Hindawi
Security and Communication Networks
Volume 2019, Article ID 8085303, 15 pages
https://doi.org/10.1155/2019/8085303

Research Article

RMMDI: A Novel Framework for Role Mining Based on the Multi-Domain Information

Wei Bai, Zhisong Pan, Shize Guo, and Zhe Chen

Command & Control Engineering College, Army Engineering University of PLA, Nanjing 210014, China

Correspondence should be addressed to Zhisong Pan; hotpzs@hotmail.com

Received 20 November 2018; Revised 14 March 2019; Accepted 24 April 2019; Published 11 June 2019

Guest Editor: Rohit Ranchal

Role-based access control (RBAC) is widely adopted in network security management, and role mining technology has been extensively used to automatically generate user roles from datasets in a bottom-up way. However, almost all role mining methods discover the user roles from existing user-permission assignments, which neglect the dependency relationships between user permissions. To extend the ability of role mining technology, this paper proposes a novel role mining framework based on multi-domain information. The framework estimates the similarity between different permissions based on the fundamental information in the physical, network, and digital domains and attaches interdependent permissions to the same role. Three simulated network scenarios with different multi-domain configurations are used to validate the effectiveness of our method. The experimental results show that the method can not only capture the interdependent relationships between permissions, but also detect user roles and permissions more reasonably.

1. Introduction

Access control is a fundamental concern in network security management. Role-based access control (RBAC) has become the dominant model for both commercial and research fields [1, 2]. The key point of RBAC is to determine proper roles to capture business needs, which is named as role engineering. There are mainly two kinds of approaches to find user roles: top-down and bottom-up. The top-down approaches always perform a deep analysis of business processes and identify user roles manually [3], while the bottom-up approaches always discover the user roles from existing datasets automatically, which are also named as role mining as they usually resort to data mining techniques [4, 5].

Existing role mining approaches mainly discover a proper user-role assignment relation $UA \subseteq USERS \times ROLES$ and a proper role-permission assignment relation $PA \subseteq ROLES \times PERMS$ from an existing user-permission assignment relation $UPA \subseteq USERS \times PERMS$. In the process, user-permission assignments are considered to be independent. However, considering a typical service authorization process, users are authorized by multiple policy control points, including gate machines, firewalls, or identity authentication systems. Those systems are always configured separately and may grant users with more permissions than they deserve. For example, users, who are authorized to enter certain space, may have the opportunity to use the terminals belong to other users in the same space; users can connect the server behind the firewall remotely to bypass the access control lists and access unauthorized services; users can use the assigned passwords to crack similar passwords for other unauthorized services, etc. In a word, if the interdependent relationships are not taken into consideration, users with certain roles would get extra permissions, introducing security vulnerabilities into network systems.

To address the above-mentioned issues, this paper proposes a novel role mining framework named as RMMDI from the perspective of network security management. Instead of mining user roles from user-permission assignments, the framework discovers user roles from the fundamental information in multiple domains, including the physical domain, network domain, and digital domain. The framework is aimed at outputting a flat RBAC state that divides user permissions into several disjoint subsets. The user permissions in one set tend to be interdependent while the permissions in different sets tend to be independent. If a permission

set is assigned to a user role, a user assigned some roles is unlikely to get extra permissions assigned to other roles. As such, potential security risks involved in the user-permission assignments process can be avoided.

The rest of this paper is organized as follows. In Section 2, some general works are briefly reviewed. Section 3 presents the proposed framework in detail. Section 4 shows the experimental setup and results, and Section 5 presents a comprehensive discussion. At last, Section 6 provides concluding remarks.

2. Related Work

2.1. Role Mining. RBAC has become a dominating model for access control in network security. Instead of assigning permissions to the user directly, RBAC introduces the concept of roles to make access control system more compact and comprehensive [6]. A role is defined as a collection of permissions. The key point of RBAC is to generate proper roles. In this process, the bottom-up approach named as role mining gets much more attention than the top-down approach as the latter is time-consuming and human-intensive [3].

Kuhlmann et al. first proposed the concept of role mining for finding roles from user-permission assignment data [7]. Traditional role mining approaches are mainly divided into two classes based on their output [5, 8, 9]. The first class is to output a prioritized list of candidate roles, each of which is assigned a priority value. A larger priority value means the role is more important or useful. Complete Miner (CM) and Fast Miner (FM) are two typical algorithms of the first class, which identify overlapping clusters by analyzing the subset enumeration in an unsupervised way [10]. The second class is to output a complete RBAC state under a certain cost. There are also a lot of classic algorithms in the class, for example, OFFIS Role mining tool with Cluster Analysis (ORCA) [11], Hierarchical Miner (HM) [12], Graph Optimization (GO) [13], HP Role Minimization (HPr) [14], and HP Edge Minimization (HPe) [14].

Besides those traditional role mining algorithms, there are also many important approaches that emerged in recent years. For example, Frank et al. proposed a probabilistic approach to improve the role mining process by taking account of the business information. The approach utilized the similarity between user-permission relations to detect exceptional assignments and wrong assignments [15]. Besides, entropy-based methods were used in this approach to analyze the impact of business knowledge on role mining [16]. Alessandro et al. presented an approach that allowed role engineers to leverage business information. In the role mining process, the access data was divided into smaller subsets from a business perspective firstly and then traditional methods can be used to discover roles with business meanings [5]. Iran et al. proposed a method based on formal concept lattices to discover roles with semantic meanings [12] as well as a method based on logistic PCA (Principal Component Analysis) to eliminate data noises [17]. Du X and Change X proposed two algorithms based on artificial intelligence, i.e., the genetic algorithm and ant colony optimization algorithm [18]. Dong et al. proposed both fast exact and heuristic methods based on biclique network cover to minimize role number or edge number [19].

With regard to goodness measure, several metrics have been proposed in the literature, including minimizing the number of roles [10, 20], minimizing the number of edges [13, 14, 19], minimizing the number of user-role assignment and permission-role assignment relations [13], minimizing both the number of roles and edges [21], and minimizing the administrative cost [22]. These optimization goals can be uniformly represented by the Weighted Structural Complexity (WSC) [8, 9].

Although there are a lot of effective role mining approaches, most of them neglect the relationships between user permissions. From the perspective of network security management, user permissions are not independent. A user or potential attacker may get extra permissions from the preassigned permissions, which may introduce fatal risks to network security. Hence, in the framework RMMDI, we model the interrelationships between user permissions from multi-domain configuration information and get more reasonable user roles, mitigating the vulnerabilities and strengthening network security.

2.2. Multi-Domain Information Modeling. Traditional network security analysis mainly concentrates on the network domain, with a few concerns on other domains. However, with the deepening of research on insider threat, an increasing number of studies have shown that the attacker will attack the network not only in digital ways, but also through the physical domain and social domain.

The existing methods of joint modeling of network multi-domain information mainly define multi-domain information by using the formalized methods and then make inference based on the logical rules to judge whether the system can reach the unsafe state. Probst et al. proposed a formal model for describing scenarios that span the physical and digital domain [23, 24]. Kotenko et al. proposed a model for describing attacks that use social engineering and physical access based on the preconditions and postconditions of atomic actions [25]. Scott et al. built a security model that adds a spatial relationship between the elements in the ambient calculus [26]. Dimkov presented a security model named as Portunes graph to abstract the environment of an organization into a stratified graph, which involved the information in physical, digital, and social domain information [27]. Kammuller and Probst combined formal modeling and analysis of infrastructures of organizations with a sociological explanation to provide a framework for insider threat analysis [28].

In this paper, we take possible interaction effects among multi-domain permissions into consideration, which are the basis of similar permission finding and role mining based on multiple domain information.

2.3. Multi-View Community Detection. The community is a universal property in many complex networks, which means that network nodes can be divided into small groups [26]. Traditional community detection methods only utilize single network information. And several multiview community

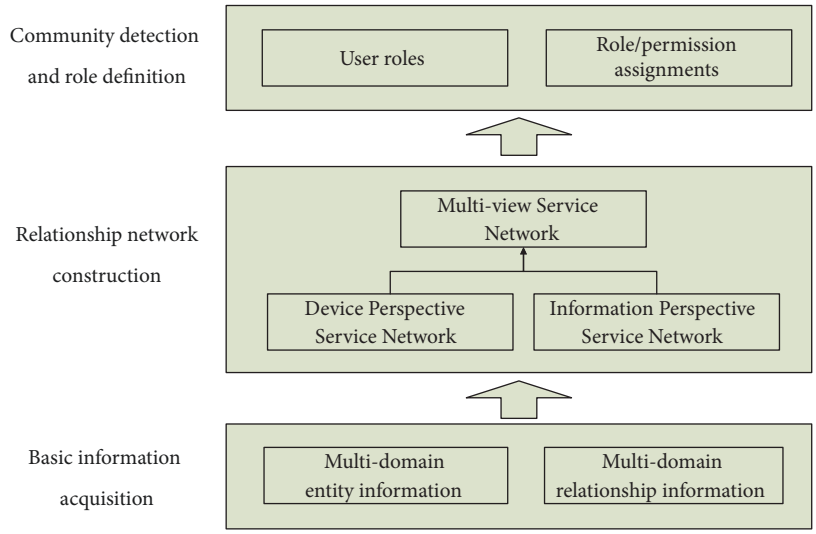


Figure 1: Role mining framework based on multi-domain information.

detection methods have been proposed, which utilize more information and achieve better performance.

Nonnegative Matrix Factorization (NMF) [29] is a classic clustering method, and several multi-view community detection methods based on NMF are proposed. Akata et al. proposed a method to jointly factorize multiple data matrices through a shared coefficient matrix [30]. Liu et al. proposed MultiNMF that regularizes the coefficient matrices learned from different views towards a common consensus for clustering [31]. He et al. extended NMF for multiview clustering by jointly factorizing the multiple matrices through coregularization [32]. Pei et al. proposed a nonnegative matrix tri-factorization (NMTF) based clustering framework with three types of graph regularization [33]. Li et al. proposed a framework based on regularized joint nonnegative matrix factorization (RJNMF) to utilize link and content information jointly to enhance the community detection accuracy [34].

In the framework RMMDI, we use the Pairwise Coregularized NMF clustering algorithm proposed in [32] to merge the information from two service networks (views). Experiments show that it can get more information than from a single view and make the role mining results more reasonable.

3. Role Mining Framework Based on Multi-Domain Information

In this paper, we proposed a role mining framework based on the multi-domain information, which is named as RMMDI. The framework is aimed at dividing possible user permissions into several disjoint subsets and assigning each subset to a user role. Then users are assigned with one or more necessary roles according to the permission they deserve. The structure of RMMDI is shown in Figure 1. The framework can be divided into three modules: basic information acquisition, relationship network construction, and community detection and role definition.

The basic information acquisition module obtains the necessary basic information from the target network, including multi-domain entity information and relationship information. The relationship network construction module constructs eight networks based on the obtained basic information, including the intermediate networks and ultimate networks. The community detection and role definition module detects permission communities on the ultimate networks by a multi-view community detection method and defines possible user roles.

3.1. Basic Information Acquisition. The basic information acquisition module is to collect network basic information, including the entities and entity relationships in the physical domain, network domain, and information domain, which are the foundation of relationship network construction.

3.1.1. Entity. There are five kinds of entities involved in the framework, i.e., space, object, service, info, and user.

Entity space represents specific physical space such as city, campus, building, or room, which is in the physical domain. All the space entities are represented as a set NS. Entity object is also located at the physical domain and represents network device like router, switch, or terminal. All the object entities are represented as a set NO. Entity service is in the network domain and represents network service like HTTP (Hypertext Transfer Protocol), FTP (File Transfer Protocol), and Email. All the service entities are represented as a set NV. Entity info is in the digital domain and represents the information like password, data, or digital file. All the info entities are represented as a set NI. Entity user represents network user. All the user entities are represented as a set NU.

3.1.2. Relationships. There are seven kinds of relationships involved in the framework, i.e., spatial similarity relationships, containment relationships, service access relationships, local management relationships, remote management relationships, service domination relationships, and info domination relationships.

Spatial similarity relationships are described by the matrix $M^{SS} \in A^{S\times S}$, where $A = \{0, 1\}$ and $S = |NS|$. $M^{SS}(i, j)$ is determined by

$$M^{SS}(i,j) = \begin{cases} 1, & \text{if } (i = j) \text{ or } \mathrm{u}(i,j) + \mathrm{u}(j,i) > \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where $u(i, j)$ is the number of users who can move from space ns_i to space ns_j and ε is the threshold value, ranging from 0 to $2|NU|$.

Device containment relationships are described by the matrix $M^{OS} \in A^{O\times S}$, where $A = \{0, 1\}$, $O = |NO|$, and $S = |NS|$. $M^{OS}(i, j)$ is determined by the following.

$$M^{OS}(i,j) = \begin{cases} 1, & \text{if object } no_i \text{ locates in space } ns_j \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Service access relationships are described by the matrix $M^{VO} \in A^{V\times O}$, where $A = \{0, 1\}$, $V = |NV|$, and $O = |NO|$. $M^{VO}(i, j)$ is determined by the following.

$$M^{VO}(i,j) = \begin{cases} 1, & \text{if service } nv_i \text{ can be reach by device } no_j \\ 0, & otherwise \end{cases} \quad (3)$$

Local management relationships are described by the matrix $M^{OV\_L} \in A^{O\times V}$, where $A = \{0, 1\}$, $O = |NO|$, and $V = |NV|$. $M^{OV\_L}(i, j)$ is determined by the following.

$$M^{OV\_L}(\mathrm{i},\mathrm{j}) = \begin{cases} 1, & \text{if device } no_i \text{ can be managed by service } nv_j \text{ locally} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Remote management relationships are described by the matrix $M^{OV\_R} \in A^{O\times V}$, where $A = \{0, 1\}$, $O = |NO|$, and $V = |NV|$. $M^{OV\_R}(i, j)$ is determined by the following.

$$M^{OV\_R}(i,j) = \begin{cases} 1, & \text{if device } no_i \text{ can be managed by service } nv_j \text{ remotely} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Service domination relationships are described by the matrix $M^{VI} \in A^{V\times I}$, where $A = \{0, 1\}$, $V = |NV|$, and $I = |NI|$. $M^{VI}(i, j)$ is determined by the following.

$$M^{VI}(i,j) = \begin{cases} 1, & \text{if the password of service } nv_i \text{ is } ni_j \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Info domination relationships are described by the matrix $M^{II} \in A^{I\times I}$, where $A = \{0, 1\}$ and $I = |NI|$. $M^{II}(i, j)$ is determined by

$$M^{II}(i,j) = \begin{cases} 1, & \text{if } \left(ni_j \longrightarrow ni_i\right) \text{ or } \left(ni_i \longrightarrow ni_j\right) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where symbol $a \longrightarrow b$ indicates that information a is dominated by information b. It means there is a service $v \in NV$, whose password is b and from which the users can get information a.

3.2. Relationship Network Construction. The relationship network construction module is to construct basic relationship networks based on the obtained basic information. As shown in Figure 2, there are eight networks to be constructed in total. The ultimate goal of this module is to form the Device View Service Network (DVSN), Information View Service Network (IVSN), and Multiview Service Network (MVSN). These three ultimate networks are used in community detection and role definition. Besides the three ultimate networks, there are other five networks involved in user-role mining, which are named as Local Management View Device Network (LMVDN), Remote Management View Device Network (RMVDN), Local Information View Device Network (LIVDN), Remote Information View Device Network (RIVDN), and Multiview Device Network (MVDN). The five intermediate networks are the foundation to construct ultimate networks. The meanings of intermediate networks and ultimate networks are described as follows.

3.2.1. Intermediate Networks. The five intermediate networks are described as undirected weighted graphs, whose adjacency matrices are constructed from the seven basic relationship matrices.

LMVDN. The LMVDN represents the similarity between devices from a spatial (local management) perspective, which means the devices located at similarity spaces are more similar than others. The network is represented by the adjacency matrix A^{OO_S}, whose values represent the similarity of two devices. The matrix is determined by the following.

$$A^{OO\_S} = M^{OS} M^{SS} \left(M^{OS}\right)^T \quad (8)$$

RMVDN. The RMVDN represents the similarity between devices from a remote management perspective, which means the devices that can be managed by similar management services are more similar than others. The network is represented by the adjacency matrix A^{OO_R}, whose values represent the similarity of two devices. The matrix is determined by the following.

$$A^{OO\_R} = M^{O\_VR} M^{VO} + \left(M^{O\_VR} M^{VO}\right)^T \quad (9)$$

LIVDN. The LIVDN represents the similarity between devices from the perspective of local management service password, which means the devices with similar local management service password are more similar than others. The network is represented by the adjacency matrix A^{OO_IL}, whose values represent the similarity of two devices. The matrix is determined by the following.

$$A^{OO\_IL} = M^{OV\_L} M^{VI} M^{II} \left(M^{OV\_L} M^{VI}\right)^T \quad (10)$$

RIVDN. The RIVDN represents the similarity between devices from the perspective of remote management service

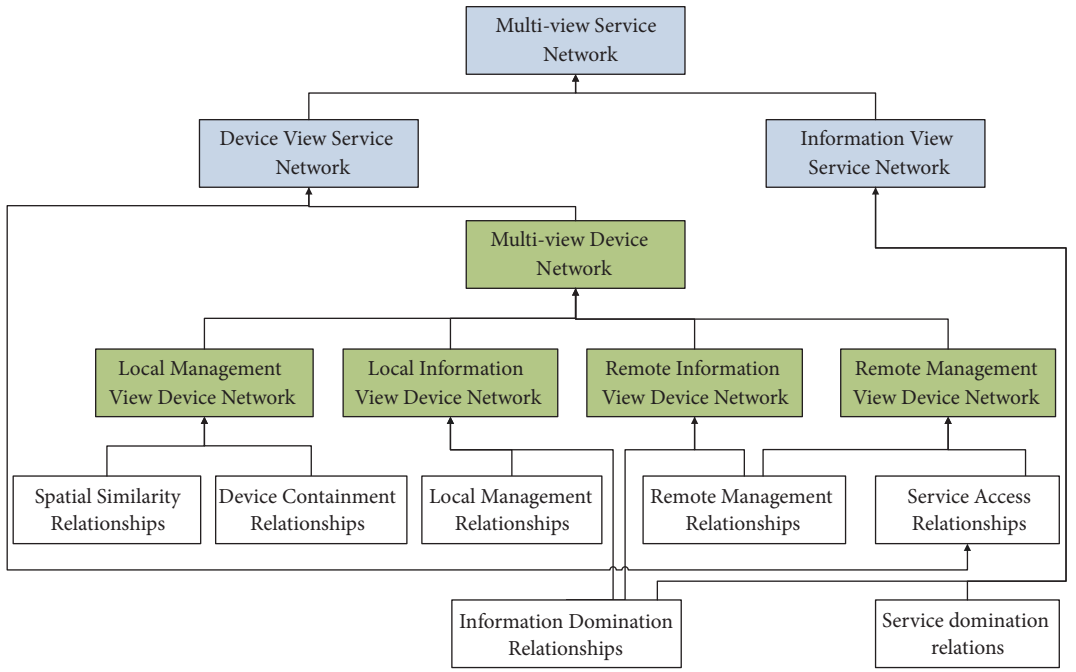


Figure 2: Relationship networks constructed in RMMDI.

password, which means the devices with similar remote management service password are more similar than others. The network is represented by the adjacency matrix A^{OO_IR}, whose values represent the similarity of two devices. The matrix is determined by the following.

$$A^{OO\_IR} = M^{OV\_R} M^{VI} M^{II} \left(M^{OV\_R} M^{VI}\right)^T \tag{11}$$

MVDN. The MVDN represents the similarity between devices from multiple perspectives, which merges the relationships from the local management perspective and the remote management perspective. The network is represented by the adjacency matrix A^{OO}, whose values represent the similarity of two devices. The matrix is determined by

$$A^{OO} = A^{OO\_S} \cdot A^{OO\_IL} + A^{OO\_R} \cdot A^{OO\_IR} \tag{12}$$

where the symbol $\cdot$ means dot product of two matrices.

3.2.2. Ultimate Networks. The three ultimate networks are also described as undirected weighted graphs, whose adjacency matrices are constructed from the seven basic relationship matrices and five intermediate networks.

DVSN. The DVSN represents the similarity of service permissions from a device perspective, which means the services accessed by similar devices are more similar than others. The network is represented by the adjacency matrix A^{VV_D}, whose values represent the similarity of two devices. The matrix is determined by

$$A^{VV\_D^*} = M^{VO} A^{OO} \left(M^{VO}\right)^T \tag{13}$$

$$A^{VV\_D} = \text{relationFilter}\left(A^{VV\_D^*}, \lambda\right) \tag{14}$$

where relationFilter(G, λ) is the function of filtering edges from the original graph, whose parameter G is the original graph and λ is the ratio of edges to be reserved.

As the matrix $A^{VV\_D^*}$ is a fully connected matrix in which the edges with small weight have negative impacts on community results, we use a function relationFilter(G, λ) to filter the low weight edges from the network. In function *relationFilter*(G, λ), for any node n in the graph G, we only reserve the top $\lambda \times \text{edgeNum}(n)$ edges with the largest weight. If two edges have the same weight, we reserve the edge between the node n and the neighbor node with a higher degree.

IVSN. The IVSN represents the similarity of service permissions from an information perspective, which means the services with a similar password are more similar than others. The network is represented by the adjacency matrix A^{VV_I}, whose values represent the similarity of two devices. The matrix is determined by

$$A^{VV\_I} = \text{relationFilter}\left(M^{VI} A^{II} \left(M^{VI}\right)^T, \lambda\right) \tag{15}$$

where relationFilter(G, λ) is the same edges filtering function in formula (14).

MVSN. The MVSN represents the similarity of service permissions from multiple perspectives, which merges the similarity relationships from the device perspective and the information perspective. The network is represented by the adjacency matrix A^{VV}, whose values represent the similarity of two devices. The matrix is determined by the following.

$$A^{VV} = A^{VV\_D} + A^{VV\_I} \tag{16}$$

Input: nonnegative matrices A^{VV_D}, A^{VV_I}. number of communities k, parameters $\lambda_D, \lambda_I, \lambda_{DI}$
Output: Service Community Division $C = \{c_1, c_2, \cdots, c_k\}$.
(1) Initialize $W^{VV\_D} \geq 0$, $H^{VV\_D} \geq 0$, $W^{VV\_I} \geq 0$, $H^{VV\_I} \geq 0$
(2) **While** Objective function does not converge and the Number of iterations is less than Threshold **do**
(3) Update H^{VV_D} according to Formula (18)
(4) Update H^{VV_I} according to Formula (19)
(5) Update W^{VV_D} according to Formula (20)
(6) Update W^{VV_I} according to Formula (21)
(7) **end while**
(8) Divide nodes to communities division $C = \{c_1, c_2, \cdots, c_k\}$ according to the coefficient matrix W^{VV_D}
(9) **return** $C = \{c_1, c_2, \cdots, c_k\}$

Algorithm 1: Permission community detection algorithm (PCDA).

3.3. Community Discovery and User-Role Definition. After building the ultimate networks, services can be divided into community relations through multi-view clustering algorithm, where all service permissions are divided into a community division $C = \{c_1, c_2, \cdots, c_k\}$. Then, for each $c_i \in C$, a role can be defined correspondingly. In this way, all network service permissions can be naturally assigned to k classes, where the possible values of k can be determined through algorithms such as maximum module degree.

In multiview service community discovery, we use the Pairwise Coregularized NMF clustering algorithm (PCoNMF) proposed in [32], which is based on regularized joint NMF. The objective function of service community discovery is formulated as follows.

$$\begin{aligned} J = {} & \lambda_D \left\| A^{VV\_D} - W^{VV\_D} \left(H^{VV\_D}\right)^{\mathrm{T}} \right\|_F^2 \\ & + \lambda_I \left\| A^{VV\_I} - W^{VV\_I} \left(H^{VV\_I}\right)^{\mathrm{T}} \right\|_F^2 \\ & + \lambda_{DI} \left\| W^{VV\_D} - W^{VV\_I} \right\|_F^2 \\ & s.t.\ H^{VV\_D} \geq 0,\ H^{VV\_I} \geq 0 \end{aligned} \tag{17}$$

The hypothesis behind PCoNMF is to regularize the coefficient matrices of the different views to a common consensus, which is then used for clustering. PCoNMF also adopts alternating optimization to minimize the objective function. The optimization works as follows: (1) fix the value of W^{VV_D} and W^{VV_I} while minimizing J over H^{VV_D} and H^{VV_I}; then (2) fix the value of H^{VV_D} and H^{VV_I} while minimizing J over W^{VV_D} and W^{VV_I}. We repeat the two steps until the iteration threshold is achieved.

According to [32], the update rules are as follows.

$$H^{VV\_D} \longleftarrow H^{VV\_D} \cdot \frac{\left(W^{VV\_D}\right)^T A^{VV\_D}}{\left(W^{VV\_D}\right)^T W^{VV\_D} A^{VV\_D}} \tag{18}$$

$$H^{VV\_I} \longleftarrow H^{VV\_I} \cdot \frac{\left(W^{VV\_I}\right)^T A^{VV\_I}}{\left(W^{VV\_I}\right)^T W^{VV\_I} A^{VV\_I}} \tag{19}$$

$$W^{VV\_D} \longleftarrow W^{VV\_D} \cdot \frac{\lambda_D A^{VV\_D} \left(H^{VV\_D}\right)^T + \lambda_{DI} W^{VV\_I}}{\lambda_D W^{VV\_D} H^{VV\_D} \left(H^{VV\_D}\right)^T + \lambda_{DI} W^{VV\_D}} \tag{20}$$

$$W^{VV\_I} \longleftarrow W^{VV\_I} \cdot \frac{\lambda_I A^{VV\_I} \left(H^{VV\_I}\right)^T + \lambda_{DI} W^{VV\_D}}{\lambda_I W^{VV\_I} H^{VV\_I} \left(H^{VV\_I}\right)^T + \lambda_{DI} W^{VV\_I}} \tag{21}$$

Hence, the permission community detection algorithm is shown as Algorithm 1.

4. Experiments and Results

In this section, we evaluate our role mining method based on the multi-domain information of a simulated network, which is the simplification of the inner network of Corporation M.

4.1. Experiment Environment. We built a simulation network for experiments, including a router, a firewall, an Intrusion Prevention System (IPS), 3 switches (Switch1, Switch2, and Switch3), 6 servers (WServer, DServer, FServer, GServer, OServer, and IServer), 3 gate machines (GM1, GM2, and GM3), and 13 terminals (T1, T2, T3, ..., T13). We used a HUAWEI S7706 as the core router, three HUAWEI S5700 as switches, a TOPSEC NGFW 4000-UF as the firewall, a TOPSEC IDP 3000 as IPS, and computers from Dell and HP as the servers or terminals. The router enabled 3-layer routing and the firewall were configured with bidirectional access control lists. All the servers and terminals were installed with different versions of Windows, including Windows 2003 Server, Windows XP, and Windows 7. We deployed an entrance guard system including 3 gate machines and a server (GServer). The gate machines used face recognition technology to determine whether a person can pass or not. An office automation system was deployed on the OServer, whose database was deployed on the DServer. We also deployed two websites and an FTP using IIS (Internet Information Services) on WServer, IServer, and FServer. Similarly, the websites depended on the same database deployed on

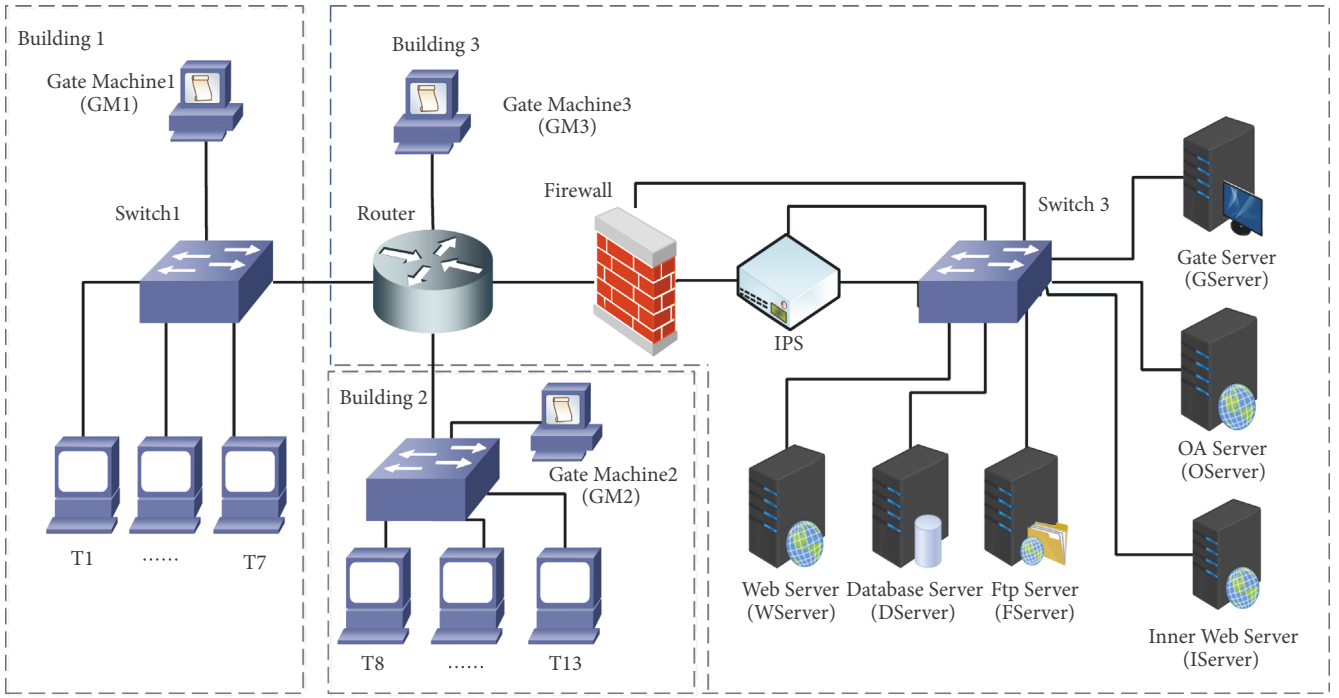


Figure 3: An example network.

DServer. The physical link relationships among devices are shown in Figure 3.

All the devices are distributed in 12 rooms in 3 buildings. 10 devices are located in building 1: terminal T1, T2, and T3 are in room 1-1; T4 and T5 are in room 1-4; T6 and T7 are in room 1-5; Switch1 is in room 1-2; and GM1 is in the hall of building 1 (room 1-3). 8 devices are located in building 2: terminals T8 and T9 are in room 2-1; T10 and T11 are in room 2-4; T12 and T13 are in room 2-5; Switch2 is in room 2-2; and GM2 is in the hall of building 2 (room 2-3). 10 devices are located in building 3: router, firewall, IPS, Switch3, and all servers are in room 3-1, and GM3 is in the hall of building 3 (room 3-2).

There were 34 services in the network, including 28 management services and 6 business services. The management services were used for device management, while the business services were used for corporation business. Each device was managed by a management service. The router and switches enabled SSH service. The servers and terminals enabled the Remote Desktop Service. In addition, the gate machines enabled web-based management interfaces. The website deployed on WServer provided a web service on port 80 named as WS_W, which was used to publish public information. The FServer provided an FTP service on port 21 named as FS_F, which was used by Network Administrators to share information. The GServer provided a data transmission service on port 8080 named as GS_T, which was used to synchronize data between GM machine and GServer. The OServer provided a web service on port 80 named as OS_W, which was used to document circulation for all users. The IServer provided a web service on port 80 named as IS_W, which was used by Server Administrators to share information. The DServer provided a database service on port 1433 named as DS_D, which was used to provide underlying support for WS_W, OS_W, and IS_W.

There were 33 passwords in the analysis. Each service, except for WS_W and OS_W, has a password. Besides, *NULL* was added to represent the empty password. All the information involved is shown in Table 1.

There were 13 users involved in analysis named from User1 to User13, who used terminals T1 to T13 and knew passwords T1_M_P to T13_M_P, respectively. Using the top-down approaches, the network security administrators had gotten 5 user roles for the business information, which were named as Ordinary User, Server Administrator, Database Administrator, Network Administrator, and Security Administrator. The role-permission assignments are listed in Table 2.

4.2. Baseline Methods. To demonstrate the effectiveness of our method, we compare our approach with two groups of baselines. The first group comprises 5 clustering methods: 2 single view methods and 3 multiview methods. The second group comprises 4 traditional role mining methods: ORCA (OFFIS Role mining tool with Cluster Analysis), CM (Complete Miner), HPr (HP Role Minimization), and HPe (HP Edge Minimization)

4.2.1. Clustering Methods. SP (Spectral Clustering). SP [35] is a classical single view clustering algorithm, which makes use of the eigenvalues of the data similarity matrix to perform dimensionality reduction before clustering in fewer dimensions. The similarity matrix is provided as an input, consisting of a quantitative assessment of the relative similarity of each pair of points in the dataset.

SymNMF. SymNMF [36] is a clustering algorithm based on NMF, which takes a nonnegative and symmetric matrix as

Table 1: Device related information.

Device	Location	Services	Password	Device	Location	Services	Password
T1	R1-1	T1_M	T1_M_P	GM2	R2-3	G2_M	G2_M_P
T2	R1-1	T2_M	T2_M_P	GM3	R3-2	G3_M	G3_M_P
T3	R1-1	T3_M	T3_M_P	Switch1	R1-2	S1_M	S1_M_P
T4	R1-4	T4_M	T4_M_P	Switch2	R2-2	S2_M	S2_M_P
T5	R1-4	T5_M	T5_M_P	Switch3	R3-1	S3_M	S3_M_P
T6	R1-5	T6_M	T6_M_P	Router	R3-1	R_M	R_M_P
T7	R1-5	T7_M	T7_M_P	Firewall	R3-1	F_M	F_M_P
T8	R2-1	T8_M	T8_M_P	IPS	R3-1	IPS_M	IPS_M_P
T9	R2-1	T9_M	T9_M_P	WServer	R3-1	WS_W WS_M	- - WS_M_P
T10	R2-4	T10_M	T10_M_P	DServer	R3-1	DS_D DS_M	DS_D_P DS_M_P
T11	R2-4	T11_M	T11_M_P	FServer	R3-1	FS_F FS_M	FS_F_P FS_M_P
T12	R2-5	T12_M	T12_M_P	GServer	R3-1	GS_T GS_M	GS_T_P GS_M_P
T13	R2-5	T13_M	T13_M_P	OServer	R3-1	OS_W OS_M	– OS_M_P
GM1	R1-3	G1_M	G1_M_P	IServer	R3-1	IS_W IS_M	IS_W_P IS_M_P

Table 2: User role-permission assignments by top-down methods.

Roles	Service Permissions
Ordinary User	WS_W, OS_W
Server Administrator	WS_M, FS_M, GS_M, OS_M, IS_M, DS_M, IS_W
Database Administrator	DS_D
Network Administrator	S1_M, S2_M,S3_M, R_M, FS_F
Security Administrator	F_M, IPS_M, G1_M, G2_M, G3_M, GS_T

an input. The matrix contains pairwise similarity values of a similarity graph and is approximated by a lower rank matrix instead of the product of two lower rank matrices.

PCoSpec (Pairwise Coregularized Spectral clustering) and CCoSpec (Center-wise Coregularized Spectral clustering). Two coregularization schemes are adopted in spectral clustering framework [37], PCoSpec utilizes a pairwise coregularization to enforce the eigenvectors of each pair to be similar, and CCoSpec employs the centroid-based coregularization to enforce the eigenvectors to be similar with a common center.

CCoNMF (Cluster-wise Coregularized NMF clustering). CCoNMF extends NMF for multiview clustering by jointly factorizing the multiple matrices through cluster-wise coregularization [32], which enforces the cluster similarity matrices to be similar.

RMSC (Robust Multiview Spectral Clustering). RMSC [38] is a multiview spectral clustering method based on Markov chain, which explicitly handles the possible noise in the transition probability matrices associated with different views.

4.2.2. Mining Baseline Methods. ORCA. ORCA [11] is the first role mining algorithm, which uses the hierarchical clustering technology to discover user roles. The algorithm defines each permission as an initial cluster first, then merges the clusters, and forms a role hierarchy.

Complete Miner (CM). CM [10] is another classic role mining algorithm proposed in 2006. It starts by creating an initial set of roles for the distinct user-permission sets, then computes all possible intersection sets of the initial roles, and outputs a list of candidate roles.

HP Role Minimization and HP Edge Minimization. HP Role Minimization (HPr) and HP Edge Minimization (HPe) [14] are the role mining algorithms based on minimum biclique coverage. HPr tries to find a minimal set of roles that override the user-permission assignment relationship, while HPe uses a heuristic method to find the smallest number of edges of an RBAC system.

4.3. Experiments Setup

4.3.1. Scenarios Construction. To validate our framework and method, we built 3 scenarios named as Scenario1 (S1), Scenario2 (S2), and Scenario3 (S3) based on the basic experimental environment shown in Figure 3 and assigned users one or more user roles, which is shown in Table 3. We can find out that each user in S1 was assigned only 1 user role, while they were assigned 2 user roles in S2. In S3, users working

Table 3: User-role assignments in different scenario.

Scenario	Ordinary User	Server Administrator	Database Administrator	Network Administrator	Security Administrator
S1	User1, User2, User3	User4, User5	User6, User7	User8, User9, User10,User11	User12, User13
S2	All Users	User7, User8	User11, User12, User13	User4, User5	User9, User10
S3	User1, User2, User5, User6, User9, User10, User13	User4, User7	User8	User11	User12

in one room may be assigned different user roles, which may introduce more vulnerabilities to network security.

For each scenario, we first configured the gate machines and firewall according to Tables 2 and 3. Spatial access control lists were added on gate machines, making users have physical access to devices they managed or used, while network access control lists were added on the firewall, making the terminals have network access to the target services.

It should be noted that there were potential conflicts among multi-domain configurations on the semantic level. Take the user User4 in S1 as example. User4 was a Server Administrator and should not access service DS_M and the firewall had forbidden T4 to access service DS_D directly, but T4 was permitted to access service WS_M and there was no firewall between WServer and DServer. Thus, User4 can use T4 to log in WServer remotely first and then access service DS_D (he can get the password DS_D_P from the configuration files on WServer). This is a typical semantic conflict between the network access control lists. Similarly, as User4 had the ability to access DS_D physically by entering the room 3-1, there is another conflict between the network access control list and the spatial access control list. Those conflicts may result in extra permissions for users.

Then, we extracted basic information from the network and established the necessary relationship matrices. Note that there were 16 space entities, 28 object entities, 34 service entities, 32 information entities, and 13 user entities in all the 3 scenarios. The relationship matrices $M^{OS} \in \mathbb{R}^{28\times16}$, $M^{VI} \in \mathbb{R}^{34\times32}$, $M^{OV\_L} \in \mathbb{R}^{28\times34}$, and $M^{OV\_R} \in \mathbb{R}^{28\times34}$ were the same in all three scenarios, where $|M^{OS}|_0 = |M^{OV\_L}|_0 = |M^{OV\_R}|_0 = 28$ and $|M^{VI}|_0 = 34$. The matrices $M^{SS} \in \mathbb{R}^{16\times16}$ and $M^{VO} \in \mathbb{R}^{34\times28}$ varied from scenario to scenario, depending on the configurations of gate machines or firewall. For each space pair (S_1, S_2), we counted the users who can move from S_1 to S_2 under each scenario and set the parameter $\varepsilon = 14$ when constructing the matrix M^{SS}. The results showed that $|M^{SS}|_0 = 46$ in S1, $|M^{SS}|_0 = 46$ in S2, and $|M^{SS}|_0 = 42$ in S3. We used the scanner NMAP to get the accessibility relationships between devices and services, establishing the matrix M^{VO}. The results showed that $|M^{VO}|_0 = 549$ in S1, $|M^{VO}|_0 = 566$ in S2, and $|M^{SS}|_0 = 548$ in S3.

Finally, we detected the user roles by RMMDI and compared the results with the two groups of baseline methods. On the one hand, we performed the role mining baseline methods based on the user-permission assignment (UPA) matrices constructed from the firewall configurations and compared the results with RMMDI. On the other hand, we studied the best parameters for each clustering method and then compared the effectiveness of RMMDI with the clustering baseline methods. Accuracy and normalized mutual information (NMI) [31–33, 36] were adopted to evaluate the community detection effectiveness of different parameters, whose values both range from 0 to 1 and a higher value means better effectiveness. We calculated the accuracy and NMI of different clustering methods with ground truth after studying the best parameters for each clustering method. In the experiments, we constructed two ground truths manually. In one ground truth, we divided 21 service permissions into 5 roles according to Table 2. In the other ground truth, we combined the roles "Database Administrator" with "Server Administrator" and classified 21 service permissions into 4 roles. For one community detection result, we compared it with the two ground truths and calculated metrics separately.

4.4. Result

4.4.1. Role Mining Results. Firstly, we performed the baseline role mining methods based on the firewall configurations. As the firewall only conducted the network access control lists, it can only reflect the accessibility between devices and services. Since each terminal was assigned to a user, we can get the 3 different UPA matrices from it. As we wanted to find disjoint service subnets, we used $W = \langle 1, 1, 1, \infty, \infty\rangle$ as the optimization objective. In order to save space, only the permission divisions are shown in Table 4. We found that the permission divisions were almost the same as those in Table 2, which meant that the role mining methods can find no errors from the top-down approach.

Then, we also performed the RMMDI on all scenarios with the role number k = 5,100 times for each scenario. The majority of the results were different from the result shown in Table 2. The most frequent result (222 in 300 times) is listed as Table 5. Comparing with Table 2, we counted the number of inconsistent classification results of each service. All 18 services were classified inconsistently for 572 times in total. And the top 2 services with the most inconsistent

Table 4: Role mining results of baselines on all scenarios.

Scenario	Method	Role	Permissions
S1	ORCA CM HPr HPe	Role1	WS_W, OS_W
		Role2	WS_M, FS_M, GS_M, OS_M, IS_M, DS_M, IS_W
		Role3	DS_D
		Role4	S1_M, S2_M,S3_M, R_M, FS_F
		Role5	F_M, IPS_M, GS_T
S2	HPr	Role1	WS_W, OS_W
		Role2	WS_W, OS_W, WS_M, FS_M, GS_M, OS_M, IS_M, DS_M, IS_W
		Role3	WS_W, OS_W, DS_D
		Role4	WS_W, OS_W, S1_M, S2_M, S3_M, R_M, FS_F
		Role5	WS_W, OS_W, F_M, IPS_M, GS_T
S2	ORCA CM HPe	Role1	WS_W, OS_W
		Role2	WS_M, FS_M, GS_M, OS_M, IS_M, DS_M, IS_W
		Role3	WS_W, OS_W, DS_D
		Role4	S1_M, S2_M, S3_M, R_M, FS_F
		Role5	F_M, IPS_M, GS_T
S3	ORCA CM HPr HPe	Role1	WS_W, OS_W
		Role2	WS_M, FS_M, GS_M, OS_M, IS_M, DS_M, IS_W
		Role3	DS_D
		Role4	S1_M, S2_M,S3_M, R_M, FS_F
		Role5	F_M, IPS_M, GS_T

Table 5: The most common result of RMMDI when k=5.

Method	Role	Permissions
RMMDI	Role1	WS_W, OS_W
	Role2	WS_M, FS_M, GS_M, OS_M, IS_M, DS_M, IS_W, DS_D
	Role3	GS_T
	Role4	S1_M, S2_M,S3_M, R_M, FS_F
	Role5	F_M, IPS_M

classification times were DS_D (282 times) and GS_T (258 times).

Finally, we changed the role number $k = 4$ and performed the experiments 100 times under each scenario. The most common results are shown in Table 6.

4.4.2. Parameter Study Results. We studied the parameters used in RMMDI as well as the baseline clustering methods. We performed a series of experiments for a series of different parameters and tried to find out the optimal parameters. The experiments were conducted under S2 with role number $k = 4$.

We first studied the parameters used in baseline methods, including λ_{p_spec} in PCoSpec, $\lambda_{c\_spec\_D}$ and $\lambda_{c\_spec\_I}$ in CCoSpec, and λ_{rmsc} in RMSC. With each parameter, we set $\lambda = 0.05$ and studied 30 different values between 0.005 and 100. When $\lambda_{p\_spec} = 0.15$, $\lambda_{c\_spec\_D} = 8$, $\lambda_{c\_spec\_I} = 1$, and $\lambda_{rmsc} = 0.15$, the methods had relatively better effectiveness on the dataset.

Then, we studied the parameters in PCoNMF and CCoNMF. There are 3 parameters: λ_D, λ_I, and λ_{DI}. λ_D and λ_I represent the weights of view A^{VV_D} and A^{VV_I}, while λ_{DI} is the regularization parameter. We studied 27 different ratios of λ_D to λ_I between 10 and 0.02, as well as 10 different values of λ_{DI} between 0.1 and 10. The experiments were conducted 50 times for each pair. The results are shown in Figures 4 and 5. We found that the parameters λ_D and λ_I had little impact on both the accuracy and NMI when $\lambda_D/\lambda_I < 1$ and $0.5 < \lambda_{DI} < 1.5$, so we used $\lambda_D = 1$, $\lambda_I = 3$, and $\lambda_{DI} = 0.9$ in the study of λ and the clustering experiments shown in Section 4.4.3.

Finally, we studied the parameter λ used in RMMDI. We performed an experiment with a series of λ from 0.05 to 1 with step size 0.5 and observed their impacts on the clustering

Table 6: The most common result of RMMDI when k=4.

Method	Role	Permissions
RMMDI	Role1	WS_W, OS_W
	Role2	WS_M, FS_M, GS_M, OS_M, IS_M, DS_M, IS_W, DS_D
	Role3	S1_M, S2_M,S3_M, R_M, FS_F
	Role4	F_M, IPS_M, GS_T

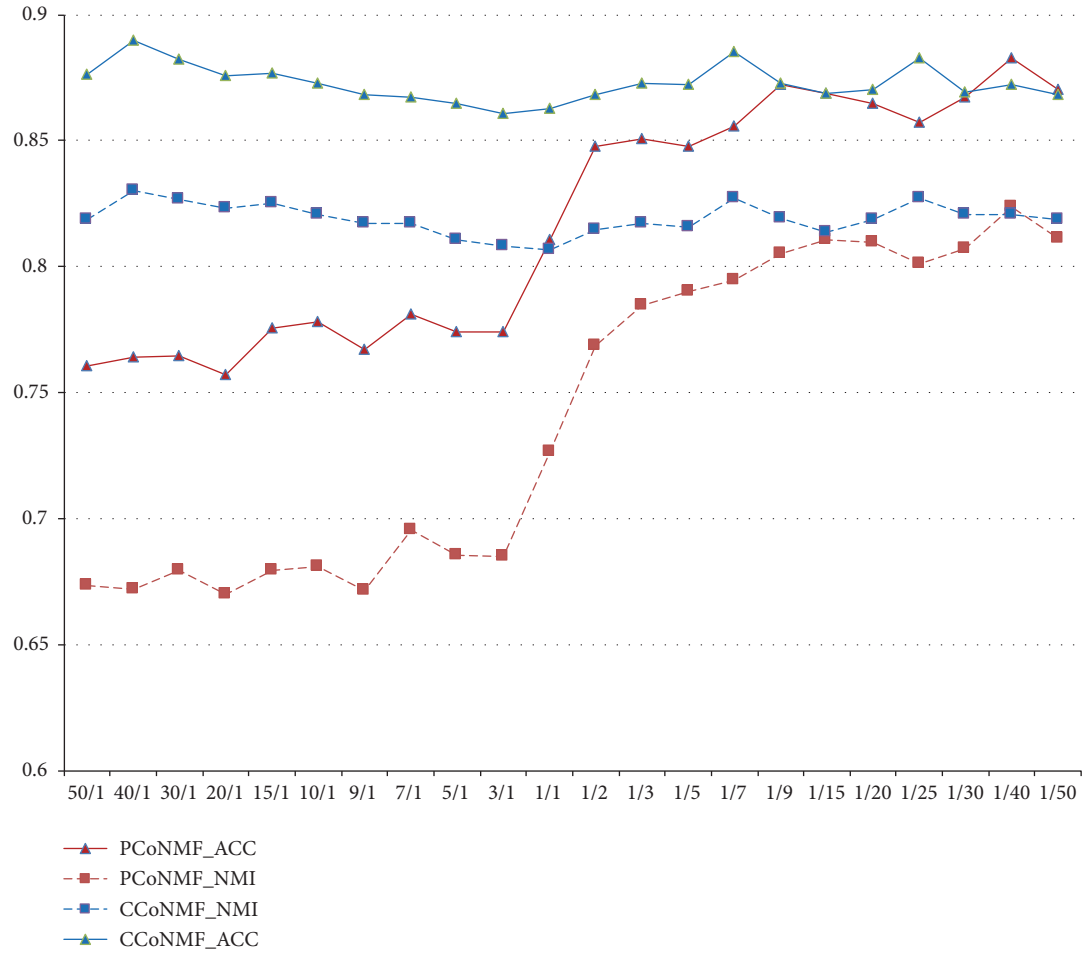


Figure 4: Evaluating the accuracy and NMI of PCoNMF and CCoNMF on varying λ_D/λ_I.

effectiveness (shown in Figure 6). The other parameters were set as mentioned in the previous paragraph.

We found that the curves showed a downward trend in whole, and the accuracy and NMI got greater values when λ was around 0.3, which was used for the experiments shown in Section 4.4.3.

4.4.3. Clustering Results. We also conducted experiments to compare the effectiveness of RMMDI with the clustering baseline methods. We performed all algorithms 200 times on each scenario and compared results with the ground truth shown in Table 4. All the other parameters were set as the optimal values mentioned in Section 4.4.2. The results are listed in Tables 7–9, in which the results of SP and SymNMF were the larger result of 3 different inputs, A^{VV}, A^{VV_D}, and A^{VV_I}. Most of those values were from the input A^{VV_I}.

5. Discussion

We propose a novel user role framework, which uses multiple domain information to mine user roles other than the preassigned user-permission assignment matrix.

It is proved that the framework is suitable for role mining. For the three scenarios used in the experiment, different users are assigned to different user roles. One user may be assigned one or more user roles, and one user role may be assigned to several users. For the results listed in Tables 7 and 8, we can find that the accuracy of the proposed framework is greater

Table 7: Accuracy for different methods on 3 scenarios.

Scenario	SP	SymNMF	PCoSpec	CCoSpec	PCoNMF	CCoNMF	RMSC
S1	0.9010	0.9229	0.7121	0.9057	*0.9381*	0.9190	0.6952
S2	0.8981	0.9276	0.6675	0.9072	*0.9428*	0.9365	0.5762
S3	0.9210	0.9181	0.7070	0.9035	*0.9365*	0.9333	0.6333

Table 8: NMI for different methods on 3 scenarios.

Scenario	SP	SymNMF	PCoSpec	CCoSpec	PCoNMF	CCoNMF	RMSC
S1	0.8605	0.8571	0.5827	0.8414	*0.8738*	0.8579	0.6382
S2	0.8506	0.8703	0.4869	0.8497	*0.8879*	0.8744	0.479
S3	0.8692	0.8538	0.5958	0.8464	*0.8774*	0.8719	0.4890

Table 9: Runtime for different methods on 3 scenarios (s).

Scenario	SP	SymNMF	PCoSpec	CCoSpec	PCoNMF	CCoNMF	RMSC
S1	0.0178	0.0109	0.1367	0.1247	0.8533	*1.1568*	0.0580
S2	0.0145	0.0091	0.0973	0.1108	0.8518	*1.0866*	0.0440
S3	0.0127	0.0085	0.0929	0.1031	0.8494	*1.2394*	0.0230

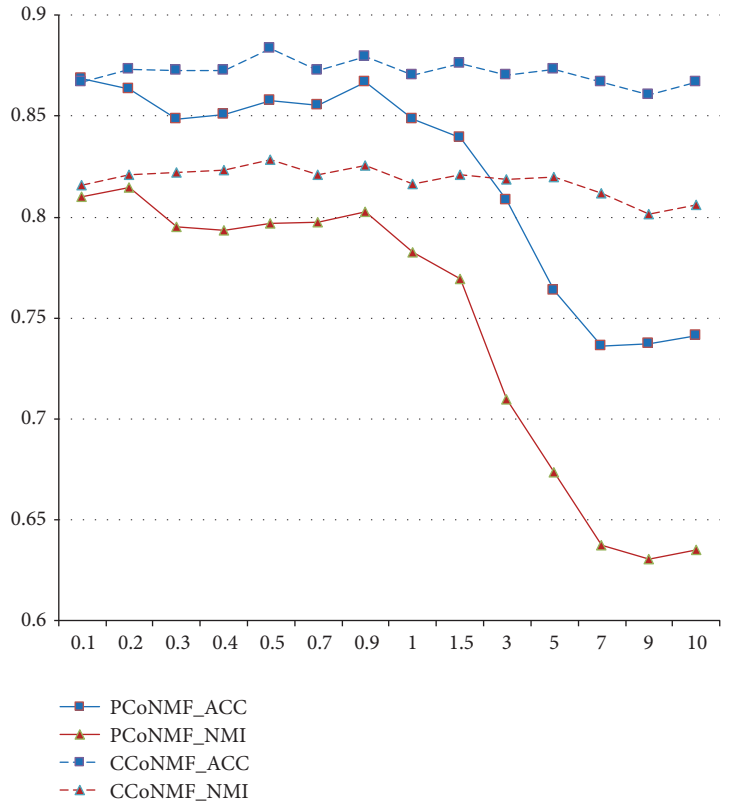
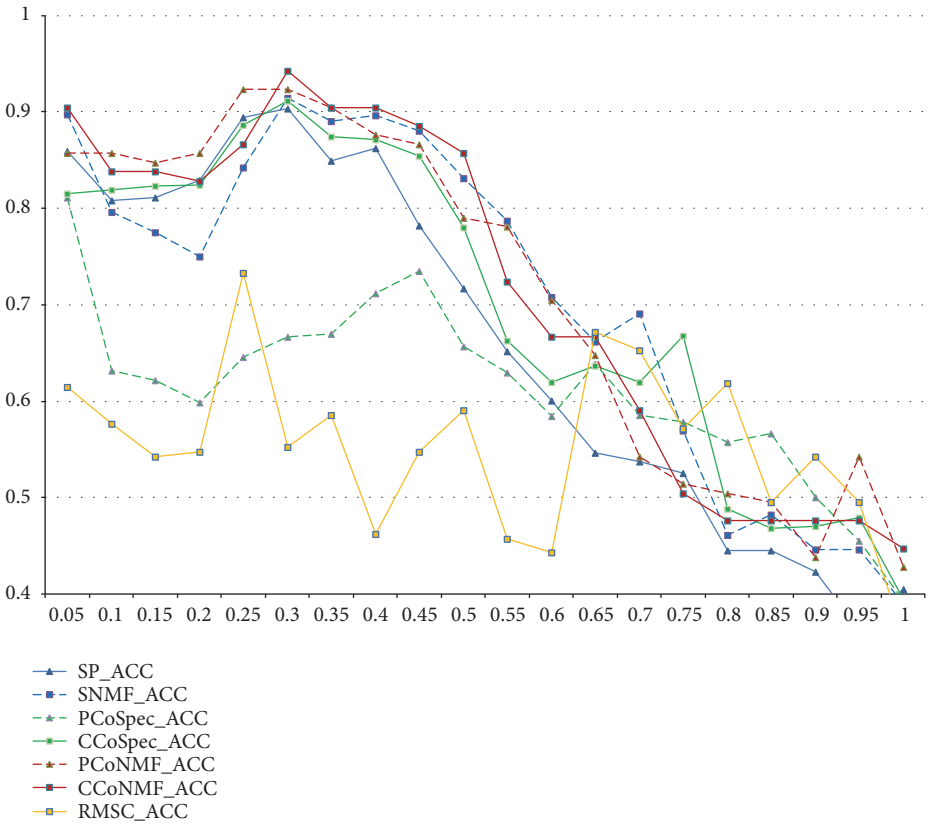


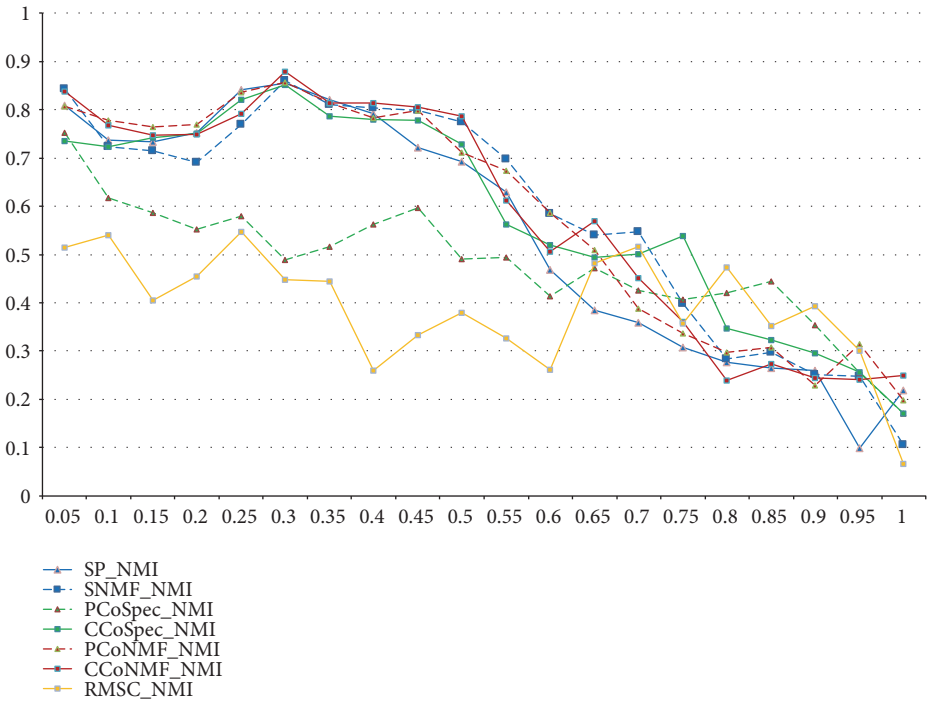
Figure 5: Evaluating the accuracy and NMI of PCoNMF and CCoNMF on varying λ_{DI}.

than 93.5% in all the three scenarios, while the NMI is greater than 87.0%. It means that the framework can detect user roles from the multiple domain configuration information successfully.

More importantly, it is also demonstrated that the framework has the ability to find interdependent relationships between permissions, avoiding potential errors. From the experimental results in Section 4.4.1, we find that RMMDI tends to integrate user roles "Database Administrator" and "Server Administrator". Analyzing user potential permissions, it can be found that all Server Administrators can access service DS_D as they can both reach the service from



(a) Evaluating the accuracy of RMMDI on varying λ



(b) Evaluating the NMI of RMMDI on varying λ

Figure 6: Evaluating the accuracy and NMI of RMMDI on varying λ.

other servers and get its password from configuration files in WServer. Therefore, it may be more appropriate to integrate user roles "Database Administrator" and "Server Administrator". This trend cannot be found by traditional methods.

It is also proved that the performances of different clustering methods vary in the framework. As shown in Tables 7 and 8, the accuracy and NMI of PCoNMF are always the best among all the three scenarios, 30% better than the worst method. It means that a reasonable clustering method will promote the effectiveness of the framework significantly. Compared with the single view clustering methods, PCoNMF promotes the accuracy and NMI by more than 1%, which means the reasonable utilization of information from multiple views will get more structure information than from single view. Both the PCoSpec and RMSC have a lower accuracy or NMI, which means the multiview methods based on spectral clustering may not be suitable to the datasets.

There are 4 parameters involved in the RMMDI in total and it is important to select proper values to the parameters. The first two parameters λ_D and λ_I are the weights of views A^{VV_D} and A^{VV_I}. From the results in Figure 4, we find that the view A^{VV_I} has more structure information than A^{VV_D} in the experiments and it is reasonable to set a larger λ_I and a smaller λ_D. The third parameter λ_{DI} is the regularization parameter that indicates the degree of community proximity between the two perspectives. A too low λ_{DI} will not establish the connection between two views. Nevertheless, as the view A^{VV_I} has more structure information than A^{VV_D}, a too big λ_{DI} will reduce the accuracy of the algorithm. Therefore, a moderate parameter λ_{DI} ($0.5 < \lambda_{DI} < 1.5$) will make the algorithm perform better. The last parameter λ is used in the function relationFilter(G, λ), which means to reserve the top $\lambda \times edgeNum(n)$ edges with the largest weight. From the results shown in Figure 6, we find that it is vital to reserve an appropriate proportion of links. A big λ will reserve more low weight links and the existences of low weight links will impact the effectiveness of the framework. However, a low value of λ will lost a lot of useful structure information, which will have an impact on the effectiveness of the algorithm too.

6. Conclusion

In this paper, a novel framework for role mining based on multi-domain information named as RMMDI is proposed. The key idea of the framework is to mine user roles from multiple domain information rather than existing user-permission assignment matrices. In the framework, information from the physical domain, network domain, and digital domain is used to find the relationships between user permissions, and multi-view community detection methods are used to integrate information from different domains. Experiments on 3 simulated network scenarios demonstrate that RMMDI can capture the interdependent relationships between permissions and perform user-role mining more effectively and reasonably.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the grants from the National Key R&D Program of China (Project No. 2017YFB0802800).

References

[1] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models," *The Computer Journal*, vol. 29, pp. 38–47, 1996.

[2] A. Colantonio, R. D. Pietro, A. Ocello, and N. V. Verde, "A formal framework to elicit roles with business meaning in RBAC systems," in *Proceedings of the 14th ACM Symposium on Access Control Models And Technologies*, pp. 85–94, ACM, Stresa, Italy, 2009.

[3] A. Baumgrass, M. Strembeck, and S. Rinderle-Ma, "Deriving role engineering artifacts from business processes and scenario models," in *Proceedings of the 16th ACM Symposium on Access Control Models and Technologies*, pp. 11–20, ACM, Innsbruck, Austria, 2011.

[4] A. Colantonio, R. Di Pietro, A. Ocello, and N. V. Verde, "Taming role mining complexity in RBAC," *Computers & Security*, vol. 29, pp. 548–564, 2010.

[5] A. Colantonio, R. Di Pietro, and N. V. Verde, "A business-driven decomposition methodology for role mining," *Computers & Security*, vol. 31, no. 7, pp. 844–855, 2012.

[6] S. Hachana, F. Cuppens, N. Cuppens-Boulahia, and J. Garcia-Alfaro, "Semantic analysis of role mining results and shadowed roles detection," *Information Security Technical Report*, vol. 17, pp. 131–147, 2013.

[7] M. Kuhlmann, D. Shohat, and G. Schimpf, "Role mining - revealing business roles for security administration using data mining technology," in *Proceedings of the Eighth ACM Symposium on Access Control Models and Technologies*, pp. 179–186, ACM, Como, Italy, 2003.

[8] I. Molloy, N. Li, T. Li, Z. Mao, Q. Wang, and J. Lobo, "Evaluating role mining algorithms," in *Proceedings of the ACM Symposium on Access Control Models and Technologies*, pp. 95–104, 2009.

[9] J. Jiang, X. Yuan, and R. Mao, "Research on role mining algorithms in RBAC," in *Proceedings of the 2018 2nd High Performance Computing and Cluster Technologies Conference*, pp. 1–5, ACM, 2018.

[10] J. Vaidya, V. Atluri, and J. Warner, "RoleMiner: mining roles using subset enumeration," in *Proceedings of the 13th ACM Conference on Computer and Communications Security*, pp. 144–153, ACM, Alexandria, Va, USA, 2006.

[11] J. Schlegelmilch and U. Steffens, "Role mining with ORCA," in *Proceedings of the Tenth ACM Symposium on Access Control Models and Technologies*, pp. 168–176, ACM, Stockholm, Sweden, 2005.

[12] I. Molloy, H. Chen, T. Li et al., "Mining roles with semantic meanings," in *Proceedings of the ACM Symposium on Access Control Models and Technologies, SACMAT '08*, pp. 21–30, Estes Park, Colo, Usa, 2008.

[13] D. Zhang, K. Ramamohanarao, and T. Ebringer, "Role engineering using graph optimisation," in *Proceedings of the 12th ACM*

Symposium on Access Control Models and Technologies, pp. 139–144, ACM, Sophia Antipolis, France, June 2007.

[14] A. Ene, W. Horne, N. Milosavljevic et al., *Fast Exact and Heuristic Methods for Role Minimization Problems*, 2008.

[15] M. Frank, D. Basin, and J. M. Buhmann, "A class of probabilistic models for role engineering," in *Proceedings of the 15th ACM Conference on Computer and Communications Security*, pp. 299–310, ACM, Alexandria, Va, USA, 2008.

[16] M. Frank, A. P. Streich, D. A. Basin et al., "A probabilistic approach to hybrid role mining," in *Proceedings of the Acm Conference on Computer & Communications Security*, 2009.

[17] I. Molloy, N. Li, Y. Qi et al., "Mining roles with noisy data," in *Proceedings of the ACM Symposium on Access Control MODELS and Technologies*, pp. 45–54, 2010.

[18] X. Du and X. Chang, "Performance of AI algorithms for mining meaningful roles," in *Proceedings of the 2014 IEEE Congress on Evolutionary Computation (CEC)*, pp. 2070–2076, 2014.

[19] L. Dong, Y. Wang, R. Liu, B. Pi, and L. Wu, "Toward edge minability for role mining in bipartite networks," *Physica A: Statistical Mechanics and its Applications*, vol. 462, pp. 274–286, 2016.

[20] L. Wu, L. Dong, Y. Wang et al., "Uniform-scale assessment of role minimization in bipartite networks and its application to access control," *Physica A: Statistical Mechanics and its Applications*, vol. 507, pp. 381–397, 2018.

[21] Q. Guo, J. Vaidya, and V. Atluri, "The role hierarchy mining problem: discovery of optimal role hierarchies," 2008.

[22] A. Colantonio, R. D. Pietro, and A. Ocello, "A cost-driven approach to role engineering," in *Proceedings of the Acm Symposium on Applied Computing*, 2008.

[23] C. W. Probst, R. R. Hansen, and F. Nielson, *Where Can an Insider Attack?* Springer, Berlin, Germany, 2007.

[24] C. W. Probst and R. R. Hansen, "An extensible analysable system model," in *Elsevier Advanced Technology Publications*, vol. 13, pp. 235–246, 2008.

[25] I. Kotenko, M. Stepashkin, and E. Doynikova, "Security analysis of information systems taking into account social engineering attacks," in *Proceedings of the 2011 19th International Euromicro Conference on Parallel, Distributed and Network-Based Processing*, pp. 611–618, 2011.

[26] D. Scott, A. Beresford, and A. Mycroft, "Spatial policies for sentient mobile applications," in *Proceedings of the 4th IEEE International Workshop on Policies for Distributed Systems and Networks*, p. 147, IEEE Computer Society, 2003.

[27] T. Dimkov, *Alignment of Organizational Security Policies: Theory and Practice*, University of Twente, Enschede, Netherlands, 2012.

[28] F. Kammüller and C. W. Probst, "Invalidating policies using structural information," in *Proceedings of the 2013 IEEE Security and Privacy Workshops*, pp. 76–81, 2013.

[29] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[30] A. Wendel, S. Sternig, M. Godec et al., "Non-negative matrix factorization in multimodality data for segmentation and label prediction," *Computer Vision Winter Workshop*, 2011.

[31] J. Liu, C. Wang, J. Gao et al., "Multi-View clustering via joint nonnegative matrix factorization," in *Proceedings of the 2013 SIAM International Conference on Data Mining*, pp. 252–260, 2013.

[32] X. He, M.-Y. Kan, P. Xie, and X. Chen, "Comment-based multi-view clustering of web 2.0 items," in *Proceedings of the 23rd International Conference on World Wide Web*, pp. 771–782, ACM, Seoul, Korea, 2014.

[33] Y. Pei, N. Chakraborty, and K. Sycara, "Nonnegative matrix tri-factorization with graph regularization for community detection in social networks," in *Proceedings of the 24th International Conference on Artificial Intelligence*, pp. 2083–2089, AAAI Press, Buenos Aires, Argentina, 2015.

[34] Z. Li, Z. Pan, Y. Zhang, G. Li, and G. Hu, "Efficient community detection in heterogeneous social networks," *Mathematical Problems in Engineering*, vol. 2016, Article ID 5750645, 15 pages, 2016.

[35] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[36] D. Kuang, S. Yun, and H. Park, "SymNMF: nonnegative low-rank approximation of a similarity matrix for graph clustering," *Journal of Global Optimization*, vol. 62, pp. 545–574, 2015.

[37] A. Kumar, P. Rai, and H. Daumé, "Co-regularized multi-view spectral clustering," *Advances in Neural Information Processing Systems*, 2011.

[38] R. Xia, Y. Pan, L. Du et al., "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp. 2149–2155, AAAI Press, Québec, Canada, 2014.