**Second Assignment- Weak spots identification**

Written by:  
Or Cohen : 206585515   
Shlomi Lantser : 209322379

**About SECSVM Classifier:**The SECSVM classifier is a machine learning algorithm designed to detect malware on mobile devices. It is designed to address the limitations of the Drebin classifier, which relies on static analysis and lacks dynamic analysis, making it vulnerable to targeted attacks. The authors of the article propose an attack model and framework to systematically evaluate and improve the security properties of the SECSVM classifier against targeted attacks. The attack model considers three main dimensions: the attacker's goal (security violation and attack specificity), the attacker's knowledge (of the training data, feature extraction/selection algorithm, and learning algorithm), and the attacker's capability (attack influence and data manipulation). The authors then evaluate a number of potential attack scenarios based on this model and demonstrate the effectiveness of the SECSVM classifier in defending against them.

**Weak spots of SECSVM Classifier: (**[**Classifier article**](https://arxiv.org/pdf/1704.08996.pdf)**)**

* One weak spot of the SECSVM classifier is its vulnerability to encryption and obfuscation attacks. As the classifier relies on static code analysis, it is not able to effectively detect malware that is encrypted or heavily obfuscated. This weakness is due to the limitations of the feature representation used by the classifier, rather than the learning algorithm itself.  
  According to the article, the vulnerability of the SECSVM classifier to encryption and obfuscation attacks is a result of the limitations of the feature representation used by the classifier.  
  The article states that "it is clear that also Sec-SVM can be defeated by more sophisticated encryption and obfuscation attacks. However, it is also worth remarking that this is not a vulnerability of the learning algorithm itself, but rather of the chosen feature representation."
* Another weak spot of the SECSVM classifier is its potential poor performance under certain types of attacks, such as mimicry attacks in which malware samples are significantly modified to closely replicate benign data, or knowledge-based attacks in which the attacker has a high level of knowledge about the system.

**How do I manage to attack him :**

* To attack the SECSVM classifier using encryption and obfuscation, I would attempt to create malware samples that are heavily encrypted or obfuscated to evade detection by the classifier. This could potentially be achieved using specialized software or techniques designed to obscure the code of the malware.
* To attack the SECSVM classifier using mimicry or knowledge-based attacks, I would attempt to modify malware samples in such a way that they closely replicate benign data or gather a high level of knowledge about the system to better understand its weaknesses. However, as mentioned in the article, these types of attacks may be difficult to execute successfully due to the inherent limitations of the feature representation and the difficulty of significantly modifying malware samples without compromising their functionality or leaving traces of manipulation.

Both points are mentioned in the article. Static code analysis relies on analyzing the code of a program or application to detect malicious intent. However, if the code has been encrypted or obfuscated, it can be difficult for a classifier to accurately detect malicious intent.  
Mimicry attacks involve modifying malware samples to closely replicate benign data, which can also lead to a decrease in detection accuracy by a classifier. Both attacks can be considered evasion attacks, as they involve manipulating data at test time to evade detection by the classifier.