4 November 2018 Study · Mobile Security

Similarity of Source Code in Applications

The full text is currently only available in the Deutsch version. A translation will follow.

Archived. This content is no longer updated. It remains available for reference.

Version	2018-10-23
Source	https://technology2.a-sit.at/en/publications/similarity-of-source-code-in-applications/

In order to gain a more abstract understanding of the behavior of program code, this project uses current scientific approaches to identify and classify patterns in code.

For the security-oriented analysis of applications, it has always been essential to identify all components occurring in the programs and their designated tasks. With the understanding of the functionality gained through this, it can finally be determined whether applications violate established security principles or, apart from that, whether they implement the functionality that would be expected at all. However, the complexity and scope of individual applications have increased dramatically in recent years, making it difficult to understand the processes and interrelationships involved in code reviews.

The goal of the project was to identify patterns in code by applying current scientific approaches in order to derive a more abstract understanding of the behavior of source code. For this it is essential to find suitable methods to process even very large amounts of code in such a way that a semantic understanding and classification of patterns is possible. The knowledge gained from this project should provide a solution for identifying semantically similar code fragments.