Improving the Ransomware Detection/Mitigation Proof-of-Concept With Machine Learning

Song, Eric (School: Yorktown High School)

Ransomware attacks render data unusable and demand a ransom in order to restore data integrity. These attacks cost billions of dollars every year at the individual, industrial, and national security levels. As ransomware poses a big threat to computer systems today, research in the field is active. However, there is greater emphasis with regard to prevention, rather than incident response. Preventative systems, while effective, can be bypassed by attackers, given enough time. Incident response aims to address this issue by mitigating attacks as they happen. Key extraction can be used for incident response when dealing with ransomware. By intercepting the ransomware's encryption key, restoring data can be done without payment of ransom. Random forest models were constructed to test the feasibility of using machine learning in combination with this technology, due to the time-sensitive nature of key extraction. To do so, a large dataset comprised of 30,000 samples was developed by testing binaries and outputting behavioral data. Afterwards, 100 random forest models were built with different random states to test the efficacy of random forests. The models demonstrate that machine learning can be incorporated into the system with a better detection rate than the previous proof-of-concept (99.3% versus 66.7%). Additionally, the efficiency of the random forest is viable for use with key extraction. In the future, this integration can be further developed to become more optimized and have better capability in defeating advanced ransomware. Additionally, different models can be used to analyze their advantages over the random forest in ransomware detection.