Detecting Insiders Threat in Collaboration Information Systems through Access Log Information

SATHIYAVATHI.S[1], PG Student

Computer Science and Engineering

University College of Engineering

Trichirappalli.

Email:Id:sathya.sindhu90@gmail.com

JAYAMALA.R[2],

Asst. Professor

Department of CSE

University College of Engineering

Trichirappalli.

Abstract

Collaborative information systems (CIS) enable users to coordinate efficiently over shared tasks. They are often deployed in complex dynamic systems that provide users with broad access privileges, but also leave the system vulnerable to various attacks. Techniques to detect threats originating from beyond the system are relatively mature, but methods to detect insider threats are still evolving. A promising class of insider threat detection models for CIS focus on the communities that manifest between users based on the usage of common subjects in the system. In our proposed system, ABE is introduced to achieve the privacy in maintaining the personal health record of the subject. This Personal health record (PHR) is going to implement an emerging patient-centric model of health information exchange, which is often outsourced to be stored at a third party, such as cloud providers. Yet, issues such as risks of privacy exposure, scalability in key management, flexible access and efficient user revocation, have remained the most important challenges toward achieving fine-grained, cryptographically enforced data access control. Our system can easily detect anomaly based on the deviation occurring the user's access pattern.

1. INTRODUCTION

Collaborative information systems (CISs) allows group of users to communicate and cooperate over common tasks in order to make a decision. They have long been called upon to support and coordinate activities related to the domain of "computer supported and cooperative work" (Benaben et al., 2006; George et al., 1990). On the Internet, for instance, the notion of CIS can be typified in wikis, video conferencing, document sharing and editing, as well as dynamic bookmarking (Gruber et al., 2007). In addition, hospitals have adopted electronic health record (EHR) systems which is used to decrease healthcare costs, strengthen care provider productivity, and increase patient safety (Menachemi and Brooks, 2008) using vast quantities of personal medical data various approaches have been developed in order to address the insider threats.

It is believed that the greatest security threat to information systems can be origin from insiders (Probst et al., 2006; Schultz et al., 2002; Stolfo, 2008; Tuglular and Spafford, 1997). Recognizing that access control is necessary, but it is not sufficient to guarantee protection, the anomaly detection methods have been proposed to detect deviation from expected behavior in collaborative environments. In this paper, we introduce a framework to detect anomalous insiders from the access logs of a CIS by leveraging the relational nature of system users as well as the meta information of the subjects accessed.

Users who are all found to deviate significantly from expected behavior are considered to be anomalous. To accomplish this assessment, projecting the users onto the spectrum of communities to measure the distance between each user and their neighbors in the network. The greater the distance between the user and their neighbors, will defined that the greater the likelihood that the user is anomalous.

There are some specific contributions of this work.

- **Centric model**. Service allows the patient to create, manage and control their personal health data in one place through the web, which has made the storage, retrieval and

sharing of the medical information .This paper propose a new patient-centric framework and a suite of mechanisms for data access control to PHRs stored in semi-trusted servers. It consists of three major participants such as PHR owner, Users and Cloud server.

- **Central Authority (CA)**. There are multiple owners who may encrypt according to their own ways, possibly using different sets of cryptographic keys. So that each user can obtain the keys from every owner.
- **Attribute Based Encryption (ABE)**. Access policies are expressed based on the attributes of users or data, which enables the patient to selectively share her PHR among a set of users by encrypting the file under a set of attributes, without the necessity to know the complete list of users.

This paper is organized with CADS and the extension of MetaCADS, and the specific community extraction and anomaly detection are described in section 2, which is as follows.

2. MetaCADS

The MetaCADS is the extension of CADS is called as *community anomaly detection system* which is a framework.

2.1 Community Extraction

The community extraction infers the relationships observed between users and the subject's records in the CIS access logs.

2.1.1 Network Construction

This network model consists of three important key roles such as users, subjects and categories. One user can access many subjects and constructs the tree based on graph based modeling over the accessed subject's .More than one user can access the same subjects. This graph is constructed from access logs of the every user.

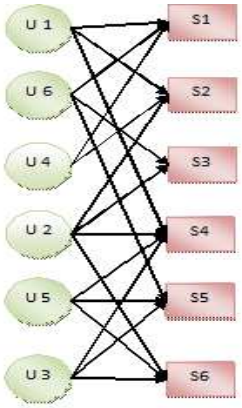

Fig.1.Example Network Construction

2.2 Detection of Anomaly

Anomaly detection is also known as Outlier detection. These outliers are referred to as abnormal behavior in a network. These abnormal behaviors are deviated largely when compared with its neighbor in the network. Deviation is measured by means of nearest neighbor radius, through these deviation measurements anomalous insiders can be detected in the CIS system. The normal users are likely to exhibit significantly smaller radius deviation scores than abnormal users. Meta-CADS exhibit larger deviations than CADS in the large user's environment.

2.2.1 Nearest Neighbor Discovery

K-means Nearest Neighbor classifier is used to detect each user behaviors in Collaborative information system based on access logs of the each user. This behavior is mentioned as the radius of user. By comparing the radius of the each user to the neighbor

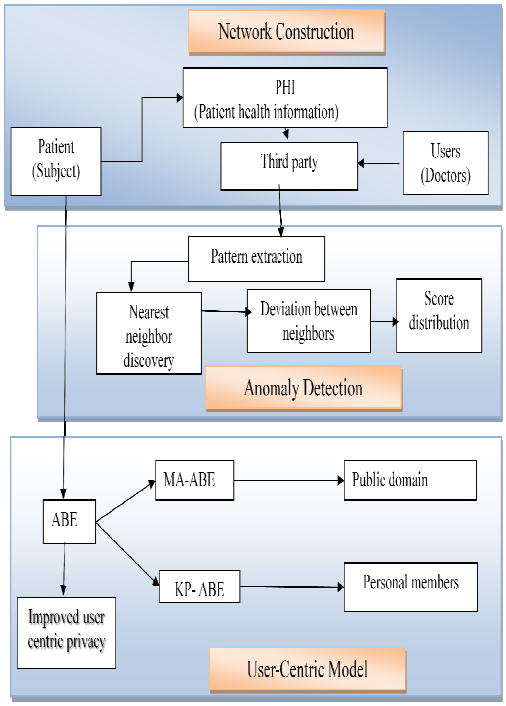

Fig.2. An architectural overview of User-Centric model

One, deviated neighbor can be detected. In CADS system, it is difficult task to find the anomalous behavior by only considering radius. Therefore, anomalous insiders can be caught by means of deviation. The distance can be calculated as,

$$DIS\ (i, j) = \sqrt{\qquad \lambda_q(Z(q,i) - Z(q,j))^2 / \qquad}$$

From this equation, we determine an appropriate value of *k*. Then the distance can be calculated using this value of *k.*

Algorithm 1. Minimization of the network community
Profile.
Input: *DIS*, a distance matrix

Output: *k*, the number of nearest neighbors

1: $k \leftarrow |U|$ {Initialize to all possible neighbors}
2: for $i = 1$ to $|U|$ do

3: $N = \{\}$

4: for $j = 1$ to $|U|$ do

5: $N \leftarrow N \cup i - nn_j$
{the *i*-nearest neighbor network for user u_j}

6: end for

7: for $j = 1$ to $|U|$ do

8: if $\psi(gj,N,i) < k$ then

9: $k \leftarrow i$ {the conductance function}

10: end if

11: end for

12: end for

2.2.2 Determining Deviation from Nearest Neighbors

Anomalous users cannot be detected through radius alone and hence we need to calculate the deviation of a node's radius from its *k* -nearest neighbors to measure the degree to which it is anomalous.

Calculating the deviation of their radius for a user as

$$Dev(u_i) = \sqrt{\sum u_j \in knni\ (rj - r)^2\ / k - 1}$$

Where $r = \sum u_j \in knni\ rj / k$

From this equation we can compute the deviation score which is that normal users are likely to exhibit significantly smaller radius deviation scores than abnormal users. This deviation score is used to determine the deviation among the users in the network.

2.3 Achieving privacy using ABE

There are two types of ABE techniques are used such as MA-ABE and KP-ABE.MA-ABE stands for Multi Authority Attribute Based Encryption in public domain. The MA-ABE scheme allows any polynomial number of independent authorities to monitor attributes and distribute secret keys. It is used in public domain to improve the security and avoid key computation problem. KP-ABE stands for Key Policy Attribute Based Encryption. They make access to PHRs based on access rights assigned by the PHR owner. The PHR owner uses a KP-ABE system to manage the secret keys and access rights of users in their Personal Domain. The insiders try to view the patient information; they need to enter an appropriate key. If the key is wrong automatically an alert message is sent to the particular owner. It is easy to identify the anomalous insiders for the first attempt itself.

2.4 Patient Centric Management

The Patient Centric management is introduced because of the security threat, therefore information of the each patients (subject) are maintained by themselves rather than health care centre. It is achieved by means of two algorithms called MA-ABE (Multi Authority Attribute Based Encryption) and KP-ABE (Key Policy Attribute Based Encryption). This will analyze the security of the proposed PHR sharing solution. It achieves data confidentiality by proving the enhanced MA-ABE scheme.

3 RELATED WORKS

In general, security mechanisms are of two types which are designed to address the insider threat. The first is to prevent the illegal behavior by modeling access rules for the system and its users. The second is to detect the illegal activity post hoc by reviewing patterns of user behavior. In this section we review prior research in these areas and relate them to the needs and challenges of CIS. We identified that information leakage may be revealed when information is shared between organizations; in which case trusted computing and digital rights management frameworks may be feasible solutions. However, in this paper, our attention is on the threats posed by authenticated individuals in a single organization.

3.1 Insider Threat Prevention

Formal access control frameworks are designed to indicate how resources in a system are made available to legitimate users. Most access control frameworks settle on if a request to the system is permitted based on a set of predefined rules. Access control frameworks have been comprehensive to deal with complex workflows by accounting for teams (Georgiadis et al., 2001), tasks (Park et al., 2001; Thomas and Sandhu 1997), and contextual cues (Peleg et al., 2008). These frameworks believe that the system is static and can be obviously modeled, but the dynamic nature of modern CIS make it complicated to apply these principles in such a situation. Additionally, collaborative systems necessitate a much broader definition of context, and the nature of collaboration cannot always be simply partitioned into tasks related with usage counts.

Experience-based access management (EBAM) (Gunter et al., 2011) is a prospective way to describe the nature of the modern organizations and the goal is to progress an access control configuration based on patterns extracted from the system's audit logs. Moreover, we desire to note that access control and role engineering is more complicated by the fact that not all users are equally trustworthy. Based on this surveillance, there have been some investigations for combining trust management models with access control frameworks (Chakraborty and Ray, 2006; Cheng et al., 2007; Crampton and Huth, 2010; Lee and Yu, 2009). These approaches allocate users to role based on their level of trust. At present, there is some evidence regarding how such approaches can be applied in real system. These models are require complex calculations and may utilize more resources than available.

In many cases, access control systems make available users with the opportunity to "break-the-glass" even though they do not have sufficient access rights. However, this approach is only possible when the number of broken glass instances (i.e., policy exceptions) is comparatively small. However, there is an evidence to recommend that the complexity of CIS, such as EHRs, result in broken glass as the standard, rather than the exception. As an example, we refer to a break-the-glass model which was shown in a consortium of hospitals in the Central Norway Health Region (Probst et al., 2006). In this case, users were assigned to an initial set of access rights and could invoke break-the-glass. However, in this study, users accessed more or less 54 percent of 99,352 patients' records through break-the-glass in a particular month and 43 percent of the 12,258 users invoked the right. Overall more than 295,000 break-the-glass instances were recorded. Obviously, this is more cases than an administrator can assess and indicates that automated auditing strategies are still essential.

3.2 Insider Threat Detection

The preceding set of approaches struggle to define "zones" in which a user can access and act upon subjects in a system. However, users can hand over illicit behaviors in the zones in which they are unrestricted to function. In this case, there are generally two classes of malicious insiders (Stolfo et al., 2008) : 1) masqueraders and 2) traitors. The masqueraders are the most well-known example of an insider. They have little knowledge of the system and they have predictable behavior. They may be a user that searches for information to utilize or they may be users whose accounts have been negotiated. On the other hand traitors have absolute knowledge of the system and its policies. A traitor may show normal behavior and still be responsible for malicious acts.

The problem focused in this paper is similar to that of detecting masqueraders. A number of prominent approaches have been anticipated to address this type of intruder. The first one is nearest neighbor anomaly detection techniques (Liao and

Vemuri, 2002; Pokrajac et al., 2007; Sun et al., 2005; Tang et al., 2002), which can be used to measure the distances between instances by evaluating their relationship to "close" instances. If the instance is not satisfactorily close, then it may be classified as an anomaly. However, social structures in a CIS are not unambiguously defined and have to be inferred from the exploitation of system resources. If distance measurement procedures are not adjusted to the way in which social structures have been constructed, the distances will not correspond to the structures well therefore detecting anomalous insiders are difficult.

The second approach is based on spectral anomaly detection which estimates the principal components from the covariance matrix of the training dataset of "normal" events. The testing segment involves the comparison of each point with the components and an anomaly score can be assigned based on the point's distance. This model can reduce noise and redundancy, however, collaborative systems are team oriented, which can decrease the performance of the model. The detection of traitors is a challenging task because it requires the discovery of subtle and major changes from a user's normal activities. Yet, this is an area grown for new research and a number of approaches have been recently proposed to deal with this type of insider threat (Boxwala et al., 2011; Chen et al., 2011; Kim et al., 2011).

4 CONCLUSIONS

In this paper, we proposed CADS, a *community anomaly detection system* to detect anomalous insiders in a CIS that utilizes a relational framework. To envisage which users are anomalous, CADS calculates the deviation of users based on their nearest neighbor networks. Considering partially trustworthy third party to fully realize the patient-centric concept, patients shall have complete control of their own privacy through encrypting their health records files to allow fine-grained access. The framework addresses the unique challenges brought by multiple owners and users, in that it greatly reduces the complexity of key management while enhance the privacy guarantees compared with previous works. Additionally, we propose patient centric and ABE (Attribute Based Encryption) to achieve privacy, therefore avoids the problem of security threat to the patient information. Our model is based on the surveillance that "normal" users tend to form communities, unlike illicit insiders.

REFERENCES

[1] F. Benaben, J. Touzi, V. Rajsiri, and H. Pingaud, "Collaborative Information System Design," Proc. Int'l Conf. Assoc. Information and Management, pp. 281-296, 2006.

[2] J. George, G. Easton, J. Nunamaker, and G. Northcraft, "A Study of Collaborative Group Work with and without Computer-Based Support," Information Systems Research, vol. 1, no. 4, pp. 394-415, 1990.

[3] S. Chakraborty and I. Ray, "TrustBac: Integrating Trust Relationships into the RBAC Model for Access Control in Open Systems," Proc. 11th ACM Symp. Access Control Models and Technologies, pp. 49-58, 2006.

[4] N. Menachemi and R. Brooks, "Reviewing the Benefits and Costs of Electronic Health Records and Associated Patient Safety Technologies," J. Medical Systems, vol. 30, no. 3, pp. 159-168, 2008.

[5] C. Probst, R.R. Hansen, and F. Nielson, "Where Can an Insider Attack?," Proc. Workshop Formal Aspects in Security and Trust, pp. 127-142, 2006.

[6] E. Schultz, "A Framework for Understanding and Predicting Insider Attacks," Computers and Security, vol. 21, no. 6, pp. 526-531, 2002.

[7] S. Stolfo, S. Bellovin, S. Hershkop, A. Keromytis, S. Sinclair, and S.W. Smith, Insider Attack and Cyber Security: Beyond the Hacker. Springer, 2008.

[8] T. Tuglular and E. Spafford, "A Framework for Characterization of Insider Computer Misuse," Unpublished paper, 1997.

[9] C. Georgiadis, I. Mavridis, G. Pangalos, and R. Thomas, "Flexible Team-Based Access Control Using Contexts," Proc. Sixth ACM Symp. Access Control Models and Technologies, pp. 21-27, 2001.

[10] J. Park, R. Sandhu, and G. Ahn, "Role-Based Access Control on the Web," ACM Trans. Information System Security, vol. 4, no. 1, pp. 37-71, 2001.

[11] R. Thomas and S. Sandhu, "Task-Based Authorization Controls (TBAC): A Family of Models for Active and Enterprise-Oriented Autorization Management," Proc. IFIP 11th Int'l Conf. Database Security, pp. 166-181, 1997.

[12] M. Peleg, D. Beimel, D. Dori, and Y. Denekamp, "Situation-Based Access Control: Privacy Management via Modeling of Patient Data Access Scenarios," J. Biomedical Informatics, vol. 41, no. 6, pp. 1028-1040, 2008.

[13] C. Gunter, D. Liebovitz, and B. Malin, "Experience-Based Access Management: A Life-Cycle Framework for Identity and Access Management Systems," IEEE Security and Privacy Magazine, vol. 9, no. 5, pp. 48-55, Sept./Oct. 2011.

[14] T. Gruber, "Collective Knowledge Systems: Where the Social Web Meets the Semantic Web," J. Web Semantics, vol. 6, no. 1, pp. 4-13, 2007.

[15] P. Cheng, P. Rohatgi, C. Keser, P.A. Karger, and G.M. Wagner, "Fuzzy Multi-Level Security: An

Experiment on Quantified Risk- Adaptive Access Control," Research Report RC24190 (W0702-085), IBM, 2007.
[16] J. Crampton and M. Huth, Towards an Access-Control Framework for Countering Insider Threats. Springer, 2010.
[17] D. Pokrajac, A. Lazarevic, and L. Latecki, "Incremental Local Outlier Detection for Data Streams," Proc. IEEE Symp. Computational Intelligence and Data Mining, pp. 504-515, 2007.
[18] Y. Liao and V.R. Vemuri, "Use of k-Nearest Neighbor Classifier for Intrusion Detection," J. Computer Security, vol. 21, no. 5, pp. 439-448, 2002.
[19] A. Lee and T. Yu, "Towards a Dynamic and Composable Model of Trust," Proc. 14th ACM Symp. Access Control Models and Technologies, pp. 217-226, 2009.
[20] J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos, "Neighborhood Formation and Anomaly Detection in Bipartite Graph," Proc. IEEE Fifth Int'l Conf. Data Mining, pp. 418-425, 2005.
[21] J. Tang, Z. Chen, A. Fu, and D. Cheung, "Enhancing Effectiveness of Outlier Detections for Low Density Patterns," Proc. Sixth Pacific- Asia Conf. Knowledge Discovery and Data Mining, pp. 535-7548, 2002.
[22] A.A. Boxwala, J. Kim, J.M. Grillo, and L.O. Machado, "Using Statistical and Machine Learning to Help Institutions Detect Suspicious Access to Electronic Health Records," J. Am. Medical Informatics Assoc., vol. 18, pp. 498-505, 2011.s
[23] Y. Chen, S. Nyemba, W. Zhang, and B. Malin, "Leveraging Social Networks to Detect Anomalous Insider Actions in Collaborative Environments," Proc. IEEE Ninth Intelligence and Security Informatics, pp. 119-124, 2011.
[24] J. Kim, J. Grillo, A. Boxwala, X. Jiang, R. Mandelbaum, B. Patel, D. Mikels, S. Vinterbo, and L. Ohno-Machado, "Anomaly and Signature Filtering Improve Classifier Performance for Detection of Suspicious Access to EHRs," Proc. Ann. Symp. Am. Medical Informatics Assoc., pp. 723-731, 2011.