# Role Mining: Survey and Suggestion on Role Mining in Access Control

Jinsuo Jia[1], Jianfeng Guan[1(✉)], and Lili Wang[2]

[1] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
{jjs,jfguan}@bupt.edu.cn

[2] Academy of Military Sciences, People's Liberation Army, Beijing 100141, China
wanglili2_2006@163.com

**Abstract.** With the increasing attacks of Network, various security defense mechanisms especially access control mechanism have become research hot-spots, in which Role-Based Access Control (RBAC) as one of the most popular mechanisms has been applied in many fields. However, the booming of various applications and huge users result in the difficulty of defining roles in advance. Therefore, lots of research efforts are focusing on role mining, which has an important impact on improving the function and performance efficiency of RBAC. By investigating and analyzing the related literature in terms of role mining, the development status of role mining technology can be divided into two aspects: the research of extended elements of role mining system and the improvement of existing role mining algorithms. Therefore, this paper summarizes and compares the advantages and disadvantages of ten role mining mechanisms with the objective to find the optimal role mining method via comprehensive comparison, and gives appropriate suggestions. In order to evaluate the role mining more comprehensively, the evaluation metrics included in each role mining mechanism are defined. Finally, this paper analyzes the problems and challenges of role mining, and gives the suggestions for further development.

**Keywords:** Role mining · Access control · Role-based access control · Problems and challenges

## 1 Introduction

With the rapid development of the information technologies, the usage of Internet has increased dramatically in every aspect of life. In the past three decades, Internet security issues such as the CIH virus in 1998, the Melissa virus in 1999, I love you virus ins 2000, the shockwave virus in 2003, Panda burning incense in

Supported in part by the National Basic Research Program of China (973) under Grant No. 2013CB329102, and in part by the Natural Science Foundation of China (NSFC) under Grant No. 61003283.

I. You et al. (Eds.): MobiSec 2019, CCIS 1121, pp. 34–50, 2020.
https://doi.org/10.1007/978-981-15-9609-4_4

2006, the conficker worm in 2008, the flashback virus in 2011, WannaCry bitcoin ransomware in 2017, and so on [1,2] have led to continuous exploration of cybersecurity protection mechanisms. Cybersecurity has seriously threatened people's daily production and life, and brought huge losses [3]. Currently, network security protection for intranets mainly includes: firewall, intrusion detection system and access control, in which access control is the first gate to protect the network [4].

In access control system, the most classic access control models are Discretionary Access Control (DAC), Mandatory Access Control (MAC) and Role-based Access Control (RBAC) [5]. The DAC model allows the owner of the object to determine the access rights of the subject to the object, which is mainly used in commercial systems and some civil organizations, such as common operating systems (Windows, UNIX systems), firewalls such as Access Control List (ACL). DAC may lead to illegal access and causing security risks. The MAC model identifies subject and object in the system according to the security level. The corresponding resources cannot be accessed without the corresponding security level. MAC is mainly applied to multi-level military security systems, such as the ministry of defense system, wartime command system and so on. However, this model will lead to the inflexibility of user access.

In order to solve the problem that DAC is too loose and the MAC is too strict, the RBAC model is proposed by David F. Ferraiolo and D. Richard Kuhn in 1992, they introduced the concepts and definitions of RBAC and described a non-autonomous access control method. In 1996, Sandhu *et al.* described a new RBAC reference model framework which systematically addresses the various components of RBAC and their interactions [6]. In 2000, NIST published an unified RBAC standard. Richard Kuhn *et al.* submitted a proposal to the 5th Role-Based Access Control ACM Symposium, evaluated and revised by NIST. In 2004, NIST RBAC model of American National Standards Institute and International Information Technology Standards Committee (ANSI/INCITS) was adopted as the US National Standard 359-2004 [7]. So far, RBAC model has been formed and entered the field of security application. In 2012, NIST RBAC was revised to INCITS 359-2012. Users in RBAC obtain the permissions corresponding to the role by obtaining the role. Therefore, RBAC is highly flexible and suitable for large-scale systems. The recently researches are focusing on Attribute-based access control (ABAC) [8] which is more complex and requires more processing power and time.

The development of RBAC has developed more than two decades, and it has derived many versions. The development of RBAC has also extended to various industries including medical system security, digital energy grid [9], software engineering, IoT [10,11], blockchain [12], and cloud computing [13], even Space and Terrestrial Integrated Network (STIN) [14], and has yielded fruitful achievements [15–17]. The most critical issue in RBAC is to mine accurate and appropriate roles to cope with the explosive application requirements. However, the generation of roles set is inefficient. For this reason, many researchers are committed to the study of role mining.

Role mining is one of the most important mechanisms and methods in the RBAC model. Without considering other aspects, the higher the efficiency of the role mining algorithm is, the better the performance of the RBAC mechanism will achieve. Therefore, this paper reviews and summarizes the literature of existing role mining in detail, and combines the above-mentioned application fields to discuss and supplement the cutting-edge research results, applications and suggestions of role mining.

The main contributions of this paper are as follows:

- We investigate and classify the current mainstream role mining technologies and find that the research of role mining mainly focuses on two major aspects. One is to introduce new elements into role mining mechanism to construct a new role mining model to improve mining efficiency and accuracy of role sets, and the other is to improve the performance of the existing role mining methods to make them perform better in more complex environments.
- We compare typical role mining models and their application scenarios, and analyze their advantages and disadvantages.
- We summarize the problems existing in the current role mining field and the challenges faced by the future development, and predict the future development direction of role mining, and provide reasonable suggestions.

The rests of this paper are organized as follows. Section 2 investigates the current status of role mining. Section 3 compares the performance of each role mining mechanism and summarizes the evaluation metrics. Section 4 summarizes the problems and challenges in the current role mining technology development. Section 5 predicts future development directions and gives suggestions.

## 2 Status of Development of Role Mining

Complex and diverse data information makes the various management system functions different. Therefore, deploying the RBAC model in various systems and finding a qualified set of roles is a difficult task. In June 2013, Aldo gave the definition of role mining [18], which can be defined as the process of analyzing user-to-resource mapping data to determine or modify user rights of RBAC in the enterprise. The roles in a given system environment are specifically divided according to work content, requirements, and responsibilities. The ultimate goal of role mining is to achieve secure and efficient system management based on the role which users play in the organization. Role mining can be done in three ways, the first being a top-down approach [18–20], the second being the bottom-up approach [18,21], and the third being based on the example [18]. In the bottom-up approach, users are assigned existing roles based on their own work content and responsibilities. A top-down approach is based on the role of the user's work content and responsibilities. By way of example, the system administrator defines roles that are consistent with the user's responsibilities and work content. However, the existing role mining models are inefficient, and they have to be changed by adding time, probability, graphics, and a mix of elements to improve its efficiency.

### 2.1 Researches on Extended Elements of Role Mining Mechanism

The extended element research of role mining mechanism refers to changing the existing role mining strategy and adding a certain element such as time, probability, graphics, weight and so on to extend the existing model. This way can be exemplified by the attributes and characteristics of the added elements, taking time as an example. This model can reflect the key characteristics of time in the generation and evolution of the role, so that the role can be generated over time, and can also decay over time.

**Time-Based Role Mining Mechanism.** Time-based Role Mining Mechanism (TRMM) selects the appropriate role set according to the change of time element. At present, many researches add TRMM to RBAC to form a new role set access control model that changes with time. Bertino *et al.* added time elements to a role-based access control model called Temporal Role-Based Access Control Model (TRBAC) [22]. The TRBAC model supports the enabling, disabling, and operational time dependencies of role over a period of time to realize temporal controllability. Similarly, Mitra *et al.* [23] also introduced time variables into the RBAC model, but their approach was to create a set of roles at each point in time, which allows RBAC to migrate over time instead of creating a role from scratch to deploy TRBAC. While in an earlier research, Mitra *et al.* officially defined the Time Role Mining Problem (TRMP) [24]. According to the user's permission assignment, from an existing set of tense to find a set of optimal role set. The method includes enumerating candidate roles and selecting greedy heuristic algorithm to select a set of minimum role sets which can find a set of the best set of characters from an existing set of tenses based on the user's permission assignment. Similarly for TRMP, Pan *et al.* also proposed a temporal approximation-based role mining approach for TRBAC [25], in which they focus on role mining with approximate time consistency rather than fixed-time nodes. In this way, the available time roles can be extended to non-fixed time nodes, so that the role mining can be applied more universally.

**Probability-Based Role Mining Mechanism.** Probability-based Role Mining Mechanism (PRMM) refers to select the best set of roles based on probabilistic statistical methods, which is highly adaptable and can choose the optimal set in any data set. The use of probabilistic methods to solve the problem of role mining is also a research hot-spot. Mario Frank *et al.* redefined role mining as a probability problem [26] to select the best role set. Probabilistic role mining mechanism does not depend on the advantages and disadvantages of data sets, and probabilistic statistics can be used to select the most appropriate role from the data sets. Compared with other models, this mechanism can be widely used to generate roles in various data sets, and it has strong generalization ability. However, this method is only suitable for rough role selection, not for precise role set mining. Therefore, Alessandro Colantonio proposed a new method [27] which introduces the availability and similarity metrics to measure the expected complexity of bottom-up analysis results and improves the calculation efficiency using two fast probability algorithms. This method is applicable to large organizations with hundreds of thousands of users and permissions.

**Graph-Based Role Mining Mechanism.** Graph-based Role Mining Mechanism (GRMM) divides user groups and permissions based on the implementation of graph elements, which can solve the problem of understanding the semantic representation of the role set difficulty. GRMM is a creative idea to solve role mining problem based on graph elements. Colantonio *et al.* graphically represented user privilege allocation to quickly analyze and motivate meaningful roles [28]. Graphic elements can be used to solve the problem of no semantics or difficult recognition in the process of role mining. However, the calculation of graphs is a more complicated process. Consultant algorithm and extraction algorithm can reduce the complexity and the operation process. The advisor algorithm can heuristically solve NP-complete problem in graph operation process [29]. The adviser algorithm does not need a predefined role set, and uses a visual-inspired character set to represent user permission assignment. This method provides a new idea for the application of graph in role mining, and tries to reduce the computational complexity. However, compared with the role mining method of time and probability, it is still very immature.

**Weight-Based Role Mining Mechanism.** Weight-based Role Mining Mechanism (WRMM) implements the extraction of important roles by mapping the priority of roles with weight values, and it can achieve different permissions for different users by mapping role priorities. The weight value measures the importance of role. Ma *et al.* proposed a weight-based role mining algorithm [30], which proposes the concept of weights to reflects the impact of weights in the system by mapping permissions, roles, and weight values. The weight value can correspond to the attribute of operation, sensitivity of the object, and the user attribute associated with the permission. The weight of permissions is calculated by exploiting the similarity between user and permissions. Compared to the traditional methods, it can scan the database's permission set based on the weight. However, in the process of calculating the weight of authority, this method needs to associate other attribute information, which may increase the complexity of calculation process and prolong the long time consumption.

**Mixed-Based Role Mining Mechanism.** Mixed-based Role Mining Mechanism (MRMM) combines multiple role mining methods to select the optimal role set. For example, Frank *et al.* proposed a hybrid model mining algorithm based on probabilistic elements, which quantitatively analyses the correlation between any type of business information and a given role, and merges the relevant business information into a probabilistic model of a hybrid model mining algorithm [31]. Mixed role sets are mined through combination of probability and statistics method and objective function of enterprise information. The experimental results show that this method can generate roles corresponding to business information well. However, the computational process is complex, and it is difficult to deal with large-scale complex systems. Zhai *et al.* proposed a hybrid method of role mining algorithm [32], which requires a top-down approach to defining the set of roles and then mining the candidate roles through a bottom-up approach. The weighted structure complexity is used as an indicator of system optimization and performance evaluation. This approach requires a predefined set of roles to increase the amount of work and time spent compared to other methods. Both mixing methods have certain advantages under their specific conditions, but they are poor in portability and universality.

### 2.2 Improvement of Role Mining Mechanism

In some special scenarios, some functions of the existing model cannot meet the requirements of current scene, so it is necessary to improve the function of a certain aspect of the existing character mining model. This paper selects several issues with high current attention including exploring roles that support overlapping permissions, mining "dirty data" and "noisy data" roles, compatible with existing role set methods, reducing the complexity of role mining systems, and finding the best set of roles. The role mining process consists of three steps. The first step is the search for role attributes, also known as the preprocessing stage. The second step is to create and run a role mining task, also known as the role detection phase. The final step analyzes the role mining results, configure and save the role [18], also known as the post-processing stage [33].

**Role Mining Mechanism with Noisy Data.** Role mining mechanism with noisy data, which is abbreviated as NdRMM, removes redundant data through noise processing to determine the optimal role set more accurately. Noise data processing belongs to the pre-processing stage of role mining, which removes erroneous data and transforms them into executable data sets. Molloy *et al.* cleaned the data before inputting data [34], and introduced a method of noise identification using (non-binary) rank reduction matrix decomposition. Experimental results show that it is effective in noise reduction. The process of mining roles is divided into two steps: eliminating noise and generating candidate roles. The evaluation results have also shown that this method can find a set of roles that are very close to the noise-free data. Therefore, this method is superior to the method of directly mining noise data.

**Role Mining Mechanism with Overlapping Privileges.** Role mining mechanism with overlapping privileges, which is abbreviated as OpRMM, can solve the role set problem of overlapping regions in the role mining process. OpRMM is critical to the process of determining the set of roles. Jaideep Vaidya *et al.* proposed an unsupervised method called RoleMiner which is used to mine roles from existing permissions. The essential task of role mining is to cluster users with the same (or similar) permissions [35]. Role mining needs to identify overlapping sets. The roles are those with overlapping permissions, which are implemented by counting the intersections between the initially discovered clusters through subset enumeration. This process is mainly for the role detection phase and is used for the determination of role set.

**Role Mining Mechanism with Minimum Perturbation.** Role mining mechanism with minimum perturbation, which is abbreviated as MpRMM, is used for the migration process of role mining systems. MpRMM can achieve minimal changes to existing systems. Most of the role mining methods are not compatible with existing roles, and all roles are defined from the beginning, which cannot be changed for the RBAC system that has been implemented. Takabi *et al.* proposed the definition of a mining hierarchy with minimal perturbations [36], and defined a heuristic algorithm called StateMiner which can maximize the proximity of the deployed RBAC state and the optimal state of the RBAC state. Zhai *et al.* used this algorithm as an metric to approximate the original character set, and introduced a similarity calculation algorithm [32]. On this basis, they proposed Minimum Disturbance Hybrid Role Mining algorithm, analyzed its complexity, and the evaluation results show that the accuracy and efficiency are significantly improved.

**Role Mining Mechanism with Reducing Complexity.** Role mining mechanism with reducing complexity, which is abbreviated as RcRMM, can reduce the complexity of complex role mining systems. RcRMM simplifies complex hierarchies, which facilitates the selection of role sets. Colantonio *et al.* proposed a solution to reduce the complexity of role mining [37] which can be divided into three steps. First, each role is assigned a weight. Second, the role-user privilege assignment which does not belong to roles whose weights exceed a given threshold is determined. Final, the role mining problem is limited to the role-user privilege assignment problem in the previous step. This solution is derived from graph theory, which allows role miners to select stable roles through context-simplified role selection tasks. To reduce the complexity of RBAC systems, and to define the concept of weighted structural complexity metrics, Molloy *et al.* [38] proposed a role mining algorithm for mining lower structural complexity RBAC systems. HierarchicalMiner and AttributeMiner are able to generate less complex RBAC states while retaining semantically meaningful roles and discovering new roles with semantic meaning. HierarchicalMiner has the ability to mine the role of maximizing system performance and generate an excellent character hierarchy.

**Role Mining Mechanism for Optimal Role Set.** Role mining mechanism for optimal role set, which is abbreviated as OrsRMM, can be used to implement approximate solutions to NP problems through heuristic algorithms. Guo *et al.* considered that the role hierarchy should assume the authority to mitigate security management, but no concept of optimal hierarchy has been proposed.

Therefore, They defined a formal indicator of the optimal role level mining structure [39]. The optimal concept is based on the role hierarchy as a graph and find the best role hierarchy, with the minimum number of edges to calculate the transitive closure. A heuristic method based on RoleMiner is proposed to achieve this goal. Vaidya *et al.* also introduced two different variants of Role Mining Problem (RMP) on how to find the correct role. One is delta-Approx RMP and the other is minimum noise RMP. Besides, they also showed that RMP is a NP-complete problem [21] and revealed the connection between several recognized problems in data mining and analysis role mining. After that, Igor Saenko and Igor Kotenko proposed a heuristic optimization method based on genetic algorithm (GA) [40] to solve RMP which develops a heuristic solution with the ability to find an accurate set of roles. By using chromosomes and genes in genetic algorithm to complete the crossover, mutation and selection process, a more appropriate set of minimal roles can be determined. As an algorithm for solving RMP problems, this method has high performance and efficiency, however, it is difficult to determine the number of population and active role set under special circumstances.

## 3 Performance Evaluation Metrics of Role Mining Mechanisms

Section 2.1 summarizes five mechanisms, including TRMM [22–25], PRMM [26,27], GRMM [28,29], WRMM [30], MRMM [31,32] which change existing role-based access control models by extending elements, and improve the efficiency and functionality of the new model by adding the characteristics of the elements. While Sect. 2.2 summarizes NdRMM [34], OpRMM [35], MpRMM [32,36], RcRMM [37,38], OrsRMM [21,39,40] five mechanisms which aim to select the optimal role set or improve the accuracy of role set. In this section, we will construct the evaluation metrics of the above ten role mining mechanisms.

In December 2016, Dong *et al.* proposed a data-centric model for predicting the best role mining results [41] without running any role mining algorithms. Different from Dong's algorithm, Molloy *et al.* compared different role mining algorithms [42] and proposed a framework to optimize and upgrade the role mining hierarchy, and evaluated the performance of different algorithms.

After analyzing the evaluation models and quality evaluation indexes of different role mining results, this chapter compares the ten role mining mechanisms in Sect. 2 from the aspects of introduce element, mechanism characteristic, application scenario and disadvantage to obtain the performance comparison of the role mining mechanisms in Table 1.

The metrics used in the comparison are shown as follows. Introduced Element: represents the mediation values for the mapping process between roles, permissions and users added to select a high-quality set of roles in the new model. Mechanism Feature: refers to the outstanding performance of various mechanisms compared with other mechanisms, such as flexibility, generalization ability, etc. Application Scenario: refers to the appropriate scenario for various role mining mechanisms, in which scenario the application can exert the maximum

**Table 1.** Performance comparison of role mining mechanisms.

| Type | Introduced element | Mechanism feature | Application scenario | Disadvantage |
|---|---|---|---|---|
| TRMM [22–25] | Time | Increase flexibility. Reduce the number of user rights allocation and role mining time | It is suitable for systems where the role is time dependent | The system is highly biased and difficult to apply in other systems |
| PRMM [26,27] | Probability statistics | It has strong generalization ability and strong expansibility | It is widely applicable to role mining of various data sets and relatively large-scale systems (about 100,000 users) | It is not applicable to an RBAC system with non-redefined probabilities |
| GRMM [28,29] | Graphics Mapping | Intuitive way to visualize user rights allocation; Quickly analyze and inspire roles without predefining the role set | A system that intuitively identifies meaningful roles in data | It does not work for systems that cannot be visualized graphically |
| WRMM [30] | Weight value | Greatly reduces the data processing times and quickly generates roles | A system that considers the different nature and importance of each permission, and identify permission sets with a small number of users | Weak sensitivity to roles with indistinguishable permissions |
| MRMM [31,32] | Business information | Visualization of role information and strong generalization ability | A role application with visual requirements associated with business information | The integration of business information will lead to an increase of pre-processing time |
| NdRMM [34] | Matrix | High Accuracy | Role mining with noisy data | The set of roles that completely cover the system cannot be found |
| OpRMM [35] | User rights assignment | It is highly effective and accurate | For data sets with overlapping permissions and high noise | Long preprocessing time and no semantic information |
| MpRMM [32,36] | – | High compatibility and slight disturbance | A system which has already deployed the RBAC | The high hierarchical complexity of RBAC system |
| RcRMM [37,38] | Weight value | Identify excellent roles and reduce system complexity | RBAC system with high complexity | The system performance heavily depends on the setting of the weight threshold |
| OrsRMM [21,39,40] | Graphics or Noise | Lighten management burden and generate a set of characters with high accuracy | It need to optimize redundant, complex and inefficient RBAC systems | Long execution time in data preprocessing and role generation |

performance. Disadvantage: refers to the imperfections of the existing mining mechanism. These metrics can inspire researchers to absorb the characteristics of mechanisms, and apply suitable scenarios, and improve the shortcomings of role mining, and enable the corresponding mechanisms to play the greatest role. Through in-depth study on the process of role mining mechanism, the quality evaluation metrics of role mining results is obtained in Table 2. In this Section, ten quality evaluation metrics were obtained to evaluate the role mining results. In Table 2, each mechanism that contains the corresponding evaluation metric is denoted as "Yes" and abbreviated as "Y". The mechanism excluding the corresponding evaluation metric is denoted as "No" and abbreviated as "N". Next, we will explain what each evaluation metric represents.

- Implementation Complexity (IC) denotes the complexity of the implementation process of a new model.
- Pre-Processing Time (PPT) represents the pre-processing time for various mediation values and roles, permissions, user mappings, and transformations in the new model. The longer the preprocessing time is the worse the performance and the shorter the time is the better the performance.
- Role Generation Time (RGT) represents the time required to generate the set of roles required by the model after pre-processing the mapping and transforming the relationship.
- Role Quality (RQ) represents the pros and cons of generating role quality. Role quality is equal to number of permissions/number of roles. The higher the ratio is the better the roles generated. Conversely, the worse.
- Extensibility (Ex) represents the ability of existing systems to extend other functions to their existing models. The application scenario, data model algorithm and so on in each model are judged.
- Compatibility (Co) refers to the migration of new models to other hardware and software systems. Or whether it can be covered with other RBAC systems, and whether to start from scratch.
- Similarity (Si) represents the degree of similarity of Roles' functions and strengthens the similarity between permissions by using similarity matrix.
- Intuition (In) indicates whether managers can directly identify and understand role meanings.
- Relevance (Re) refers to whether the new model needs training of large data sets.

## 4 Open Issue and Challenge

The RBAC model is currently the most widely used access control system, and its classic model brings new research ideas to researchers. However, there are many factors affecting the efficiency of RBAC model. Researchers are also committed to solving these difficulties to improve the efficiency of the RBAC system, and strive to maximize the efficiency of the model. The role integrity and role management efficiency generated by the role mining process is undoubtedly a

**Table 2.** Quality evaluation metric of role mining results.

| Type | IC | PPT | RGT | RQ | Ex | Co | Si | Ln | Re |
|---|---|---|---|---|---|---|---|---|---|
| TRMM [22–25] | Y | Y | Y | Y | N | Y | N | N | Y |
| PRMM [26,27] | Y | Y | N | N | Y | Y | Y | N | Y |
| GRMM [28,29] | Y | Y | N | Y | N | N | Y | Y | Y |
| WRMM [30] | Y | N | Y | Y | N | Y | Y | N | N |
| MRMM [31,32] | Y | Y | N | Y | Y | N | N | Y | Y |
| NdRMM [34] | Y | Y | Y | Y | N | N | Y | N | Y |
| OpRMM [35] | N | Y | N | N | N | Y | N | N | N |
| MpRMM [32,36] | Y | N | N | Y | Y | Y | N | N | Y |
| RcRMM [37,38] | Y | N | Y | Y | N | N | N | N | Y |
| OrsRMM [21,39,40] | Y | N | N | Y | Y | N | N | N | Y |

key technology for the efficiency of the RBAC model and the core project of the RBAC system. This paper explores and discusses the open issues related to current role mining techniques.

### 4.1 Minimizing the Role Set Problem

In all the literature reviewed in previous section, the exploration of the best roles set is undoubtedly the largest problem which has been proven to be NP-hard. In order to optimize the number of roles that cover all current user privilege assignments, the known minimum number of roles [43] can also be modeled as the Graph Coloring Problem (GCP). The current research result shows that the optimal number of role is roughly concentrated around its expected value [44–46].

### 4.2 RBAC Migration Cost Problem

The huge workload of RBAC policies migration is a main obstacle to the adoption of RBAC systems in large organizations. Some migration-related role mining algorithms can significantly reduce the cost of the migration process. For example, Xu and Stoller used a strategy mining algorithm to parameterize the role, so that the parameterized results are added to the candidate role set to complete the migration process [47]. Molloy *et al.* also studied the migration of non-RBAC systems to RBAC systems, and applied data mining to role mining to make up for the high-cost top-down approach in the migration process, which can simplify complex character hierarchies and dig out excellent character sets [48]. The migration problem of RBAC system can also be viewed as to reduce the complexity of role mining algorithm. This is a practical problem to deploy RBAC systems, but the design of unified algorithm or strategy to achieve low-cost migration is difficult due to the differences of the deployment environments

[49]. Therefore, Pan *et al.* proposed a model of high flexibility and applicability from the perspective of reducing the structure of the RBAC system, which can reconfigure the RBAC system with minimal structural complexity and perturbations [50]. However, the actual application effect needs to be continuously explored by researchers in the future.

### 4.3 Role Mining Problem with Semantic Information

System administrators generally do not want to assign a role that is completely incomprehensible. Since the role mining process requires multiple iterations [51], most of the intermediate results are not semantic. There are only a few studies of role mining algorithms that contain specific semantic information. Rosen-Zvi *et al.* proposed a technique for extracting information about authors and topics [52] which is essentially a statistical model based on probability, and the probabilistic theme is extended to include author information and using Markov to learn the author's subject from the data in an unsupervised process. This was an attempt to solve the semantic information problem and has achieved well results. Semantic information can be abstracted into attribute information. Besides, Molloy *et al.* proposed a role mining method for response authority usage and user attributes [53], which provided several models that can find a causal relationship with permission usage, including user attributes that are arbitrarily combined by this information, and a mining algorithm for the association of natural and semantic information for role mining.

### 4.4 Role Mining Results Evaluation Criteria

The merits of role mining results need to have an accurate evaluation criteria [51]. Zhang *et al.* have done related research work [54] and used five algorithms to verify the validity of the role mining results. The TF-IDF algorithm is used to mine the semantic tags for each role. However, a general criteria to evaluate the quality of role mining results is still missing. For the quality assessment of the role mining results, we can refer to the assessment [55] that ABAC model attributes are automatically extracted and the assessment using a calculation of expansion attribute strategy [56]. The definition and extension of these evaluation criteria are of guiding significance for future research.

After discussing the literature in the field of role mining, this paper summarizes the existing issues and challenges:

- *The selection of accurate and efficient role sets.*
  Since the RBAC system deployment environment is very different and the mining of role sets is recognized as an NP-Hard problem, there is currently no determination algorithm that can derive the most efficient role set according to the corresponding scenarios. Although many researchers have proposed their own heuristic algorithms, they can only be applied to specific scenarios.
- *Find the evaluation of quality criteria for role mining results.*
  There is still lack of uniform and accurate metrics to measure the pros and

cons of role mining results. The evaluation metrics of role mining results summarized in this paper provide references and suggestions for current researchers, and need further improvement and expansion.

- *Reduce the complexity of role mining algorithm.*
  The combination of multi-dimensional technology will lead to an exponential boom in system complexity, which needs to be reduced in the deployment of live scene systems (constrained environments). It is a difficult task to reduce the complexity of role mining while ensuring the maximum efficiency of RBAC system.
- *Dynamic update of the roles.*
  Since the efficient RBAC systems require more accurate and broader set of roles, the role mining algorithms need to be updated constantly. With the updating of RBAC system, the number of roles will explode. So, it is a great challenge to update the roles in time.
- *Role semantic information mining.*
  In many current algorithms and technologies, many roles become unrecognized after many iterations, which poses great difficulties in identifying and understanding character sets. It is important to ensure that role semantic information is highly identifiable during role mining. How to generate a role with accurate semantic information is yet to be further studied.

## 5 Conclusion

This paper summarizes the literature on role mining in RBAC system in the past few years, and compares the performance of role mining mechanism, and summarizes the commonly used metrics to evaluate the quality of role mining results. Through the in-depth study of various role mining mechanisms, we propose the following suggestions and predict the future development direction of role mining.

- The development of future role mining is bound to develop towards the direction of big data [57]. By combining role mining with big data, more accurate role sets can be obtained.
- The role mining will develop from a single system to a comprehensive multi-dimensional system especially the heterogeneous network [58]. The combination of role mining technology with other access control mechanisms or other security technologies will enhance its own security [59,60], reduce system complexity, and enhance the ability to resist attacks.
- Although role mining technology has a lot of research in some areas, it is still blank in many application areas. Taking the Internet of Things as an example, the IoT environment needs to implement access control in a low-power scenario. The identity information of sensor nodes [61] is equivalent to user groups, and some scenes can derive corresponding roles. Mining the identity information of these sensor nodes to establish the mapping relationship with the role can also implement access control in the context of the Internet of Things.

- An assessment framework for comprehensive, accurate, and efficient role mining results needs to be established. Because of the difference of evaluation scenarios, there is no unified evaluation model. In the future, it is possible to explore the establishment and improvement of an evaluation framework applicable to the results of role mining in any case.

**Acknowledgments.** The authors would like to thank the anonymous reviewers for their valuable comments which helped them to improve the content, organization, and presentation of this paper.

## References

1. Yan, W., Mestha, L.K., Abbaszadeh, M.: Attack detection for securing cyber physical systems. IEEE Internet Things J. **6**(5), 8471–8481 (2019)
2. Weinberger, S.: Top ten most-destructive computer viruses, 19 2012 (2012). Smithsonian.com
3. Cybersecurity Unit, Computer Crime & Intellectual Property Section Criminal Division U.S. Department of Justice. A framework for a vulnerability disclosure program for online systems. https://www.justice.gov/criminal-ccips/page/file/983996/download. Accessed 21 May 2019
4. Guan, J., Zhang, Y., Yao, S., Wang, L.: AID shuffling mechanism based on group-buying auction for identifier network security. IEEE Access **7**, 123746–123756 (2019)
5. Lipner, S.B.: The birth and death of the orange book. IEEE Ann. Hist. Comput. **37**(2), 19–31 (2015)
6. Sandhu, R.S., Coyne, E.J., Feinstein, H.L., Youman, C.E.: Role-based access control: a multi-dimensional view. In: Tenth Annual Computer Security Applications Conference, pp. 54–62, December 1994 (1994)
7. Sandhu, R.S., Ferraiolo, D.F., Kuhn, D.R.: The NIST model for role-based access control: towards a unified standard. In: Fifth ACM Workshop on Role-Based Access Control, RBAC 2000, Berlin, Germany, 26–27 July 2000, pp. 47–63 (2000)
8. Servos, D., Osborn, S.L.: Current research and open problems in attribute-based access control. ACM Comput. Surv. **49**(4), 65:1–65:45 (2017)
9. Gritti, C., Önen, M., Molva, R., Susilo, W., Plantard, T.: Device identification and personal data attestation in networks. J. Wirel. Mob. Netw. Ubiquit. Comput. Dependable Appl. (JoWUA) **9**(4), 1–25 (2018)
10. Liu, Y., Quan, W., Wang, T., Wang, Y.: Delay-constrained utility maximization for video ads push in mobile opportunistic D2D networks. IEEE Internet Things J. **5**(5), 4088–4099 (2018)
11. Kotenko, I., Saenko, I., Branitskiy, A.: Applying big data processing and machine learning methods for mobile Internet of Things security monitoring. J. Internet Serv. Inf. Secur. (JISIS) **8**(3), 54–63 (2018)
12. Di Pietro, R., Salleras, X., Signorini, M., Waisbard, E.: A blockchain-based trust system for the Internet of Things. In: Proceedings of the 23rd ACM on Symposium on Access Control Models and Technologies, pp. 77–83. ACM (2018)
13. Liu, Y., Xu, C., Zhan, Y., Liu, Z., Guan, J., Zhang, H.: Incentive mechanism for computation offloading using edge computing: a Stackelberg game approach. Comput. Netw. **129**, 399–409 (2017)

14. Yao, S., Guan, J., Yan, Z., Xu, K.: SI-STIN: a smart identifier framework for space and terrestrial integrated network. IEEE Netw. **33**(1), 8–14 (2018)
15. Moriano, P., Pendleton, J., Rich, S., Camp, L.J.: Stopping the insider at the gates: protecting organizational assets through graph mining. J. Wirel. Mob. Netw. Ubiquit. Comput. Dependable Appl. (JoWUA) **9**(1), 4–29 (2018)
16. Perera, M.N.S., Koshiba, T.: Achieving strong security and member registration for lattice-based group signature scheme with verifier-local revocation. J. Internet Serv. Inf. Secur. (JISIS) **8**(4), 1–15 (2018)
17. Valenza, F., Lioy, A.: User-oriented network security policy specification. J. Internet Serv. Inf. Secur. (JISIS) **8**(2), 33–47 (2018)
18. Aldo, M.S.: Strategic role engineering approach to visual role based access control (V-RBAC). Int. J. Comput. Appl. Eng. Sci. **3**(2), 84 (2013)
19. Narouei, M., Takabi, H.: Towards an automatic top-down role engineering approach using natural language processing techniques. In: Proceedings of the 20th ACM Symposium on Access Control Models and Technologies, pp. 157–160. ACM (2015)
20. Roeckle, H., Schimpf, G., Weidinger, R.: Process-oriented approach for role-finding to implement role-based security administration in a large industrial organization. In: Proceedings of the Fifth ACM Workshop on Role-Based Access Control, pp. 103–110. ACM (2000)
21. Vaidya, J., Atluri, V., Guo, Q.: The role mining problem: finding a minimal descriptive set of roles. In: Proceedings of the 12th ACM symposium on Access Control Models and Technologies, pp. 175–184. ACM (2007)
22. Bertino, E., Bonatti, P.A., Ferrari, E.: TRBAC: a temporal role-based access control model. ACM Trans. Inf. Syst. Secur. **4**(3), 191–233 (2001)
23. Mitra, B., Sural, S., Vaidya, J., Atluri, V.: Migrating from RBAC to temporal RBAC. IET Inf. Secur. **11**(5), 294–300 (2017)
24. Mitra, B., Sural, S., Atluri, V., Vaidya, J.: Toward mining of temporal roles. In: Wang, L., Shafiq, B. (eds.) DBSec 2013. LNCS, vol. 7964, pp. 65–80. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39256-6_5
25. Pan, N., Sun, L., Zhu, Z., He, L.: A temporal approximation-based role mining approach for TRBAC. In: 2017 3rd IEEE International Conference on Computer and Communications (ICCC), pp. 2366–2370. IEEE (2017)
26. Frank, M., Buhman, J.M., Basin, D.: Role mining with probabilistic models. ACM Trans. Inf. Syst. Secur. (TISSEC) **15**(4), 15 (2013)
27. Colantonio, A., Di Pietro, R., Ocello, A., Verde, N.V.: A new role mining framework to elicit business roles and to mitigate enterprise risk. Decis. Support Syst. **50**(4), 715–731 (2011)
28. Colantonio, A., Di Pietro, R., Ocello, A., Verde, N.V.: Visual role mining: a picture is worth a thousand roles. IEEE Trans. Knowl. Data Eng. **24**(6), 1120–1133 (2011)
29. Liu, Y., Wu, H., Xia, Y., Wang, Y., Li, F., Yang, P.: Optimal online data dissemination for resource constrained mobile opportunistic networks. IEEE Trans. Veh. Technol. **66**(6), 5301–5315 (2016)
30. Ma, X., Li, R., Lu, Z.: Role mining based on weights. In: Proceedings of the 15th ACM Symposium on Access Control Models and Technologies, pp. 65–74. ACM (2010)
31. Frank, M., Streich, A.P., Basin, D., Buhmann, J.M.: A probabilistic approach to hybrid role mining. In: Proceedings of the 16th ACM Conference on Computer and Communications Security, pp. 101–111. ACM (2009)
32. Zhai, Z., Wang, J., Cao, Z., Mao, Y.: Hybrid role mining methods with minimal perturbation (in Chinese). J. Comput. Res. Dev. **50**(5), 951–960 (2013)

33. Fuchs, L., Meier, S.: The role mining process model-underlining the need for a comprehensive research perspective. In: 2011 Sixth International Conference on Availability, Reliability and Security, pp. 35–42. IEEE (2011)
34. Molloy, I., Li, N., Qi, Y.A., Lobo, J., Dickens, L.: Mining roles with noisy data. In: Proceedings of the 15th ACM Symposium on Access Control Models and Technologies, pp. 45–54. ACM (2010)
35. Vaidya, J., Atluri, V., Warner, J., Guo, Q.: Role engineering via prioritized subset enumeration. IEEE Trans. Dependable Secure Comput. **7**(3), 300–314 (2008)
36. Takabi, H., Joshi, J.B.D.: StateMiner: an efficient similarity-based approach for optimal mining of role hierarchy. In: Proceedings of the 15th ACM Symposium on Access Control Models and Technologies, pp. 55–64. ACM (2010)
37. Colantonio, A., Di Pietro, R., Ocello, A., Verde, N.V.: Taming role mining complexity in RBAC. Comput. Secur. **29**(5), 548–564 (2010)
38. Molloy, I., et al.: Mining roles with multiple objectives. ACM Trans. Inf. Syst. Secur. (TISSEC) **13**(4), 36 (2010)
39. Guo, Q., Vaidya, J., Atluri, V.: The role hierarchy mining problem: discovery of optimal role hierarchies. In: 2008 Annual Computer Security Applications Conference (ACSAC), pp. 237–246. IEEE (2008)
40. Saenko, I., Kotenko, I.: Genetic algorithms for role mining problem. In: 2011 19th International Euromicro Conference on Parallel, Distributed and Network-Based Processing, pp. 646–650. IEEE (2011)
41. Dong, L., Wu, K., Tang, G.: A data-centric approach to quality estimation of role mining results. IEEE Trans. Inf. Forensics Secur. **11**(12), 2678–2692 (2016)
42. Molloy, I., Li, N., Li, T., Mao, Z., Wang, Q., Lobo, J.: Evaluating role mining algorithms. In: Proceedings of the 14th ACM Symposium on Access Control Models and Technologies, pp. 95–104. ACM (2009)
43. Wu, L., et al.: Uniform-scale assessment of role minimization in bipartite networks and its application to access control. Phys. A: Stat. Mech. Applications. **507**, 381–397 (2018)
44. Colantonio, A., Di Pietro, R., Ocello, A., Verde, N.V.: A probabilistic bound on the basic role mining problem and its applications. In: Gritzalis, D., Lopez, J. (eds.) SEC 2009. IFIPAICT, vol. 297, pp. 376–386. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-01244-0_33
45. Blundo, C., Cimato, S.: A simple role mining algorithm. In: Proceedings of the 2010 ACM Symposium on Applied Computing, pp. 1958–1962. ACM (2010)
46. Huang, H., Shang, F., Zhang, J.: Approximation algorithms for minimizing the number of roles and administrative assignments in RBAC. In: 2012 IEEE 36th Annual Computer Software and Applications Conference Workshops, pp. 427–432. IEEE (2012)
47. Xu, Z., Stoller, S.D.: Mining parameterized role-based policies. In: Proceedings of the Third ACM Conference on Data and Application Security and Privacy, pp. 255–266. ACM (2013)
48. Molloy, I., et al.: Mining roles with semantic meanings. In: Proceedings of the 13th ACM Symposium on Access Control Models and Technologies, pp. 21–30. ACM (2008)
49. Ye, W., Li, R., Gu, X., Li, Y., Wen, K.: Role mining using answer set programming. Future Gener. Comput. Syst. **55**, 336–343 (2016)
50. Pan, N., Sun, L., He, L.-S., Zhu, Z.-Q.: An approach for hierarchical RBAC reconfiguration with minimal perturbation. IEEE Access **6**, 40389–40399 (2017)
51. Mitra, B., Sural, S., Vaidya, J., Atluri, V.: A survey of role mining. ACM Comput. Surv. **48**, 1–37 (2016)

52. Rosen-Zvi, M., Chemudugunta, C., Griffiths, T., Smyth, P., Steyvers, M.: Learning author-topic models from text corpora. ACM Trans. Inf. Syst. (TOIS) **28**(1), 4 (2010)
53. Molloy, I., Park, Y., Chari, S.: Generative models for access control policies: applications to role mining over logs with attribution. In: Proceedings of the 17th ACM Symposium on Access Control Models and Technologies, pp. 45–56. ACM (2012)
54. Zhang, X., Han, W., Fang, Z., Yin, Y., Mustafa, H.: Role mining algorithm evaluation and improvement in large volume android applications. In: Proceedings of the First International Workshop on Security in Embedded Systems and Smartphones, pp. 19–26. ACM (2013)
55. Alohaly, M., Takabi, H., Blanco, E.: A deep learning approach for extracting attributes of ABAC policies. In: Proceedings of the 23rd ACM on Symposium on Access Control Models and Technologies, pp. 137–148. ACM (2018)
56. Morisset, C., Willemse, T.A.C., Zannone, N.: Efficient extended ABAC evaluation. In: Proceedings of the 23rd ACM on Symposium on Access Control Models and Technologies, pp. 149–160. ACM (2018)
57. Colombo, P., Ferrari, E.: Access control in the era of big data: state of the art and research directions. In: Proceedings of the 23rd ACM on Symposium on Access Control Models and Technologies, SACMAT 2018, Indianapolis, IN, USA, 13–15 June 2018, pp. 185–192 (2018)
58. Guan, J., Sharma, V., You, I., Atiquzzaman, M., Imran, M.: Extension of MIH for FPMIPv6 (EMIH-FPMIPv6) to support optimized heterogeneous handover. Future Gener. Comp. Syst. **97**, 775–791 (2019)
59. Squicciarini, A.C., Rajtmajer, S.M., Zannone, N.: Multi-party access control: requirements, state of the art and open challenges. In: Proceedings of the 23rd ACM on Symposium on Access Control Models and Technologies, SACMAT 2018, Indianapolis, IN, USA, 13–15 June 2018, p. 49 (2018)
60. Liu, B., Guan, J., Jiang, Z.: A policy management system based on multi-dimensional attribution label. In: You, I., Leu, F.-Y., Chen, H.-C., Kotenko, I. (eds.) MobiSec 2016. CCIS, vol. 797, pp. 128–142. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7850-7_12
61. Lee, A.J., Biehl, J.T., Curry, C.: Sensing or watching?: balancing utility and privacy in sensing systems via collection and enforcement mechanisms. In: Proceedings of the 23rd ACM on Symposium on Access Control Models and Technologies, SACMAT 2018, Indianapolis, IN, USA, 13–15 June 2018, pp. 105–116 (2018)