# Yes, Machine Learning Can Be More Secure! A Case Study on Android Malware Detection

## ABSTRACT

To cope with the increasing variability and sophistication of modern attacks, machine learning has been widely adopted as a statistically-sound tool for malware detection. However, its security against well-crafted attacks has not only been recently questioned, but it has been shown that machine learning exhibits inherent vulnerabilities that can be exploited to evade detection at test time. In other words, machine learning itself can be the weakest link in a security system.

## EXISTING SYSTEM

During the last decade, machine learning has been increasingly applied in security-related tasks, in response to the increasing variability and sophistication of modern attacks. One relevant feature of machine-learning approaches is their ability to generalize, i.e., to potentially detect never-before-seen attacks, or variants of known ones. However, as first pointed out by Barrenoet, machine-learning algorithms have been designed under the assumption that training and test data follow the same underlying probability distribution, which makes them vulnerable to well-crafted attacks violating this assumption. This means that machine learning itself can be the weakest link in the security chain. Subsequent work has confirmed this intuition, showing that machine-learning techniques can be significantly affected by carefully-crafted attacks exploiting knowledge of the learning algorithm.

## Disadvantages

- It can be the weakest link in the security chain.
- Skilled attackers can manipulate data at test time to evade detection, or inject poisoning samples into the training data to mislead the learning algorithm and subsequently cause misclassification errors.

## PROPOSED SYSTEM

In this paper, we rely upon a previously-proposed attack framework to categorize potential attack scenarios against learning-based malware detection tools, by modeling attackers with different skills and capabilities. We then define and implement a set of corresponding evasion attacks to thoroughly assess the security of Drebin, an Android malware detector. The main contribution of this work is the proposal of a simple and scalable secure-learning paradigm that mitigates the impact of evasion attacks, while only slightly worsening the detection rate in the absence of attack. We finally argue that our secure-learning approach can also be readily applied to other malware detection tasks.

## Advantages

- It improves the security of Drebin against stealthier attacks.
- It able to retain computational efficiency and scalability on large datasets.

## SOFTWARE REQUIREMENTS

| | | |
|---|---|---|
| Front-end | : | JSP |
| Back-End | : | MySQL |
| Server | : | Tomcat Server |
| OS | : | WINDOWS 7/above |

## HARDWARE REQUIREMENTS

| | | |
|---|---|---|
| PROCESSOR | : | CORE i3 |
| RAM | : | 512MB-2GB |
| HARD DISK | : | 40GB |