Detection of Undesirable Insider Behavior

Joseph A. Calandrino[1⋆], Steven J. McKinney[2⋆], and Frederick T. Sheldon[3]

[1] Princeton University, Princeton, NJ 08544, USA
jcalandr@princeton.edu
[2] North Carolina State University, Raleigh, NC 27695, USA
sjmckinn@ncsu.edu
[3] Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA
sheldonft@ornl.gov

Insiders, those within or closely related to an organization, pose the greatest risk to an organization's information systems. Organizations grant insiders both authorized access to and knowledge of their information systems, primarily computer systems and the organization's network. In the past, insiders have abused this trust by stealing or corrupting data, committing fraud, and modifying performance reports [10]. Because these insiders may act within the bounds of their privileges, mitigation of the insider threat differs from that of external threats.

Undesirable insider behavior involves any willful or negligent misuse of resources in an organization's information systems. Numerous existing systems, such as firewalls and intrusion detection systems (IDSs), seek to mitigate the threat that parties external to an organization pose to its information systems. Unfortunately, these mechanisms often do not restrict or monitor insiders [1]. Little research beyond access control strives to mitigate the threat that malicious or apathetic insiders may introduce. These insiders present a particularly insidious problem as they may behave adversely without violating access control policies. A 2004 survey of security and law enforcement executives found that, among respondents that experienced e-crime or intrusions and could classify attacks as internal or external, an average of 29% of attacks against their organizations came from insiders [11]. This fact, combined with the established difficulty of the insider problem [7] and the ability of a single malicious insider to cause significant financial impact ($500 million in one case [1]), demonstrates the need for further research on insider threat detection systems.

This document describes the Intelligent Insider Threat Detection (I^2TD) system for monitoring and evaluating insider behavior to detect potentially malicious or otherwise undesirable activity. The system observes an insider's local system and network-based activities and is extensible to other aspects of the information system. A rule-based method immediately notifies administrators of activity known to be suspicious, and data mining tools regularly inspect compiled user behavior for anomalies that could indicate undesirable activity. Although the monitoring, rule-based, and data mining components of the system are all of critical importance, this document focuses on the data mining component.

Numerous systems have attempted to apply machine learning and statistical learning techniques to detect intrusions and other forms of fraud. Anderson et

⋆ This research was performed during an internship at Oak Ridge National Laboratory.

al. used a statistical approach to evaluate the deviation of present user behavior from past behavior [2], and Chan et al. applied data mining techniques to the similar problem of detecting credit card fraud [4]. [8, 9] describe work towards real-time intrusion detection using data mining. The underlying systems use learning agents to regularly mine data and extract patterns that serve as classifiers for real-time detection of intrusions. [8, 9] are complementary to this paper: additional mining algorithms could assist in refining the I^2TD system's set of real-time rules. This work attempts to derive additional utility beyond [8, 9] by considering characteristics of aggregated data.

The Minnesota Intrusion Detection System (MINDS) detects anomalous network connections based on characteristics of other connections over the same time period [5, 6]. MINDS does not detect anomalies in real time, but similar to [8, 9], it also mines for association rules to assist in real-time detection. Because MINDS is not real time, it may derive and analyze characterizes of aggregate data that allow more accurate evaluation. The data mining component of this work utilizes an approach similar to MINDS to assess user behavior.

The data mining component of the I^2TD system regularly retrieves recent user activity from a database and analyzes it for anomalies. Rather than mining the raw data, the system computes aggregate characteristics of user activity during the period with the goal of smoothing inconsequential deviations and producing more accurate results. Following aggregation, the system compares the characteristics to those of the user during past periods to compute an anomaly score for the most recent period. The system also offers additional information regarding the impact of various behavioral characteristics on the anomaly score. An analyst may use the anomaly score and additional derived data to determine whether a user's behavior over a given period warrants further exploration.

At this time, the data mining component considers seven derived characteristics: number of logins, number of host machines used, number of file opens caused, number of file deletes caused, number of unique files accessed (opened or deleted), number of unique files opened, and number of unique files deleted. The system presently operates under the assumption that deviations from typical login and file access patterns are an effective indicator of undesirable insider behavior. Various other behavior attributes may also be useful, such as network or database accesses. Because the system presently aggregates data on a daily basis and considers only a user's personal historical data, however, the sample space may be small. Given a small sample space, high dimensionality may scatter the data points too greatly to draw meaningful inferences. Thus, naive consideration of additional characteristics may have an adverse impact on analysis.

Like [5, 6], the data mining component of this system isolates anomalies using the local outlier factor (LOF) metric, developed by Breunig et al. [3]. Given an item in a data set, algorithms for LOF map the set to vector space and establish a group of n nearest neighbors for the item. The algorithms then determine how much more or less isolated the item is than its nearest neighbors and use the ratio to produce the item's LOF score. If the item is part of an evenly dispersed cluster, the LOF for the item will be close to one. An LOF far greater than

one may occur if the item's nearest neighbors are relatively far away yet those neighbors are members of tight clusters.

To determine whether a user's behavior on a given day is anomalous, the system examines the day's activity record in the context of previous days' records and computes an LOF score for the day. Prior to this step, however, the system computes the standard deviation of each characteristic in previous days' records and uses the results to normalize all characteristics of the records. Because the LOF algorithms look at the mapping of records to vector space, this normalization step prevents characteristics with greater deviation from having a disproportionate impact on the Euclidean distance between records.

The rationale behind the use of LOF as opposed to a single cluster or multiple clusters is that user tasks may change from day to day and some tasks may result in greater behavior variation than others. If certain task combinations arise with reasonable regularly, a stable set of clusters will emerge. During a user's first few days in the system, the data may be too sparse and widely dispersed for the data mining component to draw meaningful conclusions. Thus, management must closely monitor users' initial activity until simple patterns begin to emerge. As patterns emerge, automated analysis increases in effectiveness.

In addition to the local outlier factor metric, the data mining component determines each characteristic's relative contribution to an item's LOF and offers the results to administrators. This additional impact information not only makes analysis far quicker and easier for administrators but also raises the possibility that supervisors may be able to assist in threat detection. In general, supervisors are most familiar with a user's day-to-day tasks, and therefore, most qualified to assess the actions necessary to complete those tasks. Unfortunately, not all supervisors are equally technically literate. Additional information beyond a cryptic local outlier score, if presented properly, is a critical first step towards making results accessible to non-technical users.

Preliminary results indicate that the data mining component effectively detects anomalous behavior in synthetic data sets. Also, the additional characteristic impact information is meaningful and helpful in those tests. System administrators have the ability to use LOF scores in whatever manner they choose: an administrator may look at all users with LOF scores above a given threshold or may consider only users with the top k anomaly scores. Should an event occur that an administrator knows will influence the anomaly scores of certain users in a given manner, the administrator can filter those scores to remove false positives. Administrators may even create more complex rules, such as ignoring certain LOF scores heavily influenced by attributes that the administrator considers insignificant. The strength and adaptability of this data mining approach indicate that it may be a useful tool for uncovering novel insider threats.

The prototype implementation of the system is presently only partially complete. Extensions are necessary for the monitoring tools, and the analysis components are not optimized for wide-scale deployment. Continued testing of the analysis components of the system as new insider activity data becomes available would be useful. Comprehensive analysis of system- and network-level charac-

teristics of user behavior also may yield insight regarding the utility of these characteristics in assessing insider threat. For example, certain aspects of user behavior be too noisy to serve as reliable indicators while others may be extremely consistent over time. Finally, the system would benefit from the addition of components to mine continuously for heuristic rules to supplement the system's real-time rule-based analysis component.

Acknowledgments. Calandrino and McKinney performed this research while under appointment to the Department of Homeland Security (DHS) Scholarship and Fellowship Program, administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and DHS. ORISE is managed by Oak Ridge Associated Universities (ORAU) under DOE contract number DE-AC05-06OR23100. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of DHS, DOE, or ORAU/ORISE.

References

1. D. Anderson, D. M. Cappelli, J. J. Gonzalez, M. Mojtahedzadeh, A. P. Moore, E. Rich, J. M. Sarriegui, T. J. Shimeall, J. M. Stanton, E. A. Weaver, A. Zagonel. Preliminary system dynamics map of the insider cyber-threat problem. Proceedings of the 22nd International Conference of the System Dynamics Society. 2004.
2. D. Anderson, T. F. Lunt, H. Javitz, A. Tamaru, A. Valdes. Detecting unusual program behavior using the statistical component of the next-generation intrusion detection expert system (NIDES). Technical Report SRI-CSL-95-06. SRI International, May 1995.
3. M. M. Breunig, H.-P. Kriegel, R. T. Ng, J. Sander. LOF: Identifying density-based local outliers. Proceedings of the ACM SIGMOD 2000 International Conference on Management of Data. ACM, 2000.
4. P. K. Chan, W. Fan, A. L. Prodromidis, S. J. Stolfo. Distributed data mining in credit card fraud detection. IEEE Intelligent Systems. IEEE, 1999.
5. V. Chandola, E. Eilertson, L. Ertöz, G. Simon, V. Kumar. Data mining for cyber security. Data Warehousing and Data Mining Techniques for Computer Security. Springer, 2006.
6. L. Ertöz, E. Eilertson, A. Lazarevic, P.-N. Tan, V. Kumar, J. Srivastava, P. Dokas. MINDS - Minnesota intrusion detection system. Next Generation Data Mining. MIT Press, 2004.
7. INFOSEC Research Council. Hard problem list. November 2005.
8. W. Lee, S. J. Stolfo. Data mining approaches for intrusion detection. Proceedings of the Seventh USENIX Security Symposium. January 1998.
9. W. Lee, S. J. Stolfo, P. K. Chan, E. Eskin, W. Fan, M. Miller, S. Hershkop, J. Zhang. Real time data mining-based intrusion detection. Proceedings of the 2001 DARPA Information Survivability Conference and Exposition II. June 2001.
10. M. R. Randazzo, M. Keeney, E. Kowalski, D. Cappelli, A. Moore. Insider threat study: Illicit cyber activity in the banking and finance sector. Technical Report CMU/SEI-2004-TR-021. Carnegie Mellon Software Engineering Institute, June 2005.
11. United States Secret Service, CSO, CERT. 2004 e-crime watch survey. 2004.