A Practical Application of Machine Learning-Based Classification Techniques to Proactively Identify Insider Threats

Joshua D. Bowen

Northern Arizona University

Prescott Valley Campus

Proposal for the Arizona / Nevada Academy of Sciences

April 1, 2017

Insider incidents are on the rise, just like the high-profile security breaches such as Snowden, thousands of insider-perpetrated security breaches occur in United States businesses every day. While current commercial software can monitor, log, and prevent access to designated files and directories, it remains difficult to predict and prevent unauthorized insider usage. Due to the gaps in research in this area, the focus of this study is to more accurately predict insider threats within a terminal environment.

Linux was chosen specifically because of its ubiquity on commercial servers around the globe. Amazon’s Machine Learning (AML) service has been selected to analyze the data, reduce the necessary computing power, and to minimize human factors considerations in the design of the machine learning architecture. AML uses multinomial logistic regression for multi-class classification and uses the stochastic gradient descent optimization technique.

The method adapted is to utilize the AML software, it will be trained on a dataset comprised of normal user situations, crafted mistakes, and malicious activity. After providing the training dataset, the software will be instructed to make predictions against similar datasets to verify accuracy. It will then be tested against a human actor that will simulate multiple different roles, and test predictions in a high-fidelity simulation.

In result, should it be demonstrated that ML software can accurately identify and predict insider threats, this research could be a foundation for future cyber security software architectures. Other opportunities for research in this area would include ML applications in intruder and malware detection.

Preliminary References

“Grand Theft Data.” *Encyclopedia of Media Violence*, 2015, doi:10.4135/9781452299655.n76.

“Learning Algorithm.” *Learning Algorithm - Amazon Machine Learning*, Amazon Web Services, 2017, docs.aws.amazon.com/machine-learning/latest/dg/learning-algorithm.html

Acknowledgements

Nancy L. Jensen – Academic Credit Advisor

Dr. James A. Subach – Faculty Advisor