www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 2 Issue 9 September 2013 Page No. 2799-2802

ANALYSIS OF VULNERABILITY IN INTERNET FIREWALL USING RULE BASED ALGORITHM

[1]T.Maheswari, [2]Mr. A. Senthil Kumar

Research Scholar,
Department of computer science, Tamil University, Thanjavur-10.
E-mail: mahimphil@gmail.com
Asst.Prof in Computer Science, Tamil University Thanjavur-10.

Abstract

Collaborative information systems (CISs) are deployed within a diverse array of environments that manage sensitive information. Current security mechanisms detect insider threats, but they are ill-suited to monitor systems in which users function in dynamic teams. The community anomaly detection system (CADS), an unsupervised learning framework to detect insider threats based on the access logs of collaborative environments. The framework is based on the observation that typical CIS users tend to form community structures based on the subjects accessed. CADS consist of two components: 1) relational pattern extraction, which derives community structures and 2) anomaly prediction, which leverages a statistical model to determine when users have sufficiently deviated from communities. We further extend CADS into Meta CADS to account for the semantics of subjects. Network security applications generally require the ability to perform powerful pattern matching to protect against attacks such as viruses and spam. Traditional hardware solutions are intended for firewall routers. However, the solutions in the literature for firewalls are not scalable, and they do not address the difficulty of an antivirus with an ever-larger pattern set. Related works have focused on algorithms and have even developed specialized circuits to increase the scanning speed.

Keywords: Common Information System, Community anomaly Detection

1.Introduction

Beyond computational support, the adoption of CIS has been spurred on by the observation that such systems can increase organizational efficiency through streamlined workflows, shave administrative costs, assist innovation through brainstorming sessions, and facilitate social engagement. On the Internet, for instance, the notion of CIS is typified in wikis, video conferencing, document sharing and editing, as well as dynamic bookmarking. At the same time, CIS are increasingly relied upon to manage sensitive information. Intelligence agencies, for example, have adopted CIS to enable timely access and collaboration between groups of analysts using data on personal relationships, financial transactions, and surveillance activities. However, at the same time, the detail and sensitive nature of the information in such CIS make them attractive to numerous adversaries. This is a concern because the unauthorized dissemination of information from

such systems can be catastrophic to both the managing agencies and the individuals (or organizations) to which the information corresponds. It is believed that the greatest security threat to information systems stems from insiders. In this work, we focus on the insider threat to centralized CIS which are managed by a sole organization.

A suspicious insider in this setting corresponds to an authenticated user whose actions run counter to the organization's policies. Various approaches have been developed to address the insider threat in collaborative environments. Formal access control frameworks, for instance, have been adapted to model team and contextual scenarios. Recognizing that access control is necessary, but not sufficient to guarantee protection, anomaly detection methods have been proposed to detect deviations from expected behavior. In particular, certain data structures based on network analysis have shown promise. Review these models in depth, but wish to highlight several limitations of these approaches up front. First, access control models assume a user's role (or their relationship to a group) is known a priori. However, CIS often violate this principle because teams can be constructed on the fly, based on the shifting needs of the operation and the availability of the users. Second, the current array of access control and anomaly detection methods tend to neglect the Meta information associated with the subjects. In this paper, 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. The framework is called the community anomaly detection system, or CADS, and builds upon the work introduced in.

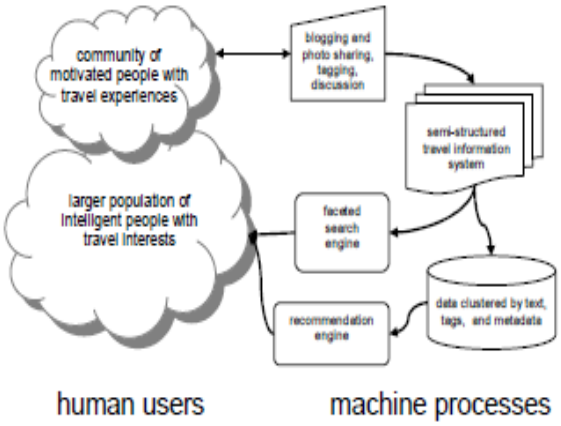
Relational patterns from access logs. A process to transform the access logs of a CIS into dynamic community structures using a combination of graph-based modeling and dimensionality reduction techniques over the accessed subjects. We further illustrate how Meta information, such as the semantics associated with subjects, can be readily integrated into the CADS framework. We call this extended framework MetaCADS.Anomaly detection from relational patterns. We propose a technique, rooted in statistical formalism, to measure the deviation of users within a CIS from the extracted community structures. Utilize a real-world data set to systematically evaluate the effectiveness of our anomaly detection framework. In particular, we study three months of real-world access logs from the electronic health record system of the Vanderbilt University Medical Center, a large system that is well integrated in the everyday functions of healthcare. In lieu of labeled anomalous users, we simulate insider threat behavior and empirically demonstrate that our models are more effective in performance than the state-of-the-art competitive anomaly detection approaches. Analysis provides evidence that the typical system user is likely to join a community with other users, whereas the likelihood that a simulated user will join a community is low. Our findings indicate the quantity of illicit insiders in the system influences which model (i.e., CADS or MetaCADS) is a more prudent solution.

2.Varying Number Of Accessed Subjects

The first set of experiments focus on the sensitivity of anomaly detection models. To begin, we mixed a single simulated user with the real users. We varied the number of subjects accessed by the simulated user to investigate how volume impacts the deviation score and the performance of the anomaly detection models in general, when the number of subjects accessed by the simulated users is small, the deviation score is low as well. However, when the number of subjects accessed is larger than 20, the deviation scores of simulated users increase significantly. This is because users with a small number of accesses do not provide sufficient information for Meta- CADS to appropriately characterize their access behaviors. Next, we set out to determine when the deviation score is sufficiently large to detect the simulated user in the context of the real users. How the number of subjects accessed by the simulated user influences the performance of the anomaly detection models. When the number of accessed subjects for the simulated user is small (e.g., one), it is difficult for all of the models to discover the user via the largest deviation score. This is expected because all of the models, except for HVU, are evidence-based. They need to accumulate a certain amount of statistical evidence before they can determine that the actions of the user are not the result of noise in the system. The performances of all models generally increase with the number of subjects accessed. However, the performance gain is relatively minor for the classification models; i.e., KNN and PCA. The false positive rate of these models is never lower than 0.4, even when the number of subjects accessed is greater than 100. By the point at which 10 subjects are accessed, HVU achieves a false positive rate of approximately 0.1 and CADS and MetaCADS are below 0.02. When the number of accessed subjects is greater than 30, HVU consistently achieves the lowest false positive rate. This is because; the majority of the real users access less than 30 subjects per day. Nonetheless, it is apparent that both MetaCADS and CADS achieve very low false positive rates when attempting to detect a single simulated user. Moreover, MetaCADS consistently achieves a smaller false positive rate than CADS. We believe this is because the assignment network facilitates a stronger portrayal of real users' communities than the access network in isolation.

3.Varying Number of Intruding Insiders

In order to assess how the number of simulated users influences the performance of the five models, we conducted several experiments when the number of simulated users was randomly generated. In these experiments, the number of subjects accessed by the simulated users was fixed at 5. We chose this number to simulate evasive maneuvering. By setting the number of subjects accessed to this level, we simulate users that attempt to avoid triggering the high volume rule. The mix rate of simulated users was varied from 0.5 to five percent. The AUC scores for the models are summarized and there are several notable observations. First, it is evident that HVU exhibits the worst performance in this setting. This is unsurprising because there are many real users that access more than five subjects in the system. Second, as in the previous set of experiments, the supervised classification models (i.e., KNN and PCA) exhibit significantly worse performance than the unsupervised relational models (i.e., CADS and MetaCADS). Third, when the number of simulated users is low (i.e., 0.5 percent), MetaCADS yields a slightly higher AUC than CADS (0.92 versus 0.91). This observation is in accordance with our results from the first experiment in which a single simulated user is mixed into the real system. However, as the number of simulated users increases, CADS clearly dominates MetaCADS. Specifically, the performance rate of CADS increases from 0.91 to 0.94, while MetaCADS decreases from 0.92 to 0.87. We believe this is because when the number of simulated users increases, they have more frequent categories in common. In turn, these categories enable simulated users to form more communities than those based on subjects along, thus lowering their deviation scores. This is an interesting observation because it suggests that if the number of intruding insiders is expected to constitute a significant number of users, the anomaly detection model will benefit from neglecting the categories associated with the accessed subjects. It can be observed that when the threshold of the deviation score is set to 0.3, most of the simulated users are detected with a low false positive rate. Depicts CADS at a mix rate of two percent, where it can be seen that a threshold of 0.6 provides relatively strong detection capability.



4.Varying Number of Simulated User and Accessed Subjects

In this experiment, we simulated an environment in which the system varied in the types of intruders to compare the anomaly detection models. Specifically, we allowed both the number of simulated users and the number of subjects accessed by the simulated users to vary. The mix rate between simulated users and the total number of users was varied between 0.5 to five percent and the number of subjects accessed per simulated user was selected at random between 1 and 150. The ROC curves of the models for three mix rates are depicted and the AUC scores are depicted. There are several findings to recognize. First, as in the previous experiment, it can be seen that the performance of the supervised classification models is significantly worse than the unsupervised models. The supervised models consistently have a lower true positive rate at all operating points. Second, unlike the previous experiment, HVU achieves comparable results to the supervised classification models. This is due to the fact that this model is correctly characterizing the intruders that access a larger number of records. Third, with respect to AUC, the same trend as earlier regarding the dominance of the unsupervised models as a function of the mix rate. Specifically, Meta CADS dominates when the mix rate is low, but CADS dominates when the mix rate is high. Notably the disparity between Meta CADS and CADS is more pronounced at the low mix rate (0.91 versus 0.88) in this setting than in the previous setting. However, at lower false positive operating points, CADS appears to dominate Meta CADS. Depict the Meta CADS and CADS deviation scores for real and simulated users as a function of the number of subjects accessed in an arbitrary day of the EHR data set. The mix rate was set to 0.5 percent for Meta CADS and two percent for CADS.

5.Role-based Access Control

With role-based access controls, access rights are grouped by role name. This approach offers significant advantages because of scalability. Each user is assigned one or more roles, and each role is assigned one or more permissions that can be given to users in that role. Users are granted membership into roles based on their competencies, credentials and responsibilities in the organization. User membership in roles can be revoked easily and new memberships established as needed. This simplifies the administration and management of permissions since roles can be updated without updating the permissions for every user on an individual basis. Moreover, the use of role hierarchies provides additional advantages since one role may implicitly include the operations that are associated with another role. A recent well known role-based approach is RBAC, which has received considerable attention as a promising way to enhance traditional mandatory and discretionary models.

6.Team–based Access Control

The TMAC model was originally proposed by Thomas in. TMAC recognized the importance of context information associated with collaborative tasks and the ability to apply this context to decisions regarding permission activation. The collaboration context of a team contains two pieces: the user context, which could be the current members (users) of a team, and the object context, which could be the set of object instances required by the team to accomplish its task. TMAC allows us to create a general structure (class/definition) of a team with role based permission assignments to object-types. However, when a team is instantiated, the user context can be used to tailor the role based permissions defined on object types to user-specific permissions on individual object instances considered to be part of a team's resources. By aligning access control to the metaphor of teams, TMAC can provide a paradigm for access control that is natural and nonintrusive to the way users work in collaborative environments. Extend the original TMAC proposal in two key directions. First, give a framework to integrate TMAC concepts with RBAC. Second, we extend TMAC to use other contextual information besides what is currently used in the user context and object context. This generalized model is referred to in the rest of the paper as C-TMAC (for context - based TMAC). Such contextual information can among others things include the time of access, the location from which access is requested, the location where the object to be accessed resides, transaction-specific values that dictate special access policies etc. This allows TMAC to model a richer set of access policies that are more closely tied to application needs.

7. Conclusions

To detect anomalous insiders in a CIS, we proposed CADS, a community anomaly detection system that utilizes a relational framework. To predict which users are anomalous, CADS calculates the deviation of users based on their nearest neighbor networks. We further extended CADS into MetaCADS to incorpate the semantics of the subjects accessed by the users. Our model is based on the observation that "normal" users tend to form communities, unlike illicit insiders. To evaluate the performance of our model, we conducted a series of experiments that compared our framework with the state-of-the-art anomaly detection methods for CIS systems. In the experiments, we mixed simulated users with the real users of a real electronic health record system. Our results illustrated that the community-based models exhibited better performance at detecting simulated insider threats. The evidence further suggested that MetaCADS is the best model when the number of intruders is relatively small, but that CADS dominates when the number of intruders increases. Since the framework is an unsupervised system, we believe it may be implemented in

real time environments with offline training. There are limitations of the system; however, and in particular, we intend to validate and improve our system with adjudication through real human experts.

8. References

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

[2] 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.

[3] Y. Chen and B. Malin, "Detection of Anomalous Insiders in Collaborative Environments via Relational Analysis of Access Logs," Proc. First ACM Conf. Data and Application Security Security and Privacy, pp. 63-74, 2011

[4] 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

[5] J. Crampton and M. Huth, Towards an Access-Control Framework for Countering Insider Threats. Springer, 2010.

[6] L. Eldenburg, N. Soderstrom, V. Willis, and A. Wu, "Behavioral Changes Following the Collaborative Development of an Accounting Information System," Accounting, Organizations and Soc.,
vol. 35, no. 2, pp. 222-237, 2010.

[7] B. Malin, S. Nyemba, and J. Paulett, "Learning Relational Policies from Electronic Health Record Access Logs," J. Biomedical
Informatics, vol. 44, no. 2, pp. 333-342, 2011.