Detecting Suspicious Access and Anomalous Insiders

When it comes to securing private data such as electronic medical records (EHRs), it is

important to not only implement preventative measures but to also have the ability to detect users

trying to access the data in malicious or otherwise inappropriate ways. According to Boxwala et

al., any usage that doesn’t fall under treatment, payment, and healthcare operations (TPO) is

considered inappropriate. They cite the example of users looking up the records of family

members or celebrities when the user is not directly involved in the care of those individuals.

More broadly, while looking at collaborative information systems (CIS) such as EHRs, Chen et

al. define authentic users that perform any actions contrary to CIS policy as suspicious. In their

paper, Boxwala et al. proposed to use linear regression and SVM machine learning models to

detect suspicious access events. They generated a feature database with 26 features believed

would be useful in predicting whether an access was appropriate. They trained the two different

models on a dataset generated with the help of security officers. Each access record in the set

was first manually classified in order to produce a standard to measure performance against.

Final evaluation was performed using a set of 58 suspicious accesses. Since all accesses in the

final evaluation set were suspicious, performance was evaluated using sensitivity alone. The

AUCs for both models were above 0.9, with 1 being the best possible score. The sensitivity of

both methods were above 75% when the cutoff probability value was set at 0.5, while their

baseline metric failed to detect any suspicious accesses. This suggests that the methods used

provide some advantage in detecting inappropriate accesses. It was noted by the authors that the

results may be slightly skewed due to the oversampling of suspicious accesses in the training set

and also biased towards suspicious accesses familiar to the security officers providing the

classification standard. Even so, the methods used above can be used to prioritize accesses for

further manual investigation, helping to make the best use of a security officer’s time. To detect

suspicious users in EHR systems, Chen et al. implemented the CADS and MetaCADS systems,

which create communities of users based on access logs. The premise is to detect suspicious

users based on their deviations from normal user patterns and networks—the distance from their

neighbors. CADS has two different parts: one to detect patterns between users and generate the

community structure, and a second to find suspicious users by calculating the distances between

users. MetaCADS adds another layer to this process by incorporating an additional assignment

network that maps users to categories of records, not just a network of accesses. To evaluate

CADS and MetaCADS, these approaches were compared to other established metrics (KNN,

PCA, and HVU) in three settings: a single simulated user accessing a range of records, a range of

simulated users with a constant number of records accesses, and a range of simulated users

accessing a range of records. Overall, CADS and MetaCADS performed better than the other

methods, although their performance against each other changed based on the number of

suspicious users in the system. The authors described some limitations on the methods, namely

that the performance of MetaCADS is variable based on the percentage of suspicious users in the

system and also that some useful information has not yet been incorporated which could improve

the results. Additionally, unlike the Boxwala et al. study, false positives were not manually

evaluated so it is possible that the algorithm actually performed better than initially reported. In

any case, the CADS method is able to leverage relational information to automatically detect

suspicious users, providing an unsupervised and extensible threat detection system.

**References**

Boxwala AA, et.al. Using statistical and machine learning to help institutions detect suspicious access to electronic health records. J Am Med Inform Assoc. 2011;18(4):498-505.

Y. Chen, et.al. Detecting anomalous insiders in collaborative information systems. IEEE Transactions on Dependable and Secure Computing. 2012; 9(3): 332-344.