# Learn2Evade: Learning-based Generative Model for Evading PDF Malware Classifiers

Recent research has shown that a small disturbance to an input may forcibly change the prediction of a machine learning (ML) model. Such variants are commonly referred to as adversarial examples. Early studies have focused mostly on ML models for image processing and expanded to other applications, including those for malware classification. This paper focus on the problem of finding adversarial examples against ML-based PDF malware classifiers. This problem is more challenging than those against ML models for image processing because of the highly complex data structure of PDF and of an additional constraint that the generated PDF should exhibit malicious behaviour. To resolve this problem, here proposes a variant of generative adversarial networks (GANs) that generate evasive variant PDF malware (without any crash), which can be classified by various existing classifiers yet maintaining the original malicious behaviour.