

## STRESZCZENIE ROZPRAWY DOKTORSKIEJ

### DESTYLACJA KORPUSU DANYCH TESTOWYCH W PROCESIE FUZZINGU Z WYKORZYSTANIEM ALGORYTMU GENETYCZNEGO

Celem niniejszej dysertacji było opracowanie oryginalnej metody destylacji korpusów danych testowych wykorzystywanych w procesie fuzzingu. Zastosowano w tym celu wielokryterialny algorytm genetyczny epiVEGA, wzbogacający algorytm VEGA o autorski operator epigenetyczny oraz mechanizm sterowania zbieżnością.

W pracy przedstawiono historię i klasyfikację fuzz testów. W części teoretycznej przybliżono zasady działania współczesnych fuzzerów. Omówiono rolę korpusów w procesie fuzzingu, a także współczesne metody i kryteria ich destylacji. Dalsza część pracy została poświęcona algorytmom ewolucyjnym. Przybliżono klasyczne odmiany operatorów genetycznych – krzyżowania i mutacji. Opisana została rola selekcji w procesie symulowanej ewolucji. Przegląd stanu wiedzy zakończony został rozdziałem przybliżającym tematykę zjawisk epigenetycznych jako sposobu dziedziczenia pozagenetycznego. Dokonano podsumowania dotychczasowych prób odwzorowywania operatorów epigenetycznych w procesie symulowanej ewolucji.

W przedmiotowej rozprawie redukcja korpusu danych testowych została utożsamiona z problemem wielokryterialnego pokrycia zbioru (ang. *multi-criteria set cover problem MCSCP*) w przeciwieństwie do obecnie stosowanego podejścia, w którym destylacja rozumiana jest jako problem minimalnego ważonego pokrycia zbioru (ang. *minimum-weight set cover problem, MWSCP*). Zaproponowano redukcje zbioru danych testowych uwzględniającą cztery kryteria destylacji, w tym trzy klasyczne: rozmiar plików, czas ich obsługi przez badany program oraz pokrycie kodu wyrażone w liczbie aktywowanych krawędzi w grafie przepływu sterowania. Czwartym, nowo zdefiniowanym kryterium redukcji była entropia uproszczona – miara losowości plików. Rozwiązanie *MCSCP* zostało wyznaczone z wykorzystaniem wielokryterialnego algorytmu genetycznego, w którym zaimplementowano autorski operator epigenetyczny, wzorowany na biologicznym procesie acetylacji i deacetylacji histonów. Minimalne prawdopodobieństwo wystąpienia operatora epigenetycznego  $p_e$  oraz najlepszy jego wariant wyznaczono eksperymentalnie. Zaproponowany algorytm został także rozbudowany o sterowanie zbieżnością w oparciu o wartość rozstępu i kwartyłu drugiego rzędu. Moduł ten dynamicznie modyfikował prawdopodobieństwo wystąpienia mutacji  $p_m$  i krzyżowania  $p_c$  w celu uniknięcia zdominowania populacji potencjalnych rozwiązań przez zbiór pusty.

Sformułowana metoda może stanowić alternatywę dla destylatorów bazujących na algorytmach programowania dynamicznego czy sztucznej inteligencji, pozwalając jednocześnie na efektywne rozpatrywanie większej liczby kryteriów redukcji zbioru danych testowych.

Na podstawie przeprowadzonych badań potwierdzono postawioną w pracy tezę, iż wykorzystanie w procesie destylacji wielokryterialnego algorytmu genetycznego wzbogaconego o operator epigenetyczny oraz mechanizm sterowania zbieżnością pozwala na efektywną redukcję korpusu danych testowych w procesie fuzzingu.