

# Intrusion Detection using Provenance Graph Centrality Parzen-Windows

Kenny Yu  
Harvard University  
kennyyu@college.harvard.edu

R. J. Aquino  
Harvard University  
rjaquino@college.harvard.edu

CS261, Fall 2013

## Abstract

Provenance is data that describes how a digital artifact came to be in its current state. We hypothesize that intrusions on a system leave behind anomalies in the lineage of digital artifacts. We present an intrusion detection approach to find these anomalies by analyzing centrality metrics on provenance graphs. We use a Parzen-Window approach (TODO CITE) on various provenance graph centrality metrics (TODO CITE) to determine probability density estimates of normal behavior, and we use these density estimates to determine if an intrusion occurred. We used this approach to analyze *user-to-remote* (u2r) intrusions and *remote-to-local* (r2l) intrusions (TODO: include r2l?) from the 1998 DARPA Intrusion Detection data sets (TODO CITE) and achieved up to \*TODO true positive rate for intrusions\* accuracy in detecting intrusions with only \*TODO false positive rate for intrusions\* accuracy. We also present future work to extend our intrusion model to an online intrusion detection system.

## 1 Introduction

## 2 Related Work

## 3 Design and Implementation

### 3.1 Selecting Metrics

## 4 Evaluation

### 4.1 Experimental Setup

### 4.2 Results

### 4.3 Discussion

## 5 Future Work

## References

- [1] CAO, D., QIU, M., CHEN, Z., HU, F., ZHU, Y., AND WANG, B. Intelligent Fuzzy Anomaly Detection of Malicious Software. In *Internal Journal of Advanced Intelligence*, vol. 4, no. 1, pp 69-86 (December 2012).
- [2] INOUE, H. AND SOMAYAJI, A. Lookahead Pairs and Full Sequences: A Tale of Two Anomaly Detection Methods. In *2nd Annual Symposium on Information Assurance* (June 2007).
- [3] KING, S. T. AND CHEN, P. M. Backtracking Intrusions. In *SOSP'03 Proceedings of the nineteenth ACM symposium on Operating systems principles* (December 2003).
- [4] KING, S. T., MAO Z. M., LUCCHETTI, D. G., AND CHEN, P. M. Enriching intrusion alerts through multi-host causality. In *Proceedings of the 2005 Network and Distributed System Security Symposium* (February 2005).

- [5] LEI, H. AND DUCHAMP, D. An Analytical Approach to File Prefetching. In *Proceedings of the USENIX 1997 Annual Technical Conference* (January 1997).
- [6] MACKO, P., MARGO, D., SELTZER, M. Local Clustering in Provenance Graphs (Extended Version). In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management* (August 2013).
- [7] MACKO, P. AND SELTZER, M. Provenance Map Orbiter: Interactive Exploration of Large Provenance Graphs. In *TaPP'11 Proceedings of the 2nd conference on Theory and practice of provenance* (June 2011).
- [8] MARGO, D., AND SMOGOR, R. Using Provenance to Extract Semantic File Attributes. In *TaPP'10 Proceedings of the 2nd conference on Theory and practice of provenance* (February 2010).
- [9] MUNISWAMY-REDDY, K., BRAUN, U., HOLLAND, D. A., MACKO, P., MACLEAN, D., MARGO, D., SELTZER, M., AND SMOGOR, R. Layering in Provenance Systems. In *Proceedings of the 2009 USENIX Annual Technical Conference* (June 2009).
- [10] MUNISWAMY-REDDY, K., HOLLAND, D. A., BRAUN, U., AND SELTZER, M. Provenance-Aware Storage Systems. In *Proceedings of the 2006 USENIX Annual Technical Conference* (June 2006).
- [11] OFFENSIVE SECURITY, INC. The Exploit Database. <http://www.exploit-db.com>.
- [12] RAPID 7 INC. Metasploit Framework. <http://www.metasploit.com>.
- [13] SOMAYAJI, A. AND FORREST, S. Automated Response Using System-Call Delays. In *Proceedings of the 2000 USENIX Annual Technical Conference* (August 2000).
- [14] TARIQ, D., BAIG, B., GEHANI, A., MAHMOOD, S., TAHIR, R., AQIL, A., AND ZAFAR, F. Identifying the provenance of correlated anomalies. In *SAC'11 Proceedings of the 2011 ACM Symposium on Applied Computing* (March 2011).