

# Visual Role Mining Using Adviser and Extraction Algorithm in Business Environment

Merin Jose

PG Scholar

Department of Computer Science &  
Engineering  
KCG College of Technology  
Chennai, India

Merlin Cyriac

PG Scholar

Department of Computer Science &  
Engineering  
KCG College of Technology  
Chennai, India

Mereen Thomas

PG Scholar

Department of Computer Science &  
Engineering  
Anna University  
Guindy Campus, India

**Abstract**—This paper offers a new visualization approach to Role based Access Control mentioned to as Visual Role Mining. The main idea is to reduce the query generation time in the databases. So that it will provide a text file retrieval model instead of query in databases. First it formally converts the data in the databases to some text files. By the help of ADVISER algorithm and BicOverlapper tool, the data will be converted to some picture formats. Later on the entire data will be extracted by the help of an Extraction algorithm. The algorithm also helps to convert the data into appropriate visual representations which supports role engineering process.

**Keywords**—Role Based Access Control, Visual Role Mining, Visual Representation.

## I. INTRODUCTION

Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make different analysis. The first and simplest analytical step in data mining is to describe the data summarize its statistical attributes, visually review it using pictures, shapes, charts and graphs, and look for meaningful links among attributes. Data mining is increasingly popular because of some research aspects too. It involves databases, data management, data preprocessing etc.

Graphing and visualization tools are a vital aid in data preparation and their importance to effective data analysis cannot be over emphasized. Visualization works because it exploits the broader information bandwidth of graphics. It allows people to go more inside to the concepts rather than the outlook. It is easier to find out the mistakes in graphical view than other textual formats. It is similar to a smart art which gives more information than a simple description.

### 1.1 Role Based Access Control

The goal of role engineering, by Edward Coyne, is to define a set of roles that is essential, correct and efficient. In particular, role engineering requires defining roles and assigning permissions to them. Role engineering is essential before all the benefits of RBAC can be realized. Meanwhile, role engineering, considered as one of the major challenges

RBAC implementation is a time-consuming and costly process. Due to this, organizations are often reluctant to move to RBAC. Therefore, the increasing popularity of RBAC calls for efficient solutions for role engineering as results in tremendous research efforts in this area.

There are three basic approaches towards role engineering: top-down, bottom-up and hybrid. Under the top-down approach, roles are defined by proper analysis and decomposition of business processes into smaller units in a functionally independent ways. These functionalities are then associated with permissions on information systems. In other way, this approach begins with defining a particular job function and then creating a role for this job function by associating needed permissions.

Access control mechanisms are crucial design elements that aim at mediating requests to data and services. Among all models proposed in the literature, Role-Based Access Control (RBAC) has become the norm for managing permissions within commercial applications. The high-level formalism and the simplicity of its design made it an attractive and pragmatic choice for implementing access control. Under RBAC, a role is a set of permissions, while users possess the permissions to perform system functions only when they are assigned to specific roles. Because of the intuitiveness of RBAC, security policies can be easily defined by business users that do not usually have all the needed IT knowledge.

Role mining is the process of analyzing user-to-resource mapping data to determine or modify user permissions for role-based access control (RBAC) in an enterprise. In a business scenario, roles are defined according to job's peculiarity, authority and responsibility. The ultimate intent of role mining is to achieve optimal security administration based on the role each individual plays within the organization. Role mining is commonly using in different environments. Other than security it holds simplicity too.

## II. RELATED WORK

The role engineering problem through a top-down perspective was illustrated by Coyne et al. [3]. Kuhlmann et al [8] was first trying to apply existing data mining techniques to elicit roles from accessed data. He introduced the term “role mining”. After that, different algorithms explicitly designed for role engineering purposes were proposed. Molloy et al. [10] presented a comprehensive study to compare them and a brief survey on the subject. Colantonio et al. [1] recently addressed the problem of analyzing the role mining complexity by also proposing a way to reduce it. In general, this approach can be considered a complement for all the existing role engineering methodologies and tools. Indeed, it allows a good, executable, and visceral way to evaluate and select roles generated by other methodologies.

There are different role mining techniques available for clustering, classification, extraction, mining. Among them some filtering mechanisms are more common. They are scan count, divide skip, merge skip.

Another related work is proposed by Geerts F. et al.[6], where a branch-and-bound algorithm for mining large “tiles” (that is, regions of database consisting purely of one’s) is introduced. It shares with the interest on finding large tiles only indeed, here the focus on the problem of visually representing tiles. A similar problem is partially addressed in 2008. M. Frank et al.[5] show a possible way to build a matrix representation of user-permission relationships. However, this generation is limited to the special case of non overlapping roles, far from being general and optimal according to definition. Moreover, it is not applicable to generic role mining approaches.

As for visual representation of mined data, a small number of visualizers have been proposed in different literature, and most of them are not explicitly designed for a particular data. The BicOverlapper tool integrates on a set of well-known visualization techniques that represent different data information on different levels. However, typical representations for each data such as repeating rows and columns of the analyzed matrix are confusing or not suitable for role mining. Jin R et al. [7] propose a visualization algorithm that extends existing graph sorting algorithms to offer a good matrix visualization of previously defined “hyper graphs” which can be mapped to the “role” concept in the RBAC terminology. Leung and Carmichael [9] developed a visualizer for frequent item sets based on multiline calls polyline. However, frequent item sets are not the only relevant patterns for role engineering. This approach greatly differs from the actual implementation:

1. Adopt a different visualization cost metric that is more suitable for role engineering incompatible with the core of their theory.
2. Show how to obtain a matrix representation without resorting to any existing mining algorithm.

Roles are treated as sets of permissions: Each row in the list is a role. Equivalently, a user role is characterized by the set of permissions that he owns. Vaidya J et al.[13] proposed a

model, these sets are determined by the row  $i$  in  $x$ , this model is equivalent to one of the instances of this model class if no underlying probability distribution is considered. In the algorithm proposed by Vaidya J et al [12], all existing users are initially considered as candidate roles. Thus, each candidate role consists of all permissions that are assigned to a particular user. Afterwards, candidate roles are picked in a greedy manner to determine the final set of roles. A similar procedure is proposed by Santamaria G et al.[11]). But there, roles and permissions are represented as sets of users. The initial roles are constructed from existing permissions. Cherichetti F et al [4] introduces, the roles are also represented as set of permissions. Candidate roles are generated and then merged, split, or placed in a role hierarchy, as determined by a small set of given rules. Namely, an initial role is the set of users that are assigned to a given permission. Again, the initial roles are iteratively merged, split, or placed in a role hierarchy according to the cardinality of intersections of the roles.

## III. ROLE VISUALIZATION PROBLEM

Recently, there has been an increasing interest in using automated role engineering techniques. Despite much work dedicated to the design of role mining algorithms, existing methodologies deal with three main practical issues: meaning of roles, noise associated with the data, and interconnections among roles.

To address the issues, a new approach, referred to as visual role mining. User-permission patterns (i.e., RBAC roles) among each individual are managed as visual patterns. The principle behind this approach is that visual representations of roles can actually amplify cognition, leading to optimal analysis results. Visualization of the user-permission assignments is performed in such a way to remove the noise, allowing role engineers to focus on relevant patterns, purchasing their cognition capabilities. Further, connections among roles are shown as different patterns, hence providing a visual manner to discover and utilize these relations.

### 3.1 Role Visualization

Given a set of already discovered roles of interest, the task is to identify the best graphical representation for them. In particular, the representation for user-permission assignments that allows for both an intuitive role validation and a visual identification of the relationships among roles. The proposed method shows that roles are easier to recognize than describe via a binary matrix representation. The proposed method can answer questions that statistical or mining approaches cannot easily provide. It will provide an easy way to analyze data within the text files rather than databases. It will also represent data in some visual manner like charts or in other easily understandable form.

### 3.2 Binary Matrix Representation

A normal representation for this information is the binary matrix, where rows and columns correspond to users and permissions, and each cell is “on” when a certain user has a certain permission granted.

The table 1 shows the input data which contains the users and the corresponding permissions associated with it. There will be different users associated with different permissions. According to the permissions, the users will be classified into different groups. Now the roles can be retrieved according to these groups. Table2 represents the candidate roles retrieved.

Table1 Input Data

User-Permission Assignments
{ <u0,p1>, u0,p3>,<u0,p8>,<u0,p9>,<u1,p1>, <u1,p2>,<u1,p3>,<u1,p4>,<u1,p6>,<u1,p8>, <u1,p9>,<u2,p1>,<u2,p2>,<u2,p4>,<u2,p6>,<u2,p9>,<u3,p1 >,<u3,p3>,<u3,p8>,<u3,p9>,<u4,p1>,<u4,p3>,<u4,p8>,<u4,p >,<u5,p1>,<u5,p2>,<u5,p4>,<u5,p5>,<u5,p6>,<u5,p9>, <u6,p1>,<u6,p2>,<u6,p4>,<u6,p6>, <u6,p9>,<u7,p0>,<u7,p1>,<u7,p7>, <u8,p0>,<u8,p1>,<u8,p7>,<u9,p1>}

Table 2 Candidate Roles

Role	Permissions	Users
r1	{p1}	{u0, u1, u2, u3, u4, u5, u6, u7, u8, u9}
r2	{p2, p4, p6, p9}	{u1, u2, u5, u6, u9}
r3	{p3, p8, p9}	{u0, u1, u3, u4}
r4	{p0, p7}	{u7, u8}
r5	{p5}	{u5}

#### IV. PROPOSED SYSTEM

By leveraging on the observations made in the previous section, it describes a viable, fast heuristic algorithm called ADVISER (Access Data Visualizer). For a given a set of roles, this algorithm is able to provide a compact representation of them. In particular, it rearranges rows and columns of the user-permission matrix to minimize the fragmentation of each roles associated to it.

ADVISER, the more fragments in the visualization of a role, and then the role visualization cost will get increased. Reordering users but not permissions only affects the number of gaps between columns, and so do Permissions (i.e., Rows and columns are sorted independently).

According to the expectation, the visualization cost decreases as the number of samples increases. Finally, extensive applications over real and public data confirm that this approach is efficient, reliable both in terms of computational time and result quality of the product.

#### 4.2 Data Access Control

Access control is the process of controlling requests to data and services maintained by a system, determining which requests should be granted or denied.

In these module, assigning privileges to an application users for achieving Data Access controls. (i.e.,) the project is choosing a banking application to prove a visual role mining, one of a major role inside banking is Branch manager how they eliciting meaning role for employees are under consideration are fully comes under these modules.

#### 4.3 Visualization

The technique devises a new approach, referred to as visual role mining. User-permission patterns (i.e., RBAC roles) are managed as visual patterns. The principle behind this approach is that visual representations of roles can actually amplify cognition, leading to optimal analysis results. It offers a graphical way to effectively navigate the result of any existing role mining algorithm, showing at glimpse what it would take a lot of data to expound. Moreover, it allows to visually identifying meaningful roles within access control data without moving to classical role mining tools. Picturization of the user-permission assignments is done in such a way to isolate the noise, allowing role engineers to concentrate on relevant patterns, purchasing their cognition capabilities.

These patterns are usually referred to as tiles. It demonstrates that it could be easier to find more patterns if users and permissions were reordered. The focus will be in turn going to the patterns which easier analysis methods.

#### 4.4 Content Manipulation

A data manipulation language (DML) is a family of syntax elements similar to a computer programming language used for inserting, deleting and updating data in a database.

But in Role Mining concepts without using database administration it tries to perform manipulations. But it is not possible to perform entire operation based in files only few operations are considered as a role because of security reasons. Later it performs some manipulations like inserting, deleting as well as setting permissions to their employees.

A visual performance can highlight possible exceptions within data in an effective manner and a textual role representation reports on information about role-user and role-permission relationships in a less communicative fashion than a graphical representation.

The proposed system has two phases; first one is similar to a web application. There all the users can login to the system with a password and a user name. All those details will be automatically saved to the database. That is known as

application database. The application database will provide the basic information about the users as and when required. The users present in the application database are considered as the application users. Among the entire application users the main user would be the application controller. The application controller schedules the tasks for each handler. Each handler is responsible for executing different tasks according to the needs. All these tasks can be viewed, controlled and modified by the admin. The administrative agent can insert, delete or update some vital information.

The second phase contains the processing of the entire data. The data present in the database will now be converted to some text files. This can avoid the triggering time of the database to process a query. These text files will be the input of the second phase. The BicOverlapper tool will generate corresponding picture/graph formats when the text files given as the inputs. ADVISER algorithm performs data zooming techniques on the picture formats. This will give a clear idea about the entire data. The EXTRACT algorithm will generate a zoomed data along with some extraction techniques.

For example, consider a bank application; it includes bank manager, other employees like cashier, accountant, loan manager & customers who are approaching the bank for some banking services. Those details will be present in the application database. By examining the application database the bank manager will be assigning certain roles to each of the employees/handlers. The application controller will be the bank manager itself. Here comes the bank application. Later on it will be dealing with the administrator. The administration agent has the authority to control the entire system. There will be two text files generated for picture generation. These can be converted to different formats by the help of a BicOverlapper tool. The system can implement data zooming and data extraction by the help of adviser and Extraction algorithm.

#### 4.5 Branch Management

The first module of the proposed system is a branch management; here the employee is the branch manager. Manager can login to the system with some username and password. There will be a number of duties assigned to the branch manager. The duties mainly includes the adding/removing the employees. That is whenever some new employees comes to the branch then the branch manager will be adding the users, in the same way whenever that particular employee moves from the branch that user will get removed. He has the provision to view all the employee details, in addition to that he will be registering, setting permissions, work assigning ... etc.

#### 4.6 Employees & Users

Employees are those who are working in the particular branch. The employees may be cashier, accountant or loan manager. The roles assigned to them will be cash management, accounting, and loan management respectively. The cashier can go through all the transactions. The accountant is

responsible for allowing each customer to start an account by sanctioning their account number. The loan manager will be responsible to grant loan for each customer. It can be either home loan or some personal loan. All the employees have the permission to change their password, to go through the work assignments given by the bank manager.

Users are the customers those who are approaching the bank for certain services. First each customer should register for starting an account. The random number generated will be the account number that is going to be assigned to the customer. But it will be ensured by the accountant before that the user cannot log in. After allowing the users to login to his area, they have some provisions to deposit money and some transfer mechanisms. The main facility provides is that they can request for certain loans and those will be redirected to the loan manager's desk. Then the permission will be granted.

#### 4.7 Text Files Generation

This module classifies the employees and the customers into text files. All the data will be available in the database. Whenever a user is registered the data will taken as a string and it will be converted to some text formats. The same thing is happening to the customer data too. This data will be used as the input for the next module. For that some colors are assigned to distinguish each role for the employees and each loan for the customers.

#### 4.8 Picture Format Generation

This module makes use of a bicOverlapper tool. BicOverlapper is a framework to support visual analysis of expression by means of biclustering. In order to improve the visualization it provides new methods in conjunction with ADVISER algorithm which is specified earlier. The BicOverlapper tool will provide certain picture outputs by taking certain texts as inputs. This visualization technique is integrated in BicOverlapper, along with several other algorithms and techniques. The input given to the tool is the text files that previously generated. This tool will generate certain picture formats according to the input data. It follows the adviser algorithm techniques also. The output generated can Perform zooming techniques.

#### 4.9 Data Zooming & Data Extraction

The data zooming and data Extraction are done by the help of two algorithms. One is adviser algorithm and other is Extract algorithm. The Adviser algorithm is implemented in the first phase.

#### 4.10 Adviser Algorithm Description

As a heuristic, ADVISER is based on some intuitions, summarized in the following: Larger roles should be better represented. The more fragments in the visualization of a role, the higher the role visualization cost. Reordering users but not

permissions only affects the number of gaps between columns, and so do permissions.

As for the first point, one can argue that small roles can be more important from a business perspective since they likely represent administrative tasks. To focus on exceptions, large roles can be removed after their identification. Notice that searching for large-area tiles is also the choice of many other mining techniques. A detailed description follows:

1. Rows and columns are sorted independently. ADVISER decomposes the optimal matrix-permutation problem into two sub problems, that is users and permissions are sorted independently. Due to this symmetry, from now on it generically refers to rows and columns as items.
2. If some items are assigned to same roles, they are put together. For this reason algorithm puts groups of items called item sets, instead of individual items.
3. Item set positions are decided one by one. In order to facilitate a better representation of roles, item sets involving roles with larger areas are analyzed first.
4. The algorithm tries to avoid the large gaps by putting item sets close to each other when they share large roles.
5. Each item set is preferentially positioned at the beginning or at the end of already sorted item sets.
6. Item set sorting is converted to item sorting.

#### 4.11 Extraction Algorithm

Extraction algorithm is used to extract data from the text files.

```
init(picture assigned to role,type Of Mining,values)
{
  File file = new File(path);
  if(file.exists()) {
    FileInputStream fis = new FileInputStream(path);
    byte buffer[] = new byte[fis.available()];
    fis.read(buffer);
    String a = new String(buffer);
    String b = a;
    fis.close();
```

## V. RESULT AND ANALYSIS

In order to represent the visual role mining, a bank application has been taken as a model. The bank manager can login to the system and he can perform certain updation. As described the options available to the manger are view employee details,add/remove users,setting permissions,employee registration, work assignment etc... The manager is responsible for creating a cash manager. The Employee will be generated in that particular id. The manager can register and assign works to that corresponding employee. After creating each employee we have to register them into the application database, which will be added to the database. It includes employee id, password, address, mail id, and role.

Every user has the provision to change their password after their login. Every cashier will have an online desk, which includes all details of every transaction. It includes the attendance register of each employee. It adds the no of days that particular employee took leave.

There will be certain works associated with the users for getting into the system. The main step is the registration part. The user should register to start an account. All the basic information should be provided. The accountant will be allowing each customer to start an account. Hereafter the customer can use the facilities offered by the bank.

The data which is available in the application database has been converted in to text files as employee data and customer data. These files are considered as the input for the BicOverlapper tool. This data will be converted to different patterns as shown in the 6.. The color difference indicates the different roles associated with each employee. The type of loan makes the color difference in the case of customer data. The analysis part becomes easier when the size of the file increases, comparing with a database.

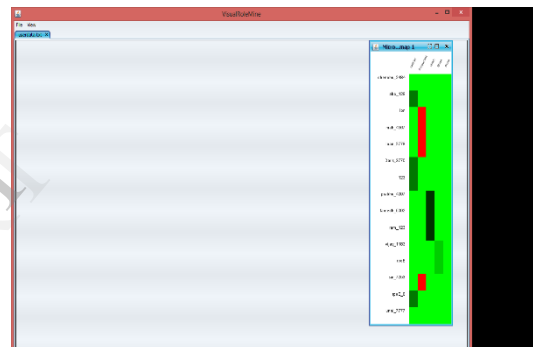


Figure 6.1 Picture Format

The Extract algorithm will be extracting the data from the text files. It is capable of representing the data into some pictorial representations like bar charts.

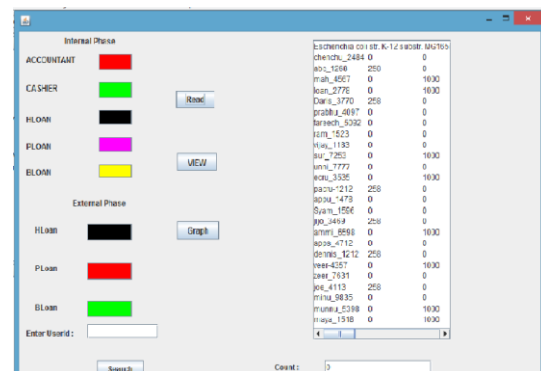


Figure 6.2 Extraction process

## VI. CONCLUSION

This paper is mainly addressing the visual role mining problem. That is, visualizing user-permission assignments in a graphical form that makes it possible to simplify the role engineering process. The proposed representation of data allows role designers to gain insight, draw conclusions, and ultimately design meaningful roles in business applications. The paper offered a formal description of the visual role mining problem. Then it demonstrated a banking environment which includes all the transactions. The people included are employees and the customers associated with the bank. Moreover, it proposed a novel algorithm called ADVISER in conjunction with a bicOverlapper tool to generate a visual representation. The bicOverlapper tool produces approximate patterns that can be used in conjunction with ADVISER to obtain high-quality visualization results. Finally, extensive applications over public data confirm that this approach is efficient, both in terms of computational time and result quality. It also described an efficient algorithm referred to as Extract algorithm. The paper introduced role engineering as a process which can greatly benefit from the visual approach proposed in earlier years. Role engineering is definitely an active research topic with a high interest from both academy and industry, as witnessed by the rich literature. Our contributions, other than being useful for role engineering, can have interesting applications in other fields as well. For instance the query generation time is one of the disadvantages associated with databases. The paper proposes a novel solution for this problem by creating a text file instead of databases. All the algorithms will be dealing with the text files only. In particular, homogeneous sub matrices indicate subsets of rows co expressed under the same conditions columns. In this case, each transaction corresponds to a row and each item corresponds to column of the matrix. As for future work, our solutions can be extended in several directions. Approximated representations of data are just some examples of possible directions to investigate. Besides partitioning, as suggested in this paper, alternative representations might be taken into account to provide a compact representation of the information.

## REFERENCES

- [1] Colantonio A, Di Pietro R, Ocello A, and Verde N. V, "Taming Role Mining Complexity in RBAC," *Computers Security*, vol. 29, pp. 548-564, 2010.
- [2] Colantonio A, Di Pietro R, Ocello A, and Verde N.V, "Visual Role Mining: A Picture Is Worth a Thousand Roles," *IEEE Transactions On Knowledge And Data Engineering*, VOL. 24, NO. 6, pp. 1120-1133, 2012.
- [3] Coyne E.J, "Role-Engineering," *Proc. ACM Workshop Role-Based Access Control (RBAC '95)*, pp. 15-16, 1995.
- [4] F. Chierichetti, R. Kumar, S. Pandey, and S. Vassilvitskii, "Finding the Jaccard Median," *Proc. 21st Ann. ACM-SIAM Symp. Discrete Algorithms (SODA '10)*, pp. 293-311, 2010.
- [5] Frank M, Basin D, and Buhmann J.M, "A Class of Probabilistic Models for Role Engineering," *Proc. 15th ACM Conf. Computer and Comm. Security (CCS '08)*, pp. 299-310, 2008.
- [6] Geerts F, Goethals B, and Mielikainen T, "Tiling Databases," *Proc. Seventh Int'l Conf. Discovery Science (DS '04)*, pp. 278-289, 2004.
- [7] Jin R, Xiang Y, Fuhry D, and Dragan F.F, "Overlapping Matrix Pattern Visualization: A Hypergraph Approach," *Proc. IEEE Int'l Conf. Data Mining (ICDM '08)*, pp. 313-322, 2008.
- [8] Kuhlmann M, Shohat D, and Schimpf G, "Role Mining—Revealing Business Roles for Security Administration Using Data Mining Technology," *Proc. Eighth ACM Symp. Access Control Models and Technologies (SACMAT '03)*, pp. 179-186, 2003.
- [9] Leung C.K.-S. and Carmichael C.L., "FpViz: A Visualizer for Frequent Pattern Mining," *Proc. ACM SIGKDD Workshop Visual Analytics and Knowledge Discovery (VAKD '09)*, pp. 30-39, 2009.
- [10] Molloy I, Li N, Li T, Mao Z, Wang Q, and Lobo J, "Evaluating Role Mining Algorithms," *Proc. 14th ACM Symp. Access Control Models and Technologies (SACMAT '09)*, pp. 95-104, 2009.
- [11] R. Santamaria, R. Theron, and L. Quintales, "BicOverlapper: A Tool for Bicluster Visualization," *Bioinformatics*, vol. 24, no. 9, pp. 1212-1213, 2008.
- [12] Vaidya J, Atluri V, and Guo Q, "The Role Mining Problem: Finding a Minimal Descriptive Set of Roles," *Proc. 12th ACM Symp. Access Control Models and Technologies (SACMAT '07)*, pp. 175-184, 2007.
- [13] Vaidya J, Atluri V, and Warner J, "RoleMiner: Mining Roles Using Subset Enumeration," *Proc. 13th ACM Conf. Computer and Comm. Security (CCS '06)*, pp. 144-153, 2006.