A Similarity-Based Machine Learning Approach for Detecting Adversarial Android Malware

Doaa Hassan[a], Matthew Might, and Vivek Srikumar
University of Utah

UUCS-14-002

[a]Computers and Systems Department, National Telecommunication Institute, Cairo, Egypt.

School of Computing
University of Utah
Salt Lake City, UT 84112 USA

October 20, 2014

Abstract

We introduce a similarity-based machine learning approach for detecting non-market, adversarial, malicious Android apps. By adversarial, we mean those apps designed to avoid detection. Our approach relies on identifying the Android applications that are similar to an adversarial known Android malware. In our approach, similarity is detected statically by computing the similarity score between two apps based on their methods similarity. The similarity between methods is computed using the normalized compression distance (NCD) in dependence of either **zlib** or **bz2** compressors. The NCD calculates the semantic similarity between pair of methods in two compared apps. The first app is one of the sample apps in the input dataset, while the second app is one of malicious apps stored in a malware database. Later all the computed similarity scores are used as features for training a supervised learning classifier to detect suspicious apps with high similarity score to the malicious ones in the database.