Anomaly detection in Windows Registry hives using machine learning

Krish Patel
Science & Engineering
Manalapan High School
Englishtown, NJ
425kpatel@frhsd.com

Jitin Jindal
Commvault
Tinton Falls, NJ
jjindal@commvault.com

Abstract

The importance of malware detection software has increased dramatically with the rise of electronic data sharing and the spread of data theft. During my internship at Commvault, an industry leader in data security and management, I created an algorithm for monitoring Windows Registry hives. The registry serves as a critical database for settings, application preferences, and system parameters, making it a frequent target of malware and malicious activities. By analyzing key-value pairs, structural relationships, and timed patterns within real-world activity on Commvault servers, I established a baseline of normal activity to detect deviations. The system integrates supervised learning models trained on labeled datasets to identify specific types of anomalies, such as trojan horses and worms. Furthermore, the model incorporates real-time logging and notifications to alert the user of potential threats. When run on a simulation of registry tampering and unauthorized key deletions, it achieves an F_1 score of 0.91, which highlights the capability of the system to detect subtle anomalies while minimizing false positives, equipping systems to address emerging threats preemptively.

Index Terms

malware detection, Windows Registry hives, anomaly detection, metadata, Commvault, internship