TODO: Intrusion Detection through Provenance

Kenny Yu
Harvard University
kennyyu@college.harvard.edu

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

CS261, Fall 2013

Abstract

TODO

1 Introduction

Provenance is metadata that tracks the history of all changes to files. In provenance-aware storage systems like PASS [6] and PASSv2 [5], the file system automatically tracks all dependencies whenever a file is created, modified, or deleted. These dependencies include the command that was executed to modify the file, the environment of the command, and all input files. The generated provenance forms a directed acyclic graph of typed nodes (e.g. files, processes, pipes) with properties (e.g. name, execution time, pid) and typed edges (e.g. forked, input, versioning).

Applications that require provenance collection typically place great emphasis on data integrity. Given this priority, a natural goal of provenance systems is to detect intrusions on the system. Somayaji et. al. have developed techniques for detecting intrusions based on system call sequences, and counteract these by exponentially delaying or aborting system calls, rendering the system useless for a malicious attacker [7][1]. These intrusions include exploiting vulnerabilities in the SSH daemon and sendmail to obtain a shell with root privileges.

Given the extensive amount of data PASS collects on file-file, file-process, and process-process dependencies, it seems plausible to detect intrusions based on provenance data collected by PASS.

====

Several existing systems intrusion detection systems using system calls

other systems to collect provenance to detect intrusions

our approach: analyze properties of dags, histograms?

2 Design

TODO

3 Evaluation

TODO

4 Conclusion

TODO

References

[1] 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).

[2] 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).

[3] 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).

[4] 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).

[5] 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).

[6] 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).

[7] Somayaji, A. and Forrest, S. Automated Response Using System-Call Delays. In *Proceedings of the 2000 USENIX Annual Technical Conference* (August 2000).