Foundations for an Access Control Model for Privacy Preservation in Multi-relational Association Rule Mining

Stanley R. M. Oliveira[1,2]
oliveira@cs.ualberta.ca
[1]Embrapa Information Technology
Andre Tosello, 209, PO Box 6041
13083-886, Campinas, SP, Brasil

Osmar R. Zaïane[2]
zaiane@cs.ualberta.ca
[2]Database Systems Laboratory
Computing Science Department
University of Alberta, Canada

Workshop on Privacy, Security, and Data Mining
ICDM - Maebashi City, Japan December 9, 2002

Outline

- Motivation
- Basic Concepts
- Privacy preservation problem in MRAR
- Requirements for a MRAR Access Control
- An Access Model for MRAR
- Related Work
- Conclusions and Future Work

Motivation

- Access control models remain a fertile area for future research;
- Broad application of MRAR;
- Advantages of relational representation;
- The need for techniques that incorporate privacy and security concerns.
- While data access control models are popular for OS and DBMS, not much has been done for protection in the context of pattern discovery.

Access Controls

- A security policy specifies who is authorized to do what.
- A security mechanism allows us to enforce a chosen security policy.
- Three main mechanisms at the DBMS level:
  - Discretionary access control
  - Mandatory access control
  - Role-based access control

Discretionary Access Control

- Based on the concept of access rights or privileges for objects (tables and views), and mechanisms for giving users privileges (and revoking privileges).
- Creator of a table or a view automatically gets all privileges on it.
  - DMBS keeps track of who subsequently gains and loses privileges, and ensures that only requests from users who have the necessary privileges (at the time the request is issued) are allowed.

Basic Concepts

Discretionary Access Control (DAC):

- Access of users to objects is at the discretion of the owner of the data;
- Proper to environments in which information sharing is more important than protection of information;
- Advantage: flexibility – widely used in commercial environments;
- Drawback: vulnerable to malicious attacks (e.g. Trojan Horses).

Basic Concepts

Mandatory Access Control (MAC):

- Based on system-wide policies that cannot be changed by individual users.
  - Each DB object is assigned a security class.
  - Each subject (user or user program) is assigned a clearance for a security class.
  - Rules based on security classes and clearances govern who can read/write which objects.
- Suitable to environments in which users and objects can be classified;
- Access of users to objects is controlled by a central authority (security administrator);
- Advantage: designed to deal with information secrecy;
- Drawback: it's not always possible to assign clearances to users or to data.

Typical Security Classes

- Objects (e.g., tables, views, tuples)
- Subjects (e.g., users, user programs)
- Security classes:
  - Top secret (TS), secret (S), confidential (C), unclassified (U): TS > S> C > U
- Each object and subject is assigned a class.
  - Subject S can ***read*** object O only if

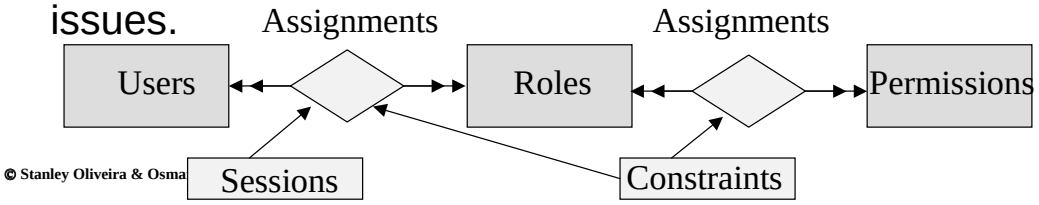
  class(S) >= class(O) (no reads in higher security)
  - Subject S can ***write*** object O only if

  class(S) <= class(O) (no writes in lower security)

Basic Concepts

Role-Based Access Control (RBAC):

- MAC and DAC are easily unified within the framework of RBAC;
- Based on the set of entities: users, roles, and permissions;
- Users are given roles and roles are assigned permissions. Objects have access permissions with regard to some roles.
- Advantage: designed to reduce complexity and cost of security administration;
- Drawback: It's not a panacea for all access control issues.



Statistical Database Security

- **Statistical databases** are used to produce statistics on various populations.
  - individual information is considered confidential.
  - users may allow to access statistical information on the population, (i. e., applying statistic functions to a population of tuples).
- Person(name, ssn, income, address, city, sex, last_ degree)
- Suppose we are allowed to retrieve only the statistical information over this relation by using **SUM, AVG, MIN, MAX, COUNT**,etc. (i.e. allow only aggregate queries. e.g., average age, rather than Joe's age).
- Statistical databases try to protect individual data by supporting only aggregate queries, but often, individual information can be inferred.

Privacy Preservation in MRAR

- Problem: If *D* is a relational database or even a data warehouse and *M* is the set of all association rules that could be mined from *D*, the goal is to provide users of different levels of access to *D* so that for each level *i*, the corresponding users are able to mine a set of association rules M_i, such that $M_i \subseteq M$.

- Goal $\Rightarrow$ classify *users* and *objects* into mining levels.

General Requirements for Access Control Models

1. General Requirements for Access Control Models

Requirement	DAC Models	MAC Models	RBAC Models
Type of Policy	Discretionary	Mandatory	Role-Based
Target System	Part OS & Part DB	OS & DB	DB
Type of Control	Access	Access & Flow	Access
Security Aspects	None	Secrecy & Integrity	Secrecy & Integrity

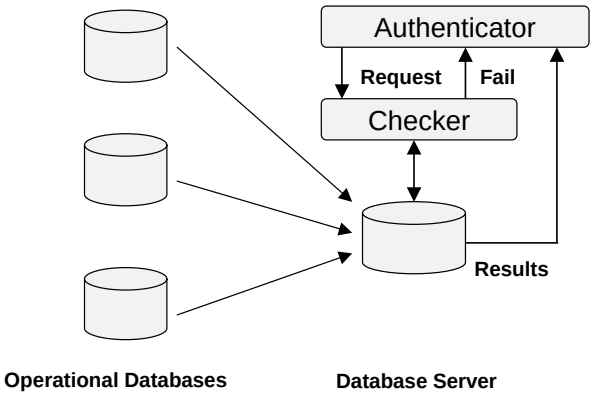
2. General Requirements for an Access Control Model for MRAR

Requirement	MRAR Model
Type of Policy	Mandatory
Target System	DB
Type of Control	Access and Flow
Security Aspects	Secrecy & Integrity

Additional Requirements for an Access Control for MRAR

- Req1: MRAR must be based on hierarchy of mining levels;
- Req2: Users associated with a mining level cannot pass rights;
- Req3: If a user is authorized to access one mining level that contains others, then the user is allowed to access the contained mining level(s);
- Req4: Users are granted rights only to access the data they need to perform their mining tasks;
- Req5: MRAR model might deal with multiple users concurrently;
- Req6: The capacity of a mining level cannot be exceeded;
- Req7: A user can never have an active mining level that is not authorized for that user;
- Req8: A user can perform an operation only if the operation is authorized for that mining level.

The MRAR Model: Structure



The MRAR Model: Definition

- The Top-MRAR is defined as follows:
  - U, O, P, and ML (users, objects, permissions, and mining level respectively).
  - permission: $O \times U \times ML \rightarrow$ {yes, no}, a function that answers if a user is given some permission for mining a particular object at a given mining level.

Multilevel Mining Relation

- Let $R(A_1{:}D_1, [ML_1], ..., A_n{:}Dn, [ML_n], T_{ML})$ be a multilevel relation schema, and for each A_i, $1 \leq i \leq n$, let D_i be the set of values associated with the domain named D_i, ML_i the mining level label for the attribute A_i, and T_{ML} the mining access level for the whole tuple. An instance of R that satisfies the domain in the schema is a set of tuples with n fields:

  $\{\rangle A_1{:}\ d_1, [ml_1], ...,\ A_n{:}\ d_n, [ml_n], t_{ML}\rangle \mid \forall i\ d_i \in D_i,\ ml_i \in ML_i;$ and $t_{ML} \in T_{ML}\}$.

The MRAR Model: Example

- Example: Three mining levels
  - *Full Mining* (FM): mining without restrictions;
  - *Specific Mining* (SM): mining affinity association rules;
  - *Restrictive Mining* (RM): mining a subset of SM.
  - Hierarchy: $Level_{FM} > Level_{SM} > Level_{RM}$
- Relational schema (Hiking Trip Store)
  - customers(cno, cname, rating, age, occupation, city)
  - items(ino, iname, price)
  - buys(cno, ino, date, qty, total)

The MRAR Model: Example

1. An example of multilevel relation

TID	CNO	ML_{CNO}	INO	ML_{INO}	DATE	QTY	TOTAL	ML_{TOTAL}	T_{ML}
100	C1	RM	I2	RM	01/05/2001	1	165.00	SM	RM
200	C1	RM	I4	RM	01/05/2001	2	60.00	SM	RM
300	C3	RM	I1	RM	01/06/2001	1	80.00	SM	RM
400	C3	RM	I3	SM	01/06/2001	1	120.00	SM	RM
500	C3	RM	I5	SM	01/06/2001	3	75.00	SM	SM
600	C4	RM	I3	SM	01/07/2001	1	120.00	SM	RM
700	C4	RM	I5	SM	01/07/2001	2	50.00	SM	SM

2. An example of multilevel relation for users in the Level RM

TID	CNO	ML_{CNO}	INO	ML_{INO}	DATE	QTY	T_{ML}
100	C1	RM	I2	RM	01/05/2001	1	RM
200	C1	RM	I4	RM	01/05/2001	2	RM
300	C3	RM	I1	RM	01/06/2001	1	RM
400	C3	RM	I3	RM	01/06/2001	1	RM
600	C4	RM	I3	RM	01/07/2001	1	RM

The MRAR Model: Properties

- Mandatory Property: access of users to objects is governed by security labels (mining levels) on the users and objects;
- Membership Property: a user is a member of only one mining level;
- Append Property: append of information is permitted without seeing it;
- Read Property: a query from a user at a given mining level can access information from the data whose label is dominated by that level;
- Mining Property: this property is completely related to *Read Property*;
- Non-Update Property: users are not allowed to alter data;
- Reclassification Property: in this case, a user at a given mining level must move to a upper level;
- Polyinstantiation Property: occurs when there are multiples instances of data at different mining access level.

Related Work

- **R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Hippocratic Databases.** In *the 28th International Conference on Very Large Data Bases*, Hong Kong, China, August 2002.
- Technology alone cannot address complex issues such as privacy;
- Hippocratic Databases: combine strength to enforce privacy:
  - Statistical databases: suppression, data swapping, etc;
  - Database security: access control, multilevel relations, etc;
  - Cryptography: collaborative work, search on encrypted data.
- Similarity between Hippocratic Databases and MRAR Model:
  - Users and objects are classified into security levels;
  - Attribute “purpose” in Hippocratic database is similar to “mining level” in MRAR Model

Related Work

- S. Jajodia, P. Samarati, M. L. Sapino, and V. S. Subrahmanian. Flexible Support for Multiple Access Control Policies. In ACM Transactions on Database Systems 26(2), 2001, pp.214-260.

Conclusions and Future Work

- Contributions:
  - Conceptual foundations and basic definitions;
  - Requirements for an access control for MRAR;
  - Design of the MRAR model considering the integration with existing technologies.
- Future Work
  - Studying new features that may be added to the model;
  - Extending the model to encompass other mining tasks (e.g. classification, clustering);
  - Analyzing the viability of integrating mining levels with roles without violating the information-flow access.

Questions?