ABSTRACT

CORPORA DISTILLATION IN FUZZING PROCESS USING GENETIC ALGORITHM

The dissertation aims to develop the original method of corpora distillation in fuzzing process. The multi-criteria genetic algorithm – epiVEGA was used for this purpose. It is the enriched VEGA algorithm with the original epigenetic operator and the additional convergence control module.

The history of fuzzing and its classification were presented in the thesis. Theoretical part of the dissertation describes how the nowadays fuzzers worklub Moreover, the role of corpora in fuzzing process, the modern methods and the criteria of distillation were discussed. The next part of the doctoral thesis focuses on genetics algorithms. The classic versions of genetic operators, a mutation and a crossover, were brought closer. The chapter describes/presents the role of selection in simulated evolution process. The review of the state of knowledge was concluded in the chapter about epigenetic phenomena as the methods of non-genetic inheritance. Previous attempts of reconstruction epigenetic operators in the process of simulated evolution were summarized additionally.

In the dissertation the reduction of corpora was interpreted as multi-criteria set cover problem (MCSCP). Currently a distillation is interpreted as minimum-weight set cover problem. The reduction was proposed with four criteria of distillation, including three classical: size of files, time of execution and number of activated edges in control flow graph of tested program. The fourth, defined by the author, was simplified entropy – the metric of files randomness. The solution of MCSCP was found with multi-criteria genetic algorithm enriched with a proprietary epigenetic operator. It was inspired by the biological phenomenon of acetylation and deacetylation of histones. The minimum probability p_e of the operator activation and the best of its variation were found experimentally. Proposed algorithm was expanded with the convergence control module based on the second-order quartile and the range. It modified dynamically portability of the mutation p_m and the crossover p_c. It was necessary for avoiding the domination of the population by empty sets.

The proposed method could be an alternative to distillators based on dynamic programming algorithms or artificial intelligence, simultaneously allowed to conduct reduction with bigger numbers of criteria.

Based on conducted researche the thesis of the dissertation was confirmed. The use of multi-criteria genetic algorithm enriched with the epigenetic operator and convergence control module allows to effectively reduce the test data for carrying out the fuzzing process.