Database Reverse Engineering and Tampering

,Ms.Aditi Dahake, Ms.Vrushali Balpande, Ms.Apurva Deshpande, Ms.AditiPande,Ms.Prajakta Bodakhe

Sant Gadge Baba Amravati University

pandeaditi88@gmail.com

Contact no :9860299026, 9763340392

Abstract- In today's modern world data is important entity. Data such as accounts, confidential documents, etc. is stored and managed by database in company. Database Reverse Engineering also works on extracting information and structure along with actual records stored in database table. Database Reverse Engineering means reconstructing ER model from existing schema, analysis and transformation. Various methods of Database Reverse Engineering are used in Digital forensic and tampering in data. Due to advance in technology growth of computer crime is increased. As the business data stored in these database is very important to investigate company dishonest act is necessary for that purpose fruitful database forensic technique should be used. Discrimination is mainly used to avoid tampering of data. It includes various strategies of discrimination.

Keywords: Database reverse engineering,Phases,Log Files,Forensic Investigation,Tampering Detection Apporach,

Discrimination, Discrimination strategy

I. Introduction

Database Reverse Engineering (DBRE) is a process of extracting database requirements from an implemented system. Traditionally, legacy systems suffer from poor design documentation and thus make the maintenance job more difficult. There are some problem in database reverse engineering like implicit structure, optimized structure, etc. The most frequent objectives of database reverse engineering are system maintenance, system extension , etc. The major goal of database reverse engineering is to reconstruct the conceptual data model of database system in the form of entity relationship diagram.

In recent database reverse engineering effort to derive a data model from table based database system. Different phases of database reverse engineering are data structure extraction, conceptualization, database structure modification. For the security of database forensic investigation is required. Secure storage of data is an everyday need for public businesses, government sectors and many institutions. For many organizations, if data is unauthorisely modified, whether by an outsider or by an internal intruder, it could cause severe problems for the company. And even for their clients as well. Therefore forensic investigation is necessary. Sometimes the original data is tampered that causes discrimination. Unfairly treating people on the basis of their belonging to a specific group, namely race, ideology, gender, etc., is known as discrimination. Mainly discrimination occurs due to favoritism.

II. Database Reverse Engineering

It is the process through which the logical and conceptual schemas of a legacy database, or of a set of files, are recovered from various information sources such as DDL code, data dictionary contents, database contents, or the source code of application. Database reverse engineering mainly works on schema extraction , analysis and transformation. Reverse engineering (RE) a piece of software consists, among others, in recovering or reconstructing its functional and technical specifications, starting mainly from the source text of the programs.

Need

The *Data* Description Language (DDL) is that part of the database management system facilities intended to declare or build the data structures of the database. The most frequent sources of problems have been identified.

Implicit structures: Such constructs have intentionally not been explicitly declared in the DDL specification of the database. They have generally been implemented in the same way as the discarded constructs mentioned above.

Optimized structures: For technical reasons, such as time and/or space optimization ,many database structures include non semantic constructs. In addition, redundant and unnormalized constructs are added to improve response time.

Awkward design: Not all databases were built by experienced designers. Novice and untrained developers, generally unaware of database theory and database methodology, often produce poor or even wrong structures.

Cross-model influence: The professional background of designers can lead to very peculiar results. For instance, some relational databases are actually straightforward translations of IMS databases, of COBOL files or of spreadsheets.

III. Database Reverse Engineering Process

Bottom-up Modeling

Build a database design based on either one of the following:

- By importing metadata directly from an existing database.
- By importing a DDL script that reflects an existing database implementation.

1.Reverse Engineer from a database or DDL script :
The resulting database is represented as a **Relational Schema** and definitions for Physical & Relational Schema objects.

2.Reverse Engineer from the Relational Schema to a higher-level schema :

The resulting schema is represented as an **ER Diagram** (or Class Diagram) and definitions for ER model objects.

IV. Motivation and Objectives

Reverse engineering is just one step in the information system life cycle. Indeed, painfully recovering the specifications of a database is not a sufficient motivation . It is generally intended to re document, convert, restructure, maintain or extend legacy applications. Here follow some of the most frequent objectives of database reverse engineering.

1.System maintenance :
Fixing bugs and modifying system functions require understanding the concerned component, including, in data-centered systems, the semantics and the implementation of the permanent data structures.

2.System extension :

This term designates changing and augmenting the functional goals of a system, such as adding new functions, or its external behaviour, such as improving its robustness.

3.System migration:

Migrating a system consists in replacing one or several of the implementation technologies. IMS/DB2, COBOL/C, monolithic/Client-server, centralized/distributed are some widespread examples of system migration.

4.Data Administration:

DBRE is also required when developing a data administration function that has to know and record the description of all the information resources of the organization.

V. Database Reverse Engineering Phases

Database reverse engineering is the process through which the logical conceptual schemas of a legacy database or of a set of files are recovered from various information sources such as data dictionary contents or source code of application. In a recent database reverse engineering effort to derive a data model from a table-based database system .In many cases , the major goal of data reverse engineering efforts is to reconstruct the conceptual data model of a database system in the form of an entity-relationship diagram. There are certain phases in database reverse engineering.

VI. Phases

There are following three phases for database reverse engineering:

- Data structure extraction
- Data structure conceptualization
- Data structure modification

1.Data structure extraction:-

The data structure extraction, extracts the complete database schema. If there is a formal DDL description of the database, this process could be greatly expedited. Otherwise, a fair amount of data analysis, program analysis, and form analysis need to be performed. There are three steps, attributes extraction ,keys extraction, and constraints extraction in this extraction process. The aim of attributes extraction is to extract semantic information for database fields through field comparison, character comparison, data analysis, and code analysis.

Attribute Extraction:-

Field comparison is to compare form fields and database fields through instances in order to obtain the true meaning of each database field. First, we find the corresponding entries of forms fields in database fields, and then use captions of form fields and the context of the form to derive the meaning of database fields. In the process of comparison, form instances are the medium. Since the value of each field in form instances is different, by comparing values in form fields and database fields, we could identify the corresponding database fields readily. In this way, we are able to extract the semantics

of most attributes in the system

.

Key Extraction:-

After primary keys are established, we can apply primary keys to extract foreign keys in order to identify association among tables. During this process, every primary key is checked whether it is referred in fields of other tables (Alhajj, 2002). The referring field is a foreign key. The criterion is that the domains of the referring fields and the referred fields must match and the values of the referring field must be a subset of those of the referred field. Such criteria can be easily tested by designing appropriate SQL constraints

.

Constraints Extraction:-

The primary objective of constraints extraction is to obtain the association cardinality between primary keys and foreign keys. If a value of the primary key in a table shows in only one record in another table with the associated foreign key, the mapping cardinality is inferred to be one to one. Otherwise the mapping cardinality is consider done to many. In fact, if the foreign key relationship is already established, one only needs to check for the uniqueness of values in the foreign key fields to determine the cardinality.

2.Data structure Conceptualization:-

The conceptualization step concerns the formalization of the conceptual model and its refinement. For EED ,the focus is on the identification of entities and relationships among these entities from the logical schema gathered from the previous step. In the process of designing table-based database, an entity might be transformed into several tables depending on its characteristics. For example, a multi-valued attribute is usually separated from the entity to become another table.

3. Database structure Modification:-

After establishing the database connection there is need to define the structure of your database. This structure is necessary to ensure that the database components can show meaningful data in your pages. There are two methods of the structure of database: There are two methods of informing for structure of your database .i.e. Automatic database structure definition Manual database structure definition:

Automatic database structure definition:-

Eg:The most common method of informing NetObjects Fusion 12 about the structure of your database is to let NetObjects Fusion 12 import the structure information from the Data Source server or file.

Manual database structure definition:-

If you are unable to connect to your database or want to add specific tables and fields, you can modify the database structure manually.

Need

Output in the form of entity relationship diagram is difficult to understand by user and it will not provide detailed information about every object. so, we can display the output in different forms.

VII. Advantages

Data extraction methods will be used for the digital forensics. Digital forensics is a branch of forensic science encompassing the recovery and investigation of material found in digital devices, often in relation to computer crime.

By acquiring these techniques it will be possible for the database user to detect database tampering and dishonest manipulation of database.

VIII. TAMPER DETECTION APPROACH

With the Database there are several things and ideas come with the database operation.

The First approach: Audit log maintain by the DBMS itself as a background. This background audit log representing individual relation and this individual relation is treated as a Transaction Time table. In DBMS we perform updating, Deletion and modification operation on data (Tuple) if this operation take long time the Audit log and Transaction time table Drill the DBMS to keep the previous tuple during this operation with their insertion and deletion/update time. During this The DBMS provide one important property with the stored Data in database that it is Modification. If want to modify the only add information at End no information is Deleted. If we change the old information that time the data get tampered.

The Second approach: The Transaction made the cryptographically hash for the modify data to generate the secure one-way hash of Transaction.

The Third approach: By using the external notarization service we notarize the hash value because of this the intruder, operating system and hardware cannot change the hash value. If the intruder, operating system and hardware makes any Changes in hash value it is very difficult to make the hash value for this change hash value regarding to the Audit Log.

The Fourth Approach: Finally the matching is performed between old hash values with rehash tuple. If hash value is same there is no problem but if matching is not occurred then we need to apply forensic analysis algorithm to find out where, when and why the tampering has been occurred.

IX. Database Forensic Investigation

Forensic investigation is a branch of forensic science relating to forensic study of database and their related metadata. Databases often contain information that may be useful during many forensic investigations. i.e. applying investigation technology to database contents and metadata. Forensic investigation focus on identifying transactions within database system. Forensic investigation is a field to investigate the computer crime. Here the forensics investigator should be able to track an attacker on the Internet. The IP address and Domain name tracing is used to detect the suspicious user.

X. Need of forensic investigation

As internet is widely used there is increased in cyber crime. So only security apporoch is insufficient therefore database forensic investigation is needed. Authenticated and authorized user access data using various mechanism provided by database server but some time authorized user makes data get tampered so system is also not secure and protected. Authorized user access database with the help of IP address and try to make some modification in database like changes in item price and changes in item quantity, this changes provides financial loss due to this issue forensic analysis is necessary. Authorized and unauthorized user detected by Tiled bitmap forensic analysis algorithm. It also helps to find at what time and possibly who and why tamper the database.

For ex. A cracker breaks into online shop's database and interchanges the cost price and selling price columns assuming a relational database. This leads to a significant loss for shop.Therefore,it is important to adopt a forensic investigation method for database system.

XI. Process of forensic investigation:

There are four major steps:

1. Collection
2. Examination
3. Analysis
4. Reporting

First of all there is need to collect the evidence from computer system. This evidence or data collection can do by various discussion related to crime. These collected evidences will have to examine. This examination pay attention to various factors related to crime. Examinar try to find the answer of “why” and “whom”. After this evidence will take place into analyses phase .After collecting various proof finally report will generate that show investigation result. Final report should consist four things:

Who did

What did

When did

How did

Log files are important to investigate crime. It is a file that lists actions that have occurred. Log file contains whole information regarding the user’s activity. These activities is written in various log files like web log, firewall log, network log. With log files it is possible to get good idea of where visitors are coming from, how often they return and navigate throught a site.

The attacker can leave the evidence behind that can be collected by certain ways by forensic tools for the purpose of further investigations. The use of computer and digital devices in the act of crime is continuously grow day by day, so this gives challenges to forensic that how they collect information from the system after an incident. Most of the digital forensic tools are commercial version, which cost is high and operated by professional forensic, so we mostly use open source forensic tools because it is easy to use and less costly.

XII. Discrimination

Discrimination can be used in various fields such as in databases, data mining, forensic investigation, etc. Discrimination is simply known as injustice. Discrimination denies the members of one group with others. Unfairly treating people on the basis of their belonging to a specific group, namely race, ideology, gender, etc., is known as discrimination. Laws are designed to prevent discrimination. Antidiscrimination laws have been adopted by many democratic governments.The problems of assessing the presence, extent, nature, and trends of discrimination and of preventing discrimination in (possibly automated) decision making are thus of primary importance. There are several decision-making tasks which are supported by information system and lend to discrimination, e.g. loan granting, education, health insurances and staff selection. Thus, these collected data are auxiliary utilized by companies for decision making purpose .The association and or classification rules can be used in making the decision for loan granting and insurance computation. The use of information systems based on various technology for decision making has attracted the attention of many researchers in the field of computer science.

XIII. Types of Discrimination:

Direct Discrimination

Direct discrimination is intentional and “directed” towards individuals, typically on the basis of their visible traits. Direct discrimination is difficult to prove, since the complainant has to demonstrate the intent to discriminate. Direct discrimination consists of rules or procedures that explicitly mention minority or disadvantaged groups based on sensitive discriminatory attributes related to group membership. In this, classification of the data is done in such a way that focuses on independent sensitive attribute. Direct discrimination can be prevented by removing discriminatory attributes from the dataset.

Indirect Discrimination

Indirect discrimination consists of rules or procedures that, while not explicitly mentioning discriminatory attributes, intentionally or unintentionally could generate discriminatory decisions. A presumption of indirect

discrimination on a group is typically based on observing that the effects of some rules or practices have put the group in an adverse position. Indirect discrimination can be prevented by removing non-discriminatory attributes from datasets.

Causes of discrimination:

1. **Prejudice**

Prejudice means to judge another person or group a priori.Prejudice leads to discrimination when it concerns unfairly or unreasonably-formed negative attitudes against a protected group.

2. **Statistical thinking**

Statistical thinking is also known as rational racism. This is the case, when employers refer directly or indirectly to the average performance of the applicant's racial group as a decision element.

3. **Unintentionally discrimination**

This type of discrimination occurs due to indifference, incorrect (execution of) procedures or practices, lack of planning and analysis of decision outcomes. Indirect and unintentional discrimination have considerable problems for data analysts to carefully take into account the effects of decisions due to their unforeseen discriminatory effects.

XIV. Strategies

1. **Affirmative**

Affirmative actions are also called positive actions.A range of policies to overcome and to compensate the problem by providing opportunities for those who are traditionally discriminated

2. **Reverse discrimination**

Reverse discrimination, sometimes identified with affirmative actions, is the disadvantage that the non-members of protected groups suffer as a result of affirmative action. It is therefore important to assess and to monitor the application of affirmative actions.

3. **Favoritism**

Favouritism occurs when individuals are treated better than other for certain reasons. Discrimination and favoritism are dual concepts: if a protected group is discriminated against in a certain context, then the remaining people in the same context are favored. Strictly speaking, however, we reserve the term favoritism for the unfair (positive) treatment of members of a specific group, and not as the implicit consequence of the discrimination of other groups.

XIV. Acknowledgment

We take this opportunity to express our gratitude and indebtedness to our guide **Prof. P. D. Bodakhe**, and H.O.D, **Prof. A. B. Raut**, Computer Science and Engineering department, who is a constant source of guidance and inspiration in preparing this work. Her constant help and encouragement helped us to complete our seminar report.We are grateful to Principal **Dr. A.B. Marathe,** for his encouragement and support.We are also thankful to all the staff members of Computer Science and Engineering department, whose suggestions helped us to complete the seminar work and those who have directly and indirectly helped for completion of the seminar report.

XV. Conclusion

Database reverse engineering gives output in the form of ER diagram which is not easy to understand. So we can modify output using various methods. If tampering is done in our database then you can detect it by forensic investigation. Forensic analysis inaugurate sat what time a crime has been identify and in this case the tampering of a database. Such analysis activities determine when the tampering occurred, and what data were altered.

REFERENCES:

[1].Alhajj, R., Extracting the extended entity-relationship model from a legacy relational database. in proc. Information Systems 28, Elsevier Science Ltd, 597–618, 29 May 2002

[2].Deepak Meena, Hitesh Gupta “Digital Crime Investigation using Various Logs and Fuzzy Rules: A Review”. International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 4, April 2013.

[3]Piyush P. Gawali1, Dr. Sunil R. Gupta “Database Tampering and Detection of Data Fraud by Using the Forensic Scrutiny Technique”. International Journal of Emerging Technology and Advanced Engineering

[4].Pavlou, Kyriacos E., And Richard T. Snodgrass. "Forensic Analysis Of database tampering.“ ACM Transactions On Database Systems.

[5].S. Ruggieri, D. Pedreschi and F. Turini, “DCUBE: Discrimination discovery indatabases".