Specializing Network Analysis to Insider Threat Detection

M. kumutha
PG Student M.Tech, CSE
Dr.M.G.R Educational and Research Institute University, Chennai.

Dr. V.N Rajavarman
Department of Computer Science
Dr.M.G.R Educational and Research Institute University Chennai.

Abstract:

Collaborative information systems (CIS) are deployed within a diverse array of environments, ranging from the Internet to intelligence agencies and also to healthcare. Such systems are applied to manage sensitive information, making them target for malicious insiders. We Introduce a Meta community-based anomaly detection system (META- CADS), an unsupervised learning framework to detect insider threats based on information recorded in the access logs of collaborative environments. We have done some enhancements, not only detect the anomalous but also protect the health records globally and locally. After the anomalous being detected and also repaired it is finally displayed with the Original format of the health records.

Keyword: collaborative information system, global and local security, Healthcare, Threats.

1. Introduction:

Collaborative information systems (CIS) are deployed within a diverse array of environments, ranging from the Internet to intelligence agencies to healthcare. It is increasingly the case that such systems are applied to manage sensitive information, making them targets for malicious insiders. While sophisticated security mechanisms have been developed to detect insider threats in various file systems, they are neither designed to model nor to monitor collaborative environments in which users function in dynamic teams with complex behavior. In this paper, we introduce a community-based anomaly detection system (CADS), an unsupervised learning framework to detect insider threats based on information recorded in the access logs of collaborative environments. CADS are based on the observation that typical users tend to form community structures, such that users with low affinity to such communities are indicative of anomalous and potentially illicit behavior. The model consists of two primary components: relational pattern extraction and anomaly detection. For relational pattern extraction, CADS infers community structures from CIS access logs, and subsequently derives communities, which serve as the CADS pattern core. CADS then use a formal statistical model to measure the deviation of users from the inferred communities to predict which users are anomalies. To empirically evaluate the threat detection model, we perform an analysis with six months of access logs from a real electronic health record system in a large medical center, as well as a publicly available dataset for replication purposes. The results illustrate that CADS can distinguish simulated anomalous users in the context of real user behavior with a high degree of certainty and with significant performance gains in comparison to several competing anomaly detection models.

2.Access Log Data Set:

Star Panel is a longitudinal electronic patient chart developed and maintained by the Department of Biomedical Informatics faculty working with staff in the Informatics Center of the Vanderbilt University Medical Center. Star Panel is ideal for this study because it aggregates all patient data as fed into the system from any clinical domain and is the primary point of clinical information management.

When possible, the logs are embellished with diagnostic billing codes assigned to the patient after the visit to their healthcare provider.

3. Network Construction:

The amalgamation of the user-subject access network and the subject-category assignment network. In the former, an edge represents that a user accessed the subject's record. In the latter, an edge represents that the subject's record is assigned to a particular category.

Prior research in social network analysis suggests it is important to represent the affinity that a user has toward a particular subject when assessing the similarity of a group. There are various aspects of a user's relationship to subjects that could be leveraged for measuring similarity.

4. Complex Category Interface:

In prior anomaly detection models, communities are based on the access network at one time only. This is appropriate when the set of users in the system is static and collaborate over distinct subjects. However, in a CIS, the set of users (e.g., care providers) and subjects (e.g., patients) are constantly rotating through the system and represent a varying set of semantic categories (e.g., diagnoses). Thus, anomaly detection should account for the dynamic nature of the system and the semantics of the subjects.

5. Community Interface:

To infer user communities, CADS performs a spectral decomposition on a relational model of the users, which Meta CADS extends to include complex categories. In preparation for the decomposition, CADS builds a matrix which is based on the access network. By contrast, Meta CADS extends the model to incorporate the assignment network, where such that the i^{th} row is the projection of user UI over the relational system.

6. Measuring Deviation From Nearest Neighbors:

Anomalous users cannot be detected through radius alone and direct application of such a measure can lead to undesirable results. Consider, in below fig user uy and the users in cluster F can be correctly classified as anomalous based on their radius. In contrast, we would fail to detect ux as an anomaly because it has a smaller radius in comparison to nodes in the F area. This bias is due to a reliance on raw magnitudes and thus we normalize the system. Rather than use raw radius, we calculate the deviation of a node's radius from those of its k-nearest neighbors to assess the degree to which it is anomalous.

7. Technique Used: MNCP (Minimization of the Network Community Profile)

To search for the k-nearest neighbors (KNNs) of a user, we adopt a modified Euclidean distance. This measure weights the principal components proportionally to the amount of variance they cover in

the system. These distances are stored in a matrix DIS of size jUj _ jUj, where indicates the distance between ui and uj. Using this measure, we determine an appropriate value for k. This is accomplished by leveraging the network community profile (NCP), a characterization of community quality based on its size. In particular, k is set to the value that minimizes NCP as defined.

Algorithm 1. Minimization of the network community profile
Input: *DIS*, a distance matrix
Output: *k*, the number of nearest neighbors

```
1: k ← |U| {Initialize to all possible neighbors}
2: for i = 1 to |U| do
3:     N = {}
4:     for j = 1 to |U| do
5:         N ← N ∪ i − nn_j
           {the i-nearest neighbor network for user u_j}
6:     end for
7:     for j = 1 to |U| do
8:       if ψ(g_j, N, i) < k then
9:           k ← i {the conductance function}
10:      end if
11:    end for
12: end for
```

Fig 1.1 Algorithm

8. Global and local attacks

We aim to design models to integrate our approach with others in the future. We intend on parameterzing such models based on local, rather than global observations. We aim to design models to integrate our approach with others in the future.

1. After detecting the anomalous it will be checked and cleared.
2. Correct form of electronic health record will be generated.
3. Intend on such models based on local and global observations
4. Aim to design models to integrate our approach with others in the future.

The anomaly that the patient test name and medicine and reason for visit get interchanged is rectified. The correct form of Electronic Health Record, without any anomaly, is generated.

Protecting the EHR from intrusion like globally (hackers from outside the organization) and locally people who working inside the organization

9. Local Parameters

A doctor or a person can access the record of a particular patient only on his record being authenticated by the administrator. If his record does not exist the access will be denied and as a message “doctor is not working in this medical centre” will be sent to the concern person. Unless the doctor makes authorized log in, despite his working there, the access to the record of a particular patient whom the doctor treats, will be denied and even the doctor himself cannot access.

Once the two constraints namely, the doctor must be working there and that doctor should have made the authorized login at the particular time, being got cleared, the access to that particular patient record is made possible

10. Global Parameter

In the event of an unauthorized person trying to access the record of particular patient, the CIS not only prevent s the unauthorized entry but also sent message as “unauthorized person cant access”. CIS alerts the administrator as outsiders tries to access the record and administrator send the message as “the

electronic health record is secured so unauthorized person cant access".

11. Conclusion:

In this paper , proposed networks, not only detect the anomalous but also protect the health records globally and locally. After the anomalous being detected and also repaired it is finally displayed with the Original format of the health records

12. References:

[1] Jason Crampton and Michael Huth, Towards an Access-Control Framework for Countering Insider Threats, Advances in Information Security Volume 49, 2010, pp 173-195.

[2] Toby Cook, Associate Partner,. Stopping insider attacks how organizations can protect their sensitive information, IBM Center for Business Michel Bobillier, Global Offering Executive 2006.

[3] Deborah L. McGuinness, Honglei Zeng, Paulo Pinheiro da Silva, Li Ding, Dhyanesh Narayanan, Mayukh Bhaowal Investigations into Trust for Collaborative Information Repositories: A Wikipedia Case Study, 2006

[4] M. Alawneh and I. Abbadi, "Preventing Information Leakage between Collaborating Organizations," Proc. 10th Int'l Conf. Electronic Commerce, pp. 185-194, 2008.

[5] Christos K. Georgiadis Ioannis Mavridis George Pangalos Roshan K. Thomas, Flexible Team-based Access Control Using Contexts, 2001

[6] Mohamed Tamer Refaei, Mohamed Eltoweissy, Adaptation of Reputation Management Systems to Dynamic Network Conditions in Ad Hoc Networks, 2010

[7] John Felix Charles Joseph, Amitabha Das, Cross-Layer Detection of Sinking Behavior in Wireless Ad Hoc Networks Using SVM and FDA. 2011.

[8] You Chen, Steve Nyemba, Bradley Malin, Detecting Anomalous Insiders in Collaborative Information Systems, IEEE Transactions on dependable and secure computing, 2012 vol. 9 no. 3, pp. 332-344.