Fine-grained Access Control using Email Social Network

by Yiru Li and Anil Somayaji, School of Computer Science, Carleton University, Ottawa, Canada, and Carrie Gates, CA Labs, CA Technologies

Abstract

We introduce the idea of applying social network analysis to organizational access control. With current access control systems, users are typically assigned access to many more resources than they need and permissions are rarely updated. We propose to refine these access policies using the dynamic social network defined by intra-organizational emails. Email social networks within organizations reveal working relationships between employees. These revealed relationships can supplement standard access control groups, providing a way to enforce finer-grained access with minimal system administrator overhead.

1. Introduction

Current access control systems are problematic; users are typically assigned access to much more information than they need and permissions are rarely updated. Ponemon in a 2008 survey found that 52 percent of the IT professionals surveyed did not believe access rights are well-managed in their organizations, and 78 percent thought employees often had privilege to information resources which are not pertinent with their job function[7]. In addition, current access control systems cannot keep pace with the change of personnel and resources within organizations. New employees join, employees leave, and new resources such as documents, spreadsheets and databases, are created. In current practice, administrators need to be informed of changes and then manually update the systems. This practice likely results in a large time lag between changes in organizations and updates of security systems, which enables exploitation of sensitive resources.

These access control problems are due to usability and administration constraints - administrators have very limited knowledge of personnel and the resources to be protected, and have constraints on work hours. Role-Based Access Control (RBAC)[4], which is designed to make the administration easy, is widely used in organizations. Its usability problems are often represented with specific terms like “role engineering”, “role discovery” and “role managemen”'. In order to discover roles for RBAC systems, administrators have to collaborate with people in different business units because of their lack of business knowledge. In reality, however, such collaboration does not always happen. Ponemon found that only 57% of those surveyed organizations said security, IT and business departments in their organizations work together[7]. Approaches that complement the defects of the human communication between administrators and business people are needed. Role-mining is the main approach described in the literature, which uses machine-learning technologies to automatically identify roles[6, 15]. Besides RBAC, there exist varied theoretical access control models, which claim tight and just-in-time access control policies. Facing the same usability problems, these models can only be applied to specific types of business where computer systems are fully developed to track the status on personnel and resources.

About the authors

Yiru Li is a Ph.D. candidate in Carleton University. Her research area is mitigating enterprise insider threats using social networks.

Dr. Anil Somayaji is an Associate Professor in the School of Computer Science at Carleton University in Ottawa, Canada. He received a B.S. (1994) in Mathematics from the Massachusetts Institute of Technology and a Ph.D. (2002) in Computer Science from the University of New Mexico. He has served as the program committee chair of the New Security Paradigms Workshop and has served on the program committees of major computer security venues including ACM CCS, USENIX Security, ACSAC, and RAID, among others. His research interests include computer security, operating systems, complex adaptive systems, and artificial life.

In this paper, we present a new way of automatically identifying groups of users who have work relationships in an organization through analysis of the organization's email social networks. Analysis of an organization's email corpus provides information and insight leading to an understanding of the communicative relationships in the social network in a particular organization[10]. Email corpora have been studied with the intention of revealing the relationships among employees, extracting a social hierarchy, ranking the major officers in an organization, grouping connected users, and reproducing the organizational structure, but no studies have discussed the possibility of applying the revealed relationships between employees to usability problems in access control systems.

Based on preliminary experiments on three users' email archives, we can identify user groups which are attributed to collaboration for common task goals. We refer these identified user groups to as Communities of Collaboration (CoCos). These identified CoCos are analogous, but are not interchangeble with the roles in RBAC. They are more like the "team" concept in the team-based access control model[13] but are smaller units than teams or subteams which are formally constructed in an organization. Identifying CoCos through email social networks is more advantageous than role-mining technologies which often use system configurations as a data source. These identified CoCos have semantic meanings and the dynamic nature of email social networks enables timely capture CoCo formation and dissolution.

We believe that CoCos can be applied to supplement access control systems by providing administrators information on users' collaboration groups. Alternatively, it can be used as a new access control model - CoCo-based access control. Compared to other access control models, CoCo-based access control has the potential to improve security in business environments where informal collaboration groups constantly form and dissolve with valuable data stored in a central repository. Specifically, we discuss two potential application areas for CoCos, supplementing standard access control systems and assessing the risk of access requests.

The paper is organized as follows. In Section 2, we discuss usability problems of access control. In Section 3, we present related work on analysis of email social networks. In Section 4, we describe preliminary experiments of extracting CoCos from three users' email archives. In Section 5, we compare identified CoCos and three ways of grouping users in three access control models. In Section 6, we discuss two possible application areas for CoCos. In Section 7, we conclude and discuss future work.

2. Usability Problems in Organizational Access Control

There exist a variety of theoretical access control models: role-based access control model[11], task-based access control model[8], team-based access control model[13], attribute-based access control model[18] and so on. Except for RBAC which is designed for easy management, most access control models are designed to make fine-grained access control policies.

All these theoretical models are difficult to implement, because the models require business knowledge of users and resources, and assume that such knowledge can be easily obtained by manual identification by administrators or automatic capture by computer systems. Even though RBAC is designed to be easy to manage, in practice it is still difficult to implement especially in large businesses. RBAC's implementation difficulties have been studied in terms of

Dr. Carrie Gates is a Distinguished Engineer, senior vice president and director of research with CA Labs, the research arm within CA Technologies. She is responsible for performing research that has the potential to impact the strategic direction of CA Technologies products and services. This is achieved through identifying opportunities within the business units at the company that can be transformed into research relationships performed in collaboration with university faculty and students, with a focus on research in the area of enterprise-level security.

She is also actively pursuing research in the areas of insider threat detection and usable security. In addition to her security research, Dr. Gates is involved in research in sense-making and network traffic analysis, and has recently done work on security architectures and cloud computing. She has given several invited talks, served on both Masters and PhD thesis committees, and is on several organizing and program committees for academic conferences in the areas of governance, forensics, insider threat, and general security.

Dr. Gates' complete biography and publications are available at ca.com/calabs.

discovering and maintaining roles. In RBAC systems, administrators define roles and assign each user to the corresponding roles. Users with the same role are granted the same set of permission to resources. In order to construct roles, administrators should perform a detailed analysis of business processes and derive roles from such analysis. The process of defining roles should be based on a complete analysis of how an organization functions and should include input from a wide spectrum of employees, including business line managers and human resource[17]. However in practice, organizations rarely follow this process because it consumes too many work hours and is slow.

The process of defining roles should be based on a complete analysis of how an organization functions and should include input from a wide spectrum of employees ... However in practice, organizations rarely follow this process because it consumes too many work hours and is slow.

To overcome the drawbacks of the manual role discovery in RBAC systems, automation was introduced - a machine learning based method called “role mining” uses data mining techniques on existing system configuration data[6, 12, 16, 15, 19]. This approach can potentially accelerate RBAC system construction to a great extent, but it is limited by the fact that a role inferred by this approach is a set of permissions that does not correspond to any real-world concepts, such as a job position or a work location. Even though in this paper we have similar intention to that of related role mining work in terms of automatically identifying user groups, user groups that we try to identify are associated with tasks in real world. In addition, these role mining techniques have been developed assuming a static environment and thus are not able to catch changes on roles[1]. Consequently, administrators still need to manually keep track of changing of users' roles, for example, adding a user to a role, removing a user from a role, and update it to the systems.

3. Related Work on Email Social Network Analysis

A social network is a graph made of individuals who are connected by a set of social relationships, such as friendship, co-working, kinship and so on. Facebook and Linkedln are two well-known on-line social networks that respectively reveal friendship relationships and professional relationships. In the literature of social network analysis, a social network is partitioned into subnetworks in order to identify user groups (also called “clusters”, or “communities”). A partitioned subnetwork will be considered to have community structure if it consists of subsets of vertices, with many edges connecting vertices of the same subject, but few edges lying between subsets.

Email corpora are of particular interest for researchers focusing on organizational theory and behavior. Analysis of email communication in an organization enables the examination of social and organizational processes in real-world over a long period of time[3]. Through organizational email logs, Tyler et al.[14] identified a kind of user community, which is supposed to represent an informal network of collaboration that naturally grows and coalesces within organizations. Conceptually, user communities they identify are similar to CoCos, but their experimental results reveal that the user communities their algorithm identifies are formal organization units such as departments, or project teams. Johansen et al.[5, 2] also used email corpora to determine associations between individuals by measuring features of volume, directionality and frequency of email, and briefly discussed potential applications for automating email management, such as topical classification, flagging important messages, and SPAM mitigation.

Email corpora are of particular interest for researchers focusing on organizational theory and behavior.

J. Diesner et al.[3] used the Enron email data set to investigate dynamics of the relationships as organizational management crisis escalated. R. Rowe et al.[10] extracted the social hierarchy from email communications, and believe that their algorithm enables ranking the major officers of an organization, grouping

similarly ranked and connected users, and accurately reproducing the organizational structure.

4. The Preliminary Experiment

4.1 Data Source

We assume that most email communication in an organization is related to work or the organization rather than personal; most people have multiple email accounts and use each email account for a different purpose. Their work email account is for work issues and other email accounts are for personal. We emphasize that the valid data sources are email archives associated with an organization's email accounts.

We assume that most email communication in an organization is related to work or the organization rather than personal; most people have multiple email accounts and use each email account for a different purpose.

Three users' email archives are used in this preliminary analysis. All these email archives are associated with the users' work email accounts. These three users have different occupations: Subject 1 is a Ph.D. student, Subject 2 is faculty at a university, Subject 3 is research staff in a large business. The following details the three users' email archives.

- Subject 1's work email archive during the year 2009, which contains around 3000 messages. In this email archive, there are about 800 email IDs.
- Subject 2's work email archive over the year 2007, which contains 7305 unique messages. In this email archive, there are about 1712 email IDs.
- Subject 3's work email archive where all email messages are related to one project during the year 2008. There are a total of 1818 messages.

4.2 Algorithm

In this experiment, we intend to identify CoCos for each individual user rather than for a formal organizational unit such as departments, labs or teams. Correspondingly, we focus on each user's email archives separately from other users'.

We designed a very simple algorithm to identify CoCos for users through their email archives, as described below:

- First, we collect all email groups from a user's email archives. An email group is composed of people whose email addresses appeared at least once together in the email headers “to”, “cc”, or “bcc”.
- Second, we filter out email groups which are not compliant with the following three rules:

  Rule 1: The total number of messages exchanged for a certain period of time in an email group is more than a threshold, which is set as 10 messages at minimum for one year.

  Rule 2: The total number of people in an email group who have ever sent email messages is at least two.

  Rule 3: The total number of people in an email group is more than two.

  Email groups which remain after filtering with these three rules are considered as CoCos.

In this preliminary experiment, the threshold in Rule 1, in which at least 10 messages must be exchanged over the period of a year, is very small for email exchanges for a specific task. The numbers of messages exchanged reflect how

important the associated tasks are to the user. For example, in Table 2, the four groups of Subject 2 with less than 15 messages (Group 11, 20, 21, and 22) are associated with tasks that are not Subject 2's main responsibilities as faculty. (Group 5 and 16 are excluded, because they are part of task groups, which will be discussed in Section 4.3.) We purposely set a small threshold because we rather capture trivial tasks than miss important ones. Also, the time span of the threshold, one year, is too long. We think that a span of 30 days would be better.

We purposely set a small threshold because we rather capture trivial tasks than miss important ones.

With Rule 2, a threshold of 2 means that there are at least two people in the email group who send at least one message. We infer from observation that groups in which there is only one email sender are very likely to be event announcements or commercial advertisements. In Rule 3, we observe that users exchange mail with almost every email ID in his email archives, and thus there are a large number of two-user groups which remain after filtering with Rule 1 and 2. We also observe that email communication between two users likely covers various topics including jobs and casual personal issues. Thus we choose to neglect email groups composed of two persons. However, this means that two-person CoCos will be missed.

These three rules as well as their thresholds are a result of our observations with the email archives. The thresholds probably need to be adjusted for email collected in different business environments. Automatically adjusting thresholds for optimum classification results is an area for future research. Better rules or algorithms may exist beyond these three rules. In this paper, our focus is on demonstrating that CoCos can be extracted from email archives,

Subj.	Total Email Grps.	Groups Rule 1	Groups Rule 1, 2	Groups Rule 1, 2, 3
1	350	22	9	7
2	1600	127	100	23
3	356	72	38	21

Table 1: Email Groups Filtered with Three Rules

rather than on algorithm optimization.

4.3 Preliminary Results

Table 1 shows numbers of email groups that remain after filtering with Rule 1, 2 and 3. Consider subject 2's email archive as an example. There are a total of 1600 email groups in his email archive. Filtering with Rule 1, 127 email groups remain. Consecutively filtering with Rule 2, 100 email groups remain out of the 127 groups. After filtering with these three rules, only 23 groups remain, which are supposed to be CoCos.

Out of subject 1's 350 email groups in total, 7 groups remain which are supposed to be CoCos. Out of subject 3's 356 email groups in total, 21 groups remain which are supposed to be CoCos.

Table 2 shows the detail of each of Subject 2's 23 email groups which remain after filtering with Rule 1, 2 and 3. The “Name” column shows user names; the “Msgs” column shows the numbers of messages exchanged over one-year period; and the “Topics” column is manually labelled by Subject 2 and shows what tasks were performed in each group. Consider the first group as an

example. Group 1 consists of five persons, exchanged a total of 109 messages, and the task of its email communication is about writing a paper.

No.	Name	Msgs	Topics
1	A-E	109	paper
2	B,E,F,G	55	G's honour's project
3	B,D,E	79	D's dissertation
4	B,C,H,I	20	paper
5	B,C,I	14	paper
6	,I,J	47	I's thesis work
7	B,E,K	28	K's MCS thesis
8	B,L,M	22	co-supervising M's Ph.D.
9	B,C,J,N	20	defense committee
10	B,O-S	24	O's Ph.D. defense
11	B,T,U	12	lab seminars
12	B,U,V	39	grant management
13	B,C,W	18	Research office discussion
14	B,X,Y	84	Consulting
15	B,X,Z	33	Consulting
16	B,X,Y,Z	14	Consulting
17	B,C,AA-AD	40	Project mtgs., reports
18	B,J,AE	40	Collaboration mtgs.
19	B,C,I,K	23	Company visit
20	B,C,AF,AG	12	Company visit
21	B,V,AH-AK	14	faculty committee
22	B,C,AA,AD,AL – AW	15	grant proposal
23	B,AX,AY	26	family event

Table 2: Detail of Subject 2's 23 Email Groups. Note that Subject 2 is in fact B; hence B is a member of all these groups.

We filtered each email group using the three rules for a CoCo and then verified the results, which checks the true positive rate. Verification is based on the conceptual definition of a CoCo - a group of users will be considered a CoCo if they collaborate and share a common task goal. We interviewed each user to verify the results.

We also verified whether each of the groups which was filtered out is not a CoCo (the false negative rate) using interviews. We asked each user, "Can topics in the email groups which remain after filtering with the three rules represent most of the main work tasks and activities that he conducted in the year"?An alternative is to randomly select a number of email groups that our algorithm does not identify as CoCos and to verify those specifically with users. We will use the latter approach in our future work.

We only evaluated the algorithm's accuracy of group discovery with the first two subjects because the evaluation requires a lot in-person interaction with each subject and the third subject was not available.

Verification on Subject 1's Email Groups

Our algorithm identifies 7 email groups as CoCos in Subject 1's email archive. Subject 1 verified that each of the 7 groups is associated with a specific work task. We conclude that false positives are zero; the 7 email groups identified with our algorithm are all CoCos. Based on Subject 1's memory, he confirms that work tasks covered by these 7 groups are the main activities which occurred during the period of the email archive, and that there are not any significant activities missing. Therefore we conclude that false negatives are zero---no CoCos are missing.

Verification on Subject 2's Email Groups

Our algorithm identifies Subject 2's 23 email groups as CoCos out of his total 1600 email groups. Through interviewing the subject, it was verified that each of the 23 groups in Table 2 is associated with a specific task. However, as explained below, not all of them are CoCos.

- Not all of the 23 groups are associated with work. Group 23 is composed of Subject 2's family, and its communication is about one family traveling event. Thus this group is falsely identified as a CoCo.
- Multiple groups are associated with the same work task. Group 4 and 5 are associated with the same task - writing a paper about web security. The three persons in Group 5 - B, C, and I - started the work earlier and later Group 4 was formed with joining of another person, H. Also, Groups 14, 15 and 16 are associated with the same task - consultation and class presentation in a company. Thus, Group 4 and 5 should be combined into one CoCo, and Group 14, 15, and 16 should be combined into another CoCo. Group 5, 14 and 15 are not complete CoCos because they only contain some of the collaborators in those tasks.

We conclude that out of the 23 groups, there should be 19 CoCos for Subject 2, and thus false positives are 4 out of 23. In terms of false negatives, we conduct the evaluation by asking a question to Subject 2, "Can topics in these 23 groups represent most of main work tasks and activities that he conducted in 2007"? Because of his affirmative answer, we infer that the 23 groups represent Subject 2's most main activity groups and there are no important work groups missing.

The difference in the accuracy between the two subjects may be important. We need more data to characterize how the accuracy rate impacts the resulting access control implementation. In this experiment, the three subjects provided us their email data sets based on the trust on us. In the next step for analyzing email archives from more subjects, to protect these subjects' privacy, we have built an email tool which will be sent to potential subjects. Each subject will run the tool with his email archives in his computer and will do evaluation on the groups discovered from his email. All we will see is accuracy rates which are sent back by the subjects. In this way, the subjects' privacy is completely protected. When it is applied to access control systems, the email analysis tool should be embedded in SMTP servers.

In the next step for analyzing email archives from more subjects, to protect these subjects' privacy, we have built an email tool which will be sent to potential subjects.

5. CoCos vs. User Groups in Existing Access Control Models

We first discuss characteristics of CoCos which are derived from our observations in Table 2 in Section 4.3. Then we compare the identified CoCos with the ways of grouping users in three access control models: RBAC[11], team-based access control (TMAC)[13], and task-role-based access control[8]. We choose these three among the many theoretical access control models because the

factors used to assign permissions are related to users' work such as roles, teams and tasks. In this comparison, we only discuss the factors relevant to assigning users to user groups. Here we specifically do not discuss how permissions would be assigned to groups; instead, we focus on the factors which cause users to be assigned the same set of permissions.

Characteristics of CoCos

- Each CoCo is associated with one specific task.
- Some tasks are derived from the same project. Group 1, 2 and 3 are all associated with the same project, Project-NN, but each group has a different task. Thus CoCos could consist of sub groups of a large project team.
- Tasks last for varying periods of time, and correspondingly, CoCos exist for various periods. For example, writing a paper can last a few months or even a year, while arranging a seminar may just take a few days.
- Not all of tasks are associated with users' main responsibilities. As faculty, Subject 2's main responsibility is supervising students, writing papers, teaching, applying for grants. Among the tasks associated to the 23 groups in Table 2 in Section 4.3, while most tasks are about Subject 2's main job responsibilities, some involve non-position specific duties such as travel arrangements.
- Members in Subject 2's CoCos are often in different geographic locations and are part of different organizations. Because most access control systems are built within a boundary, such as a lab, a department or a project team, many of Subject 2's CoCos would be cut out when applied to such access control systems. Indeed, this is why much of the data sharing he does is through email attachments rather than shared computer storage.

CoCos vs. Role-Based Access Control

In RBAC systems, roles are created for various job functions in an organization and users are assigned to roles based on their responsibilities and qualifications. Each role is associated with a set of permissions.

In RBAC systems, roles are created for various job functions in an organization and users are assigned to roles based on their responsibilities and qualifications. Each role is associated with a set of permissions.

It is recognized that there is little agreement on what RBAC means, with the result that a role is interpreted in different ways by researchers and system developers[11]. A role can represent competency in specific tasks (a pharmacist), can embody authority and responsibility (a project manager), and can reflect specific duty assignments that are rotated through multiple users (a duty physician, a shift manager).

We compare roles and CoCos through an example. Suppose a research lab in a university, which consists of two professors, three post-docs and a number of graduate students. A few roles, “every researcher”, “project X”, “fund management”, “system back-up” are defined. All of the lab members are assigned to the “every researcher” role, which is assigned permissions to access all documents about research work except for project X. A few people are assigned to the “project X” role, which is assigned permissions to access documents related to project X. The two professors are assigned to the “fund management” role, which is assigned permissions to access documents related to fund management. Two graduate students are assigned to the “system back-up role, which is assigned permissions to back up the system.

Not all of the above roles are consistent with CoCos. The roles, “project X” and “fund management” are compatible with CoCos, while two roles “every research” and “system back-up” are not. The “system back-up” role in the example is not a CoCo, because backing up the system is a routine task performed by one of two individuals every day, and there is no collaboration between them. In CoCo-context access control, permissions of backing up a system could not be assigned.

In research labs, resources can be categorized into computer system-related resources and business-related resources. In the above example, resources associated with the roles “project X” and “fund management” are business-related. Resources associated with the role “system back-up” are computer system-related. Tasks related to the computer systems can be characterized as routine, repetitive and no collaboration is needed most of the time. In contrast, tasks related to research can be characterized as dynamic with a lot of collaboration is needed. Thus we conclude that CoCo-context access control is not suitable for business environments where tasks are routine and repetitive such as system administration, customer representatives and bank tellers.

CoCos vs. Team-Based Access Control

In team-based access control[13], the notion of “team” is defined as an abstraction that encapsulates a collection of users in specific roles with the objective of accomplishing a specific task or goal. In the model, a team is associated with a set of team resources which are supposed to be only accessed by the team members. A medical clinic setting is used to demonstrate the model, where a team is a number of medical staff with a goal of treating a particular patient, and the team resources are information about the patient. Permissions associated to such a team are also discriminated by roles of medical staff, such as physicians and nurses. In the model, users who take the same roles in the same team will be assigned the same set of permissions.

CoCos are very similar to the “team” concept, because a CoCo is composed of a number of people collaborating on a specific task. The difference lies in that CoCos represent both formal and informal collaboration units, and medical teams described in team-based access control model are part of formal management structure. In some business environments, informal collaboration units can be formed and disassembled very frequently. For example, an engineering project team is composed of dozens of sub-teams, and each sub-team contains dozens of users. We call the project team and its sub-teams as part of formal management structure because they can be found in the organization chart. In the contrast, an informal collaboration unit could be a few users in a sub-team who are assigned by the team manager to finish a task.

CoCos are very similar to the “team” concept, because a CoCo is composed of a number of people collaborating on a specific task. The difference lies in that CoCos represent both formal and informal collaboration units.

CoCos vs. Task-Role Based Access Control

In task-role based access control (T-BAC)[8], what permissions should be assigned to users depends on what tasks the users are assigned to do. Tasks are defined as a fundamental unit of business activity, and as actual work units that employees are assigned, and as the smallest unit of job assignments. Oh and Park describe a sale department to illustrate their access control model. Each activity of the department such as reviewing customers' statistics, reviewing sale results, is assigned a set of permissions. T-BAC assumes that tasks can be separated before actual business activities occur. This assumption may be true in environments where tasks are repetitive. In many other environments, task separation and assignment are a dynamic business process and are not pre-

specified. For example, in a software-development company, each project being developed is likely different from previous ones and project leaders, who are supposed to have most knowledge on the projects, very likely only have a rough plan of how tasks should be separated at the start of the projects.

There are three important differences between CoCos and T-BAC in terms of tasks:

There are three important differences between CoCos and T-BAC in terms of tasks.

- The size of tasks. Unlike the task concept in T-BAC model, tasks associated to CoCos do not have to be the smallest unit of job assignments. We have observed that there exists a hierarchy structure between CoCos; a task associated to a CoCo can be a sub-task of another task associated to anther CoCo. For example, a CoCo contains 17 users with a task of dealing with a sale contract, and another CoCo contains a few of the 17 users with a task of reviewing the contract content, which is part of the task of the sale contract.
- Task separation. Unlike T-BAC model, with CoCos we assume that tasks are separated and assigned with the process of work activities.
- Knowledge of tasks. With our current algorithm of identifying CoCos, we do not have knowledge of the task associated with a CoCo. This contrasts with T-BAC model where knowledge on what each task is about is required.

6. Potential Applications

From the above discussion it can be seen that conceptually CoCo has similarities as well as important differences to the other three access control models in terms of how users are grouped. In this section we briefly describe two potential applications for email identified CoCos.

6.1 Supplementing Standard Access Control Systems

CoCos, which are identified through email archives, can be applied to off-line recommendation systems for administrators. Identified CoCos can aid administrators by supplementing business knowledge of user activities when deploying or maintaining an access control system. Administrators can extract a user's current task groups and obtain information about whom a user is working with through his email archive. Such information can help administrators determine what roles or which user groups the user should be assigned to. Email's temporal nature makes the identified CoCos also temporal, and can be used to remind administrators to adjust a user's roles or user groups after the user's job responsibilities have changed. When the system senses a large difference between the user's current roles and his CoCos, administrators would be alerted that an update of that user's roles may be appropriate.

Such information can help administrators determine what roles or which user groups the user should be assigned to.

Identified CoCos can also be applied in real-time to complement access control policies. Based on the assumption that employees who collaborate are likely to access the same resources, CoCos can be used to decide whether to grant access to a resource. For example, consider the following scenario: Due to an administrator's mistake, a senior manager is not assigned a role which is associated with permission to access sensitive documents. Over a weekend, the senior manager urgently needs to access the documents, but his access is denied by the traditional access control system. However, with access to his CoCos, his request may be granted if he is in the same CoCo with other senior management employees who have access to this resource. His common membership in CoCos with other senior managers implies that he works with

them, and is likely to need access to the same resources. Because of the possibility of error in extracting CoCos, some restrictions may be needed with granting this request; for example, he may be allowed to read but not print. Also, a record of resource access granted by common CoCo memberships would be audited by administrators.

6.2 CoCo-based Risk Measurement on Access Requests

We introduce a framework for CoCo-based risk measurement of access request authorizations. This framework is based on the assumption that users who have work relationships with each other are likely to access the same resources. A user's CoCos reflect his work relationships. Two users are considered to have a work relationship if they are common members in another user's CoCo. In this framework, risk is estimated for each access request through calculating probabilities according to relationships between employees. Probabilities reflect how much that request senders need to access the resources for their work needs. The higher the probability, the more likely the access request is for a legitimate work need, reducing the risk created by authorizing the access request.

We define three terms for use in the framework: the resource, the resource owner and the resource accessor. A resource, for example, is a file or a piece of data in a database. A resource owner is the user who creates the protected resource. A resource accessor is the user who is requesting access to the resource. The owner and the accessor each have a set of CoCos, which can be manually identified by each user or automatically identified by analysis of data sources such as email archives.

Risk for each access request is measured according to three factors: the relationship between the resource owner and resource accessor, the relationships between the accessor and other accessors who have requested access to the resource, and the relationships between the resource owner and other accessors who have requested access to the resource. Concretely, these three factors can be transformed into the following three CoCo-based questions:

Risk for each access request is measured according to three factors: the relationship between the resource owner and resource accessor, the relationships between the accessor and other accessors who have requested access to the resource, and the relationships between the resource owner and other accessors who have requested access to the resource.

1. Is the accessor in one of the owner's CoCos? If not, we infer that the accessor and the owner do not have a work relationship and that the accessor less likely needs to access the resource created by the resource owner.
2. Does anyone who is in the accessor's CoCos have a history of requesting access to the resource? If not, we infer that none of users with whom the accessor has work relationships needed to access the resource and that the accessor less likely needs to access the resource.
3. What is the ratio of users who are (or were) in at least one of the owner's CoCos and requested access to the resource in the past, to the total number of users who requested access to the resource? If the ratio is large (e.g., greater than 80%, for example), we conclude that this resource is more likely to be needed by users who have work relationships with the owner.

The risk of an access request can be calculated based on the answers to the three questions. For example, if the answer to the three questions is “no”, “no” and “large ratio”, the risk is very high because of the aggregation of the low likelihood of the accessor's work need inferred from each question. Thus the access request should be refused.

We use the above three questions as an example to present how CoCos can be used for risk measurement on access requests. There could be more CoCo-based rules or algorithms. In order to prove the effectiveness of the algorithms, experiments with real data would be needed.

Beyond the effectiveness of algorithms, challenges which need to be addressed include:

1. A user's CoCos will change over time as his tasks change. Thus a mechanism or criteria is needed to determine the birth and death of a CoCo.
2. Malicious employees who have the intention of accessing sensitive resources might interfere with CoCo formation so that they can be granted access to the resources that carry a low risk.
3. Regardless of how CoCos are identified, manually or automatically, minor identification errors are inevitable. Some tolerance of inaccuracy in CoCo identification is necessary in the framework.

This model is suitable for dynamic business environments in which sensitive resources are continuously developed by a dynamic set of users, and are stored on a centralized server. It is very difficult to set up access control in these environments. Software development companies are good examples of our targeted environment. For example, a Microsoft employee downloaded confidential information unrelated to his job in 2009[9]. Through the brief introduction of CoCo-based risk estimate of access requests, we conclude that this framework can stop information thefts like the Microsoft incident.

7. Conclusion

Many theoretical access control models purport to provide fine-grained access control policies, but few can be put into practice. The main practical problem for these theoretical access control models is that administrators lack knowledge of personnel, business processes, and resources, and automatic mechanisms for capturing this information do not exist. Automatically recovering this information is very important for improving the effectiveness of current access control systems. In this paper, we explored a new way of recovering this information. We showed how to identify a user's work relationships with other users through analyzing his email.

Many theoretical access control models purport to provide fine-grained access control policies, but few can be put into practice.

In terms of extracting work relationships, email social networks are better than other social networks such as IM or the phone network. Many people have multiple email accounts and are likely to use one for work. This natural separation in email between work and for personal use makes it easier to extract work relationships. This contrasts with mobile phones, where one phone is increasingly used for both work and personnel conversations. Land phone within an organization would also be a good source to reveal employees' work relationships if its logs exist, because it is easy to filter all calls which are not internal to the organization.

We designed a simple algorithm to identify a user's task groups, which we refer to as Communities of collaboration (CoCos). We performed preliminary experiments with three users' email archives to demonstrate that task groups could be extracted and showed how knowledge of these groups could be applied to access control. We then compared CoCos and three alternative access control models in terms of how users are grouped to assign the same set of permissions. The similarity implies that CoCos could be used to help

administrators obtain business knowledge and improve the effectiveness of current access control systems. A CoCo-based access control model could also be developed, which would be valuable in business environments where collaborations are continuously forming and dissolving. We briefly discussed the possibility of both applications. We are currently engaged in research to verify the feasibility of these applications.

References

[1] E. Bertino, S. Calo, and H. Chen. Some usability considerations in access control systems. In *Symposium on Usable Privacy and Security (SOUPS) 2008*, 2008.

[2] P. O. Boykin and V. P. Roychowdhury. Leveraging social networks to fight spam. *Computer*, 38(4):61-68, 2005.

[3] J. Diesner, T. L. Frantz, and K. M. Carley. Communication networks from the enron email corpus "it's always about the people. enron is no different". *Computational and Mathematical Organization Theory*, 11(3):201-228, 2005.

[4] V. Gligor. Characteristics of role-based access control. In *RBAC '95: Proceedings of the first ACM Workshop on Role-based access control*, page 10, New York, NY, USA, 1996. ACM.

[5] L. Johansen, M. Rowell, K. Butler, and P. McDaniel. Email communities of interest. In *In Conference on Email and Anti-Spam (CEAS07)*, 2007.

[6] 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, SACMAT '03*, 2003.

[7] E. Messmer. Access control problems highlighted. *Network World,* February 7 2008. http://news.techworld.com/security/11359/access-control-problems-highlighted/.

[8] S. Oh and S. Park. Task-role-based access control model. *Inf. Syst.*, 28:533-562, September 2003.

[9] PCWorld. Microsoft charges employees with spying, 2009. http://www.pcworld.com/article/158652/microsoft_charges_employee_with_spying.html.

[10] R. Rowe, G. Creamer, S. Hershkop, and S. J. Stolfo. Automated social hierarchy detection through email network analysis. In *WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 109-117, 2007.

[11] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. Role-based access control models. *Computer*,

29:38-47, February 1996.

[12] J. Schlegelmilch and U. Steffens. Role mining with orca. In *SACMAT '05: Proceedings of the tenth ACM symposium on Access control models and technologies*, pages 168-176, New York, NY, USA, 2005. ACM.

[13] R. K. Thomas. Team-based access control (tmac): a primitive for applying role-based access controls in collaborative environments. In *Proceedings of the second ACM workshop on Role-based access control, RBAC '97*, New York, NY, USA, 1997. ACM.

[14] J. R. Tyler, D. M. Wilkinson, and B. A. Huberman. Email as spectroscopy: automated discovery of community structure within organizations. *Communities and technologies*, pages 81-96, 2003.

[15] J. Vaidya and V. Atluri. Roleminer: mining roles using subset enumeration. In *In CCS 006: Proceedings of the 13th ACM conference on Computer and communications security*, pages 144-153, 2006.

[16] J. Vaidya, V. Atluri, and Q. Guo. The role mining problem: Finding a minimal descriptive set of roles. In *In Symposium on Access Control Models and Technologies (SACMAT)*, pages 175-184, 2007.

[17] S. Vanamali. Role engineering: the cornerstone of role-based access control, 2008.

[18] L. Wang, D. Wijesekera, and S. Jajodia. A logic-based framework for attribute based access control. In *FMSE '04: Proceedings of the 2004 ACM workshop on Formal methods in security engineering*, pages 45-55, New York, NY, USA, 2004. ACM.

[19] D. Zhang, K. Ramamohanarao, and T. Ebringer. Role engineering using graph optimisation. In *SACMAT '07: Proceedings of the 12th ACM symposium on Access control models and technologies*, pages 139-144, New York, NY, USA, 2007. ACM.

NOTICES

The information in this publication could include typographical errors or technical inaccuracies, and CA, Inc. ("CA") and the authors assume no responsibility for its accuracy or completeness. The statements and opinions expressed in this publication are those of the authors and are not necessarily those of CA.

Certain information in this publication may outline CA's general product direction. However, CA may make modifications to any CA product, software program, service, method or procedure described in this publication at any time without notice, and the development, release and timing of any features or functionality described in this publication remain at CA's sole discretion. CA will support only the referenced products in accordance with (i) the documentation and specifications provided with the referenced product, and (ii) CA's then-current maintenance and support policy for the referenced product. Notwithstanding anything in this publication to the contrary, this publication shall not: (i) constitute product documentation or specifications under any existing or future written license agreement or services agreement relating to any CA software product, or be subject to any warranty set forth in any such written agreement; (ii) serve to affect the rights and/or obligations of CA or its licensees under any existing or future written license agreement or services agreement relating to any CA software product; or (iii) serve to amend any product documentation or specifications for any CA software product.

Any reference in this publication to third-party products and websites is provided for convenience only and shall not serve as the authors' or CA's endorsement of such products or websites. Your use of such products, websites, any information regarding such products or any materials provided with such products or on such websites shall be at your own risk.

To the extent permitted by applicable law, the content of this publication is provided "AS IS" without warranty of any kind, including, without limitation, any implied warranties of merchantability, fitness for a particular purpose, or non-infringement. In no event will the authors or CA be liable for any loss or damage, direct or indirect, arising from or related to the use of this publication, including, without limitation, lost profits, lost investment, business interruption, goodwill or lost data, even if expressly advised in advance of the possibility of such damages. Neither the content of this publication nor any software product or service referenced herein serves as a substitute for your compliance with any laws (including but not limited to any act, statute, regulation, rule, directive, standard, policy, administrative order, executive order, and so on (collectively, "Laws") referenced herein or otherwise or any contract obligations with any third parties. You should consult with competent legal counsel regarding any such Laws or contract obligations.