Assignment 2

Eynav Ben Shlomo 209328970, Elon Ezra 313534133

December 15, 2022

Abstract

As part of the course "Attacks on the Android operating system" , We chose to carry out the project on the classifier found in the article "Yes, Machine Learning Can Be More Secure! A Case Study on Android Malware Detection"

1 Introduction

We tested the classifier and listed the weak points and how it can be exploited for the planned attack on the classifier. Carrying out this research is so that when we carry out the attack on the classifier later on, we can inject a malicious application into the classifier, but the classifier will classify it as innocent with the help of the types of attacks detailed in this article.

2 Vulnerability Attacks

2.1 Obfuscate Or Encrypt Attacks

According to the article, Drebin and sec-SVM use static code analysis to identify malicious test code before it is executed. However, there are situations in which malicious code is only executed when the app is running, such as when the app injects code after launching or when it connects to a network. a way to exploit this vulnerability is to obfuscate or encrypt the malicious code so that the feature extraction process does not detect it.

2.2 Mimicry Attacks

It is possible to perform a mimicry attack on apps, as the secSVM analysis method is static and features that are truly innocent can be used by attackers. Malicious malware samples can almost exactly replicate benign data, and this is possible due to an intrinsic vulnerability of the feature representation. No learning algorithm can accurately separate such data with satisfactory accuracy. For example, if an app sends requests to a specific domain, an attacker could set up their own server to replace the real domain with their own address and mediate between the app and the real server. This could be demonstrated with an app that logs in with a password and uses a server. The password and username would be sent to the attacker's server, where they could be saved and passed on to the original network.

2.3 Limited Knowledge Attacks

Limited knowledge attacks are possible, as the machine learning models used in the classification process are given a vector of features and are trained to classify them based on their maliciousness. By manipulating the features in such a way that they match the feature vectors that have passed the classifier test, an attacker could successfully pass the test. This could be done by adding or subtracting features to the malicious app that make the feature vector appear benign, although this could potentially affect the app's functionality.