A Machine Learning Approach for Classifying and Categorizing Andrid Sources and Sinks

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Field | Description  (Mayank Chaudhari) | Comment 1  (peer 1 name) | Comment 2  (peer 2 name) | Comment 3  (peer 3 name) |
| Malware Analysis Technique | Static analysis |  |  |  |
| Dataset description |  |  |  |  |
| Feature selection Technique | mannual |  |  |  |
| Feature Extraction Technique |  |  |  |  |
| Classification Technique | SVM |  |  |  |
| Accuracy or confusion matrix scores |  |  |  |  |
| Scalability and Efficiency | The method is scalable to large amount of apps |  |  |  |
| Overview | SUSI is an automated machine learning guides approach for identifying sources and sinks directly from the code of an Android API. While SISI is not able to identify each and every source or sink it resembles a practical best-effort solution and achieves a precision and recall of more than 92%. In this work researchers at first provide the definition for a source and sink. Then based on this definition the methods are labeled as source and sink. This information is then used to detect more source and sinks by using SVM. Later the identified sources are then categorized into 12 source and 15 sink categories.  DUring classification task they have more than two classes so they have used one-against-classification, There are some methods that are both source and sinks, These can be treated either source or sink. At first the methods in test set are classified as source or sink or neither. The methods labeled as neither are left form further analysis. After this step the methods source or sink methods are categorized into different categories. The feature database consist of 144 features.it was observed that considering the method signature and the syntax of its method body alone is insufficient to reliably detect source and sinks. With such features alone they were unable to obtain a precision or recall higher than 60%. The inclusion of data-flow features greatly helped for better classification. SUSi can not identify callback methods and performs well with functions with parameters and return values. |  |  |  |
|  |  |  |  |  |