like to express my sincere gratitude and appreciation to my co-supervisor Dr. Richard Watson for his advice and suggestions.

I would also like to thank the Center for Systems Biology (CSBi) for granting me half-tuition fee support. I would also like to thank the University of Dhaka, Bangladesh for the scholarship for study abroad.

I would also like to express my sincere gratitude and appreciation to the present head of the department of Maths and Computing (HOD) Dr. Stijn Dekeyser, former HOD Associate Professor Ron Addie and Dr. Richard Watson for their co-operation.

I extend my sincere and deep appreciation to my beloved wife Siuly and my son Shadman Srijon for their patience and continuous support. They have always been here with love and compassion to comfort me.

At last, but not the least, I wish to express my appreciation to my adorable parents, my brothers and my friends and relatives for their prayers, love and encouragement.

Abstract

In recent years, the phenomenal technological developments in information technology have led to an increase in the capability to store and record personal data about customers and individuals. This has led to concerns that the personal data may be misused for a variety of purposes. In order to alleviate these concerns, a number of techniques have been recently proposed in order to perform data mining tasks that are privacy-preserving. Thus the field of privacy has seen rapid advances in recent years and in the data mining environment have led to increased concerns about privacy. In this thesis, we develop efficient, effective and realistic methods in the privacy-preserving data mining field focusing on three core techniques, namely access control, data anonymization and statistical disclosure control.

In Part I, this thesis presents a model for privacy preserving access control which is based on a variety of purposes. Conditional purpose is applied along with allowed purpose and prohibited purpose in the model. It allows users to use some data for certain purposes with conditions. The structure of the conditional purpose-based access control model (CPBAC) is defined and investigated through a practical paradigm with access purpose and intended purpose. An algorithm is developed to achieve the compliance computation between access purposes and intended purposes. According to this model, more information from data providers can be extracted while at the same time assuring privacy that maximizes the usability of consumers' data. This model extends traditional access control models to a further coverage of privacy preservation in the data

mining environment. Its interior is a new structure for managing collected data in an effective and trustworthy way. This structure helps enterprises to circulate clear privacy promises and to collect and manage user preferences and consent. Finally, we inject this model with the conventional well known role-based access control (RBAC) model as RBAC is still the most popular approach towards access control to achieve database security and is available in many DBMS. The notion of applying these mechanisms to allow web sites to publish a privacy policy, and implement more nuanced management of usage information and other personal information, ultimately allows (legitimate) use of information.

In Part II, this thesis presents a systematic clustering based k-anonymization technique to minimize the information loss and at the same time assure data quality. The proposed technique adopts a system to group similar data together and then anonymize each group individually. The structure of systematic clustering problem is defined and investigated through paradigm and properties. An algorithm of the proposed problem is developed and it is shown that the time complexity is in $O(\frac{n^2}{k})$, where n is the total number of records containing individuals and their private information. Experimental results show that the proposed method attains a reasonable dominance with respect to both information loss and execution time. A way out is also shown to illustrate the usability of the algorithm for incremental datasets. Finally we extend the systematic-clustering approach to the l-diversity model that assumes that every group of indistinguishable records contains at least l distinct sensitive attribute values. The whole procedure consists of the two steps, namely a clustering step for k-anonymization and an l-diverse step.

In Part III, this thesis presents two heuristic algorithms for microdata protection in Statistical Disclosure Control (SDC). The first heuristic microaggregation algorithm works by partitioning the microdata into clusters of at least k records in a systematic way and then replacing the records in each cluster with the cen-

troid of the cluster which we refer to systematic microaggregation for SDC. The structure of the systematic microaggregation problem is defined and investigated and an algorithm of the proposed problem is developed. Experimental results show that the systematic microaggregation attains a reasonable dominance with respect to both information loss and execution time than the most popular heuristic algorithm called Maximum Distance to Average Vector (MDAV). Finally it has shown that the systematic microaggregation is highly scalable.

The second heuristic algorithm, called pairwise-systematic (P-S) microaggregation easily captures extreme values in the dataset and works by adopting simultaneously two distant groups at a time with the corresponding similar records together in a systematic way. Extensive experimental studies are conducted to show the efficiency and the effectiveness of the algorithm. The performance of the P-S algorithm is compared against the most recent microaggregation methods. Experimental results show that the P-S algorithm incurs significantly less information loss compared to the latest microaggregation methods for all of the test situations. Finally we propose a new microaggregation method where centroid is considered as median. The new method guarantees that the microaggregated data and the original data are similar by using a statistical test.

Publications Based on this Thesis

Accepted/Published Manuscripts

1. Kabir, M.E., Wang, H., and Bertino, E. (2010), "A Role-involved Purpose-based Access Control Model", Information Systems Frontiers (Revisions).

2. Kabir, M.E., Wang, H., and Bertino, E. (2010), "A Conditional Role-involved Purpose-based Access Control Model", Journal of Organizational Computing and Electronic Commerce (Revisions).

3. Kabir, M.E., Wang, H., and Bertino, E. (2010), "A Conditional Purpose-based Access Control Model with Dynamic Roles", Expert Systems with Applications (in Press).

4. Kabir, M.E., and Wang, H. (2010), "Microdata protection method through microaggragation: A median based approach", Information Security Journal: A Global perspective (in Press).

5. Kabir, M.E., Wang, H., and Yanchun, Z. (2010), "A Pairwise-Systematic Microaggregation for Statistical Disclosure Control", Proceedings of 10th International Conference on Data Mining (ICDM 2010), Sydney, Australia.

6. Kabir, M.E., Wang, H., and Bertino, E. (2010), "A Role-involved Conditional Purpose-based Access Control Model", Proceedings of IFIP EGES conference on E-Government and E-Services (EGES 2010) at the IFIP WCC world conference, Brisbane, Australia.

7. Kabir, M.E., and Wang, H. (2010), "Systematic Clustering-based Microaggregation for Statistical Disclosure Control ", Proceedings of International Conference on Data and Knowledge Engineering (ICDKE 2010), Melbourne, Australia.

8. Kabir, M.E., Wang, H., and Bertino, E. (2010), "Systematic Clustering Method for l-diversity Model", Proceedings of 21st Australasian Database Conference (ADC 2010), Brisbane, Australia.

9. Kabir, M.E., and Wang, H. (2009), "Conditional Purpose Based Access Control Model for Privacy Protection", Proceedings of 20th Australasian Database Conference (ADC 2009), Wellington, New Zealand.

Submitted Manuscripts

10. Kabir, M.E., Wang, H., and Bertino, E. (2009), "Efficient Systematic Clustering Method for k-anonymization", Acta Informatica. Submitted.

11. Kabir, M.E., and Wang, H. (2010), "Microaggregation for Statistical Disclosure Control: A Systematic Clustering-based approach", Applied Soft Computing. submitted.

Contents

List of Figures

List of Tables