



# A new role mining framework to elicit business roles and to mitigate enterprise risk

Alessandro Colantonio<sup>a,b,\*</sup>, Roberto Di Pietro<sup>b</sup>, Alberto Ocello<sup>a</sup>, Nino Vincenzo Verde<sup>b</sup>

<sup>a</sup> Engiweb Security, Roma, Italy

<sup>b</sup> Università di Roma Tre, Dipartimento di Matematica, Roma, Italy

## ARTICLE INFO

Available online 19 August 2010

### Keywords:

RBAC  
Role engineering  
Role mining  
Risk management  
Clustering coefficient

## ABSTRACT

*Role-based access control* (RBAC) allows to effectively manage the risk derived from granting access to resources, provided that designed roles are business-driven. *Role mining* represents an essential tool for role engineers, but existing techniques are not able to elicit roles with an associated clear business meaning. Hence, it is difficult to mitigate risk, to simplify business governance, and to ensure compliance throughout the enterprise. To elicit meaningful roles, we propose a methodology where data to analyze are decomposed into smaller subsets according to the provided business information. We introduce two indices, *minability* and *similarity*, that drive the decomposition process by providing the expected complexity to find roles with business meaning. The proposed methodology is rooted on a sound theoretical framework. Moreover, experiments on real enterprise data support its effectiveness.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

*Access control* is a cornerstone of enterprise risk and security management. It represents the process of mediating requests to data and services maintained by a system, and determining whether the requests should be granted or denied [11]. It is the responsibility of an access control system to ensure that only users with legitimate credentials are granted permissions to access requested resources. Hence, in an access control model the risk factor of illegitimate credentials is eliminated by construction [2]. Significant research has focused on providing formal representations of access control models. Among all models proposed in the literature, *Role-Based Access Control* (RBAC) [1] is certainly the most adopted by medium- to large-size organizations, greatly due to its simplicity: a role can be seen as a set of permissions; users, in turn, are assigned to appropriate roles based on their responsibilities and qualifications. As a result, RBAC offers great benefits to business users. A role represents a job function or a title established for a set of users within an organization. Thus, the adoption of RBAC makes it easier to define security policies by business users [16]. RBAC also implements the appropriate security engineering principles to enforce risk reduction, such as separation of duties (SoD) and least privilege [2]. Further, the use of roles minimizes

system administration effort due to the reduced number of relationships required to relate users to permissions [5].

Despite the benefits related to deploying role-based access control systems, many organizations are reluctant to adopt them, since there are still some important issues that need to be addressed. In particular, the model must be customized to capture the needs and functions of the organization. In an ideal RBAC environment, we expect roles to be well defined so that role definitions are formed with strict role boundary rules in order to enforce all the required enterprise security policies. Unfortunately, where RBAC is deployed, this rarely happens, thus leading to role misuse [2]. For this purpose, the *role engineering* discipline [9] has been introduced. However, choosing the best way to design a proper set of roles is still an open problem. Various approaches to role engineering have been proposed, which are usually classified as: *top-down* and *bottom-up*. The former requires a deep analysis of business processes to identify which access permissions are necessary to carry out specific tasks. The latter seeks to identify de facto roles embedded in existing access control information. Since bottom-up approaches usually resort to data mining techniques, the term *role mining* is often used as a synonym for bottom-up. To maximize benefits, bottom-up should be used in conjunction with top-down, leading to an *hybrid* approach. As a matter of fact, top-down may ignore existing permissions and exceptions, whereas bottom-up may not consider the business functions of an organization [19].

The bottom-up approach has attracted researchers, since it can be easily automated [22]. Indeed, companies which plan to go for RBAC usually find themselves in the situation of having a collection of several legacy and standard security systems on different platforms that provide “conventional” access control [20]. Thus, role mining is the application of data mining techniques to generate roles from the

\* Corresponding author.

E-mail addresses: [alessandro.colantonio@eng.it](mailto:alessandro.colantonio@eng.it), [colanton@mat.uniroma3.it](mailto:colanton@mat.uniroma3.it) (A. Colantonio), [dipietro@mat.uniroma3.it](mailto:dipietro@mat.uniroma3.it) (R. Di Pietro), [alberto.ocello@eng.it](mailto:alberto.ocello@eng.it) (A. Ocello), [nverde@mat.uniroma3.it](mailto:nverde@mat.uniroma3.it) (N.V. Verde).

URLs: <http://ricerca.mat.uniroma3.it/users/colanton/> (A. Colantonio), <http://ricerca.mat.uniroma3.it/users/dipietro/> (R. Di Pietro), <http://ricerca.mat.uniroma3.it/users/nverde/> (N.V. Verde).

access control information of this collection of systems. Several works prove that the role mining problem is reducible to many other well-known NP-hard problems, such as clique partition, binary matrix factorization, bi-clustering, graph vertex coloring [6,8,32] to cite a few. However, on one hand the slavish application of standard data mining approaches to role engineering might yield roles that are merely a set of permissions, namely with no connection to the business practices. On the other hand organizations are unwilling to deploy roles they cannot bind to a business meaning [5]. Indeed, such roles could have some difficulties in being inserted within the risk management framework in use within the organization. In such a case, risks are incurred to the system by users that are authorized to use their access right in an incorrect manner [2]. Moreover, when hundreds of thousands of existing user–permission assignments need to be analyzed, the number of candidate roles might be so high that trying to assign a business meaning to each of them is often impracticable. The number of candidate roles may also grow because of the “noise” within the data—namely, permissions exceptionally or accidentally granted or denied. In such a case, classical role mining algorithms discover multiple small fragments of the true role, but missing the role itself [7]. This increases the risk of designing roles that do not capture the actual business needs of the organization.

Only a few recent works value business requirements in role mining [3,5] by proposing a measure for the business meaning of roles. However, it is difficult to introduce this metric in existing role mining approaches currently found in the literature. An alternative way of leveraging business-related information to offer meaningful candidate role-sets may be by restricting the analysis to sets of data that are homogeneous from an enterprise perspective. For instance, let us suppose that a partial or coarse-grained top-down analysis identifies a certain set of users that perform the same tasks, but the analysis lacks the knowledge of which permissions are required for the execution of these tasks. In this scenario, by restricting role mining techniques to these users only—instead of analyzing the organization as a whole—, the elicited roles will be related to such tasks. Thus, it will likely be easier to assign a business meaning to the results obtained from the bottom-up approach. Moreover, by grouping users that perform similar tasks together first, and then analyzing them separately, eliciting roles with no business meaning can be avoided. Indeed, investigating analogies among groups of users that perform completely different tasks is far from being a good role mining strategy [5]. Further, it will be easier to manage resulting candidates roles, achieving two results: a simplification of the security policy enforcement process; and, a reduction of the risk related to unintentional/incorrect use of granted permissions through roles. Partitioning data also introduces benefits in terms of execution time of role mining algorithms. Indeed, most role mining algorithms have a complexity that is not linear with respect to the number of users or permissions to analyze [3,13,21,30]. Based on previous observations, several enterprise information may be used to decompose the role mining problem. Business processes, workflow tasks, and organization unit trees are just a few examples of business elements that can be leveraged. Usually, such information is already available in most companies before starting the role engineering task. However, when dealing with information from several sources, a few decisions have to be made about: what information can actually improve the role mining process; what level of detail is required; and, lastly, how to verify that each sub-problem is easily solvable using a data mining algorithm.

To address all the abovementioned issues, this paper proposes a methodology that helps role engineers leverage business information during the role mining process. In particular, we propose to divide the access data to analyze into smaller subsets that are homogeneous according to some business data, instead of performing a single bottom-up analysis on the entire organization. This eases the attribution of business meaning to automatically elicited roles and reduces the problem complexity, thus allowing for better enforce-

ment of security policies and reducing the risk related to illegal accesses. In order to select the best business information that improves the subsequent role mining process, as well as to establish how deeply the data must be partitioned, two indices, referred to as *minability* and *similarity*, are identified. Minability and similarity are both rooted on sound mathematical theory. These indices are used to measure the expected complexity of analyzing the outcome of the bottom-up approaches. Leveraging these indices allows for the identification of business information that best fits with the access control data, namely the information that induces a decomposition which increases the business meaning of the roles elicited in the role mining phase and, at the same time, simplifies the analysis. This leads to a decrease in the likelihood of making errors in role management, and consequently reduces the risk of role misuse. The paper also introduces two fast probabilistic algorithms to efficiently compute such indices, making them suitable also for big organization with hundreds of thousands of users and permissions. The quality of the indices is also formally assured. Several examples illustrate the practical implications of the proposed methodology and related tools, which have also been applied on real enterprise data. Results support the quality and viability of the proposal.

The paper is organized as follows: [Section 2](#) introduces the background required to formally describe the proposed tools and reports on related works. The adopted risk model is then described in [Section 3](#), that also maps typical risk-related concepts to RBAC entities. Minability and similarity indices are introduced and further discussed with simple examples in [Section 4](#). The proposed methodology, which is mainly based on these two indices, is proposed in [Section 5](#). In [Section 6](#) two efficient probabilistic algorithms are proposed: one to calculate similarity, the other one to calculate the minability index. Then, the viability of the proposed applications is demonstrated in [Section 7](#), showing the results of a test on real data. Finally, [Section 8](#) provides concluding remarks.

## 2. Background and related work

### 2.1. Role engineering

Before introducing the required formalism used to describe role engineering, we first review some concepts of the ANSI/INCITS RBAC standard [1] needed for the present analysis. For the sake of simplicity, we do not consider sessions, role hierarchies or separation of duties constraints in this paper. In particular, we are only interested in the following entities:

- *PERMS*, *USERS*, and *ROLES* are the sets of all access permissions, users, and roles, respectively;
- $UA \subseteq USERS \times ROLES$ , is the set of all role–user relationships;
- $PA \subseteq PERMS \times ROLES$ , is the set of all role–permission relationships.

The following functions are also provided:

- $ass\_users: ROLES \rightarrow 2^{USERS}$  to identify users assigned to a role. We consider it as derived from *UA*, that is  $ass\_users(r) = \{u \in USERS \mid \langle u, r \rangle \in UA\}$ ;
- $ass\_perms: ROLES \rightarrow 2^{PERMS}$  to identify permissions assigned to a role. We consider it as derived from *PA*, that is  $ass\_perms(r) = \{p \in PERMS \mid \langle p, r \rangle \in PA\}$ .

In addition to RBAC concepts, this paper introduces other entities required to formally describe the proposed approach. In particular, we define:

- $UP \subseteq USERS \times PERMS$ , the set of the existing user–permission assignments to be analyzed;
- $perms: USERS \rightarrow 2^{PERMS}$ , the function that identifies permissions assigned to a user. Given  $u \in USERS$ , it is defined as  $perms(u) = \{p \in PERMS \mid \langle u, p \rangle \in UP\}$ ;

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات