An Efficient Similarity-Based Approach for Optimal Mining of Role Hierarchy

Hassan Takabi and James B D Joshi
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA, USA
{hatakabi, jjoshi}@sis.pitt.edu

ABSTRACT

In order to deploy role-based access control (RBAC) system, one requires to first identify a complete, correct and efficient set of roles. This process, known as role engineering, has been identified as one of the costliest tasks in migrating to RBAC. Several approaches have been proposed that mostly use data mining techniques to discover roles. However, most of them do not consider the existing roles and try to define all roles from scratch which is not acceptable for organizations that have an RBAC system in place. In this paper, we define the problem of mining role hierarchy with minimal perturbation. In order to do this, we define a measure for goodness of an RBAC state and another measure for minimal perturbation. Our proposed algorithm presents a heuristic solution to find an RBAC state as similar as possible to both the existing state and the optimal state.

1. INTRODUCTION

In order to deploy role-based access control (RBAC), an organization must first identify a set of roles that is complete, correct and efficient, and then assign users and permissions to these roles. This process is known as *role engineering*, is the costliest component of RBAC implementation. There are two general approaches to accomplish the task of role engineering and construct an RBAC system: the *top-down approach* uses a detailed analysis of business processes, defines particular job functions, decomposes them into smaller units, and finally creates roles for these units by associating needed permissions. The *bottom-up approach* called *role mining* uses data mining techniques to discover roles from existing system configuration data in particular the permission assignments.

There have been several attempts to propose bottom-up approaches to mining roles [1, 2, 3, 4, 6]. However, there is no formal notion of goodness of a produced role set. Without considering semantic meanings, minimality may serve as a best approximation for discovering good *descriptive* roles, but generally it is not a good measure for goodness of discovered roles. Moreover, the proposed role mining techniques do not consider the existing RBAC configuration and try to define all roles from scratch. These approaches may not be acceptable for organizations that have an RBAC system in place [6]. Note that migrating to RBAC is not a one time process. Once an RBAC system is implemented and in place, maintenance of the system becomes an important issue [6]. When the initial RBAC configuration becomes messy and inefficient as a result of being used for a long time, many changes and updates, it is not a good idea to adopt a completely different RBAC system and redefine the system from scratch. Moreover, changes to the existing role set may cause disruptions to the organization's operations. Migrating to a new set of roles from the existing set of roles should cause as little disruption as possible. So, the goal is to look for a set of roles as close as possible to both the existing set of roles and the optimal set of roles.

To the best of our knowledge, there is not much in the literature to address this issue. In this paper we propose an approach that takes role migration cost and the existing RBAC state into account and identifies an RBAC state that is as close as possible to both existing RBAC state and the optimal state. In order to do this, we use the theory of formal concept analysis [7] which has been shown to provide a strong theoretical foundation for role engineering [6]. We introduce two different measures: one for optimality (i.e., structural complexity) of an RBAC state, and another for minimal perturbation (i.e., similarity of sets of roles). Our proposed algorithm presents a heuristic solution to find an RBAC state with the smallest structural complexity and as similar as possible to both the existing state and the optimal state. Our contributions in this paper are as follows: We define the problem of mining role hierarchy with minimal perturbation. In order to do this, we introduce new measures for optimality and minimal perturbation and present a heuristic algorithm to find an RBAC state as similar as possible to the existing state and the optimal state. We also present evaluations that indicate the effectiveness of the algorithm. Finally, we discuss future research directions.

The remainder of this paper is organized as follows: In Section 2, we present a brief description of the formal concept analysis and its relationship with RBAC state, whereas in Section 3 we describe the problem of mining role hierarchy with minimal perturbation. Section 4 presents our algorithm and its evaluation. Section 5 gives a direction for research in role mining. Section 6 discusses the related work. Finally, Section 7 concludes the paper.

2. THE THEORY OF FORMAL CONCEPT ANALYSIS AND RBAC STATE

In formal concept analysis, the family of concepts that complies the mathematical axioms defines a lattice, and is called a *concept lattice*. In concept lattice, each concept inherits all permissions associated with its subconcepts, and users are inherited in the other direction. Therefore, we can remove redundant permissions and users from each node. The result is called the *reduced concept lattice*. The reduced concept lattice defines a complete RBAC state. Each concept represents a role and the lattice can be viewed as the role hierarchy. It is clear that the reduced concept lattice provides the semantic relationships among concepts and has more meanings than just a set of permissions. Using the reduced concept lattice as the role hierarchy has the disadvantage that the role hierarchy may be extremely large. In the reduced concept lattice, some concepts introduce no new users, some concepts introduce no new permissions, and some concepts introduce neither new users nor new permissions. However, it is not correct to remove all concepts with no new users or new permissions. We need to have a measure to compare the different role hierarchies generated from the reduced concept lattice and identify which one is more desirable.

3. THE PROBLEM OF MINING ROLE HIERARCHY WITH MINIMAL PERTURBATION

In this section, we start by defining two different measures: one to measure goodness of an RBAC state and another one to measure perturbation. Then, we formally define the problem of mining role hierarchy with minimal perturbation.

A Measure for Goodness of an RBAC State

Given an access control configuration, many RBAC states may be consistent with it. There has to be a measurement of how good an RBAC state is in order to select among them. Several metrics have been defined in the literature to measure goodness of identified roles such as minimizing the number of roles or minimizing the administration cost of the resulting RBAC model and weighted structural complexity [2, 3, 6, 4]. Considering all those metrics, we believe the weighted structural complexity is the most general and most flexible measure that covers other measures as well. In order to have a measure, we adopt the notion of the *weighted structural complexity* as a measurement of goodness of an RBAC state. The key difference between the measure that we use and the one used by *Molloy et al.* [6] is that their measure allows direct user-permissions assignment relation ($DUPA$) in an RBAC state while we do not allow it and do not include it in the weighted structural complexity, because allowing direct user-permissions assignment defeats the purpose of RBAC.

A Measure for Minimal Perturbation

Vaidya et al. [5] use Jaccard Coefficient to measure similarity between roles and role sets. Although Jaccard Coefficient is quiet straightforward, in general, it is too simple to measure similarity of sets of roles. Moreover, they only consider permissions assigned to roles and ignores the users. We define a flexible and general measure for similarity between roles and role sets that takes into account users and permissions associated with roles as well as relations in role hierarchy with adjustable weights. First, we define a similarity measure between two roles and then extend it to two role sets.

For any two roles r_1 and r_2, first we define three different similarity measures based on their permissions, users, and hierarchy relation. Then, by combining all these measures with adjustable weights, we define the Role-Role similarity measure which is called $sim(r_1, r_2)$. Compared to similarity between roles, measuring the similarity between sets of roles is a significantly more complex task. It is not clear whether a single role corresponds to only one other role or to a set of roles. We extend prior measure to measure similarity between two sets of roles; we compute similarity for each pair of roles in both role sets, sort them and then pick the maximum similarities of pairs of roles and take their average. A key issue is deciding how many of these similarities should be taken into account specially when two role sets have different numbers of roles. Ideally, we want every role to contribute to the final similarity measure. Wa propose following approach to measure similarity between two role sets rs_1 and rs_2.

Role Set-Role Set Similarity. *For any two role sets rs_1 and rs_2 where rs_1 is the smaller role set, we compute similarity between them as follows:*
$\forall r_i \in rs_1$, find $Max_{r_j \in rs_2} sim(r_i, r_j)$ such that for all selected pairs (r_i, r_j) and (r_x, r_y), if $r_i \neq r_x$ then $r_j \neq r_y$. In this step every role in rs_1 is matched with exactly one distinct role in rs_2, but there are some roles in rs_2 that have not been matched with any role from rs_1. For those roles in rs_2 that have not been matched in the first step, we define a threshold t and consider only roles that have a similarity measure above the threshold. Finally, we take average over all of chosen similarities.
Also, the dissimilarity between two role sets is defined as follows: $dissim(rs_1, rs_2) = 1 - sim(rs_1, rs_2)$

In order to combine the two defined measures, we define a global optimization function that minimizes the weighted structural complexity of the resulting RBAC state and maximizes the similarity (minimize the dissimilarity) between identified roles and the existing roles. Using all of the prior definitions, we next define the problem of mining role hierarchy with minimal perturbation as follows:
The problem of Mining Role Hierarchy with Minimal Perturbation. The goal of the problem of mining role hierarchy with minimal perturbation is to minimize the global optimization function of the predefined optimality measure (i.e., weighted structural complexity) and the predefined perturbation measure (i.e., dissimilarity measure).

4. THE PROPOSED ALGORITHM

Our heuristic algorithm consists of two phases. In the first phase, we generate the reduced concept lattice using the deployed configuration, which gives us an RBAC state. In the second phase, we prune this lattice and select the final RBAC state. Once we have the reduced concept lattice, we should decide which roles are appropriate and which ones should be removed. Our proposed algorithm is a greedy algorithm; it iterates over all of the roles in the reduced concept lattice and perform pruning if the change will decrease combination function of weighted structural complexity of the RBAC state and the distance between the deployed role set and the role sets with and without that role. The algorithm stops when no more operation can be performed. Removing each role reduces the cost of creating the role and the associated relationships. However, we need to add back some relationships so that user-permission assignment relation and the inheritance relation remain correct. Considering the combination function we defined, we remove role r from the reduced concept lattice when the value of the combination function decreases after removing that role.

We have implemented the proposed algorithm. Due to space limit we can not show the figures. From the results, we observe that our algorithm generates significantly fewer roles than the original state. Our algorithm has smaller weighted structural complexity than the HierarchicalMiner and is closer to the optimal solution. We also observe that compared to VAG algorithm, our approach provides better results and the mined roles of our proposed similarity measure are closer to the original state.

5. CHALLENGES AND DIRECTIONS

In this section, we discuss some possible research directions in the area of role mining.

Parameterized roles. Another problem is to extract parameterized roles that correspond to categories of concepts. For example, we may create a role for students with the student id as a parameter. Using parameterized roles could dramatically reduce the number of roles and consequently the administration cost. In some situations, permissions are parameterized. For example, permission about a file, permission about a directory, permission about a database, etc. We can use this information especially in combination with other useful information to discover parameterized roles.

Separation of Duty Constraints. Another issue is to consider separation of duty constraints in role mining process. There could be some currently specified separation of duty constraints in the system; it is not clear what effects the migration will have on these constraints. So, we need to consider this aspect while aiming to minimize the perturbation.

Role mining in multi-domain environments. Another thread of research is to use role mining for integrating different access control policies in multi-domain environments. One way to address secure interoperation in multi-domain environments is centralized approach that maps each local policy into one global policy. In traditional centralized approach, the global policy is usually specified by administrators who know the functions of all the cooperating organizations. Such an approach may not be realistic in some situations due to its cost, dynamic access patterns, and difficulty to find such administrators that are familiar with all domains. We can adopt the idea of role mining to come up with an approach that automatically defines the global policy. Here, we have multiple domains each of which has its own RBAC system. Users acquire different roles from different domains based on their needs. We can utilize these RBAC systems' configurations from different domains to define the global policy. So, we observe and get the user's access history and use the same idea as role mining to create the interoperation links through the user access patterns.

6. RELATED WORK

Zhang et al. have presented a heuristic algorithm for role mining, which models an RBAC state as a graph with relationships as its edges [3]. The goal is to minimize the number of edges while maintaining the same connectivity. *Gue et al.* have proposed RH-Builder algorithm, a heuristic solution for role hierarchy construction out of the existing role set such that the number of direct relationships is minimized [4].

Molloy et al. have proposed HierarchicalMiner algorithm, which considers both the semantics of roles and system complexity [6]. However, it does not consider the existing RBAC state and defines everything from scratch. *Vaidya et al.* have defined the Minimal Perturbation problem [5]. They use a similarity metric based on Jaccard coefficient to formalize the problem and propose a heuristic solution based on the previously developed FastMiner algorithm [2]. However, their approach considers only flat roles and ignores the role hierarchy; also the measure it uses to formulate similarity is very simple and not realistic.

7. CONCLUSION AND FUTURE WORK

In this paper, we have defined the problem of mining role hierarchy with minimal perturbation. We have also defined two measures: a measure for goodness of an RBAC state and another measure for minimal perturbation, then based on these measures developed a heuristic solution to find an RBAC state that is as close as possible to both deployed RBAC state and the optimal state. Our experiments demonstrated the effectiveness of the proposed approach. Furthermore, we have discussed some potential future research directions in role mining.

8. REFERENCES

[1] J. Schlegelmilch and U. Steffens, "Role mining with ORCA", In Proc. ACM Symposium on Access Control Models and Technologies (SACMAT), pages 168-176, USA, 2005.

[2] J. Vaidya, V. Atluri, and Q. Guo, "The role mining problem: Finding a minimal descriptive set of roles", In Proc. ACM Symposium on Access Control Models and Technologies (SACMAT), USA, 2007.

[3] D. Zhang, K. Ramamohanarao, and T. Ebringer, "Role engineering using graph optimisation", In Proc. ACM Symposium on Access Control Models and Technologies (SACMAT), pages 139-144, 2007.

[4] Q. Guo, J. Vaidya, and V. Atluri, "The Role Hierachry Mining Problem: Discovery of Optimal Role Hierarchies", In Proc. 2008 Annual Computer Security Applications Conference, pages 237-246, 2008.

[5] J. Vaidya, V. Atluri, and Q. Guo, "Migrating to Optimal RBAC with Minimal Perturbation", In Proc. ACM Symposium on Access Control Models and Technologies (SACMAT), pages 11-20, 2008.

[6] I. Molloy, H. Chen, T. Li, Q. Wang, N. Li, E. Bertino, S. Calo, and J. Lobo, "Mining Roles with Semantic Meanings", In Proc. ACM Symposium on Access Control Models and Technologies (SACMAT), pages 21-30, 2008.

[7] B. Ganter and R. Wille, "Formal Concept Analysis: Mathematical Foundations", Springer, 1998.