Comparison of Binary Procedures: A Set of Techniques for Evading Compiler Transformations

Z. Radivojević, M. Cvetanović, S. Stojanović

References:

Abstract:
License violation analysis may require digital forensics in the performance of a time-consuming search in order to find out whether a binary code of a product contains a procedure that originates from a source code for which a license is required. The conducted experiment shows that the production of a binary code using an arbitrary compiler decreases results of the evaluated solutions up to 10 times. The best performing solution, among those evaluated, uses software metrics for assessing similarities between procedures and ranks procedures from the binary code according to their similarities with the target forensics procedure. This paper tries to improve the ranking by proposing five techniques for making similarities assessment more robust against compiler transformations. The proposed techniques filter stack instructions and transfer instructions, retain partial information about the instruction order, simulate inlining, and eliminate procedures that significantly differ from the searched procedure. The techniques are evaluated using a dataset based on the STAMP benchmark and re-evaluated using a dataset based on the BusyBox toolset. The evaluation shows that the use of the proposed techniques increases recall by 47 and 42% for the first and second datasets, respectively.

Keywords:
binary code analysis; license violation; dual-license; software metric; software library copyright infringement