*dr hab. inż.* ***Maciej Walkowiak***
*prof. uczelni*
*Politechnika Bydgoska im. Jana i Jędrzeja Śniadeckich w Bydgoszczy*
*Wydział Telekomunikacji, Informatyki i Elektrotechniki*

Bydgoszcz, 4 kwietnia 2023 r.

**RECENZJA ROZPRAWY DOKTORSKIEJ**

**kpt. mgr. inż. Marcina Pachnika**

***Destylacja korpusu danych testowych w procesie fuzzingu z wykorzystaniem algorytmu genetycznego***

Podstawą do przygotowania recenzji rozprawy Pana mgr. inż. Marcina Pachnika jest pismo Pana dr. hab. inż. Jana Kelnera, profesora WAT, zastępcy przewodniczącego Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Wojskowej Akademii Technicznej im. Jarosława Dąbrowskiego w Warszawie z 7 lutego 2023 roku (sygn.: WYCH\N\00062\22023), informujące o uchwałach Rady Dyscypliny Naukowej

Recenzja została przygotowana na podstawie rozprawy doktorskiej.

Opiniowana praca doktorska Pana Marcina Pachnika powstała pod naukowym kierunkiem Pana dr. hab. inż. Kazimierza Worwy, profesora WAT, przy wsparciu promotora pomocniczego Pana płk. dr. inż. Rafała Kasprzyka.

Praca została przedstawiona Radzie Dyscypliny Naukowej *Informatyka Techniczna i Telekomunikacja* Wojskowej Akademii Technicznej im. Jarosława Dąbrowskiego w trybie przewidzianym w Ustawie z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce* (Dz. U. 2018 poz. 1668 z późn. zm.). Rada Dyscypliny Naukowej podjęła uchwałę nr 5/RDN ITiT/2023 powierzając mi wykonanie recenzji pracy doktorskiej.

## 1 *Jaki jest problem naukowy (teza) rozprawy i czy został on trafnie i jasno sformułowany?*

Fuzzing jest popularny w dziedzinie bezpieczeństwa komputerowego, szczególnie w testowaniu bezpieczeństwa oprogramowania. Idea polega na dostarczaniu programowi różnych nietypowych, przypadkowych lub losowych danych wejściowych w celu identyfikacji ewentualnych błędów, awarii czy podatności na ataki. W zasadzie chodzi o przygotowanie takich nieprzewidywalnych danych wejściowych, które mogą spowodować niepożądane lub nieprzewidywane zachowanie programu.

Fuzzing jest stosowany przez testerów bezpieczeństwa do weryfikacji oprogramowania pod kątem błędów i podatności na ataki. Jest to skuteczna technika, która może ujawnić wiele problemów, szczególnie w przypadku aplikacji krytycznych.

Na rynku istnieje wiele darmowych i komercyjnych narzędzi pracujących w oparciu o omawianą ideę. W każdym przypadku jednym z niezbędnych kroków postępowania jest generowanie danych wykorzystywanych do testowania. Większość z tych narzędzi do generowania danych testowych korzysta z wcześniej przygotowanego zbioru tzw. korpusu. Im mniejszy jest korpus testowy, tym krócej trwa proces testowania. Zmniejszanie objętości korpusu nazywa się destylowaniem.

Celem pracy autorki były badania dotyczące skutecznego przygotowania korpusu danych. Cel rozprawy jest sformułowany w miarę precyzyjnie, dobrze określa obszar i zakres przeprowadzonych badań.

Autor sformułował tezę naukową rozprawy w taki sposób, iż *wykorzystanie w procesie destylacji wielokryterialnego algorytmu genetycznego wzbogaconego o operator epigenetyczny oraz mechanizm sterowania zbieżnością pozwoli na efektywną redukcję korpusu danych testowych w procesie fuzzingu.*

Teza naukowa jest sformułowana ostrożnie. Z treści pracy wynika bowiem, że efektywność redukcji ocenić można liczbowo, dlatego wydaje się, że zawarcie w tezie informacji o ilościowej mierze redukcji byłoby odważniejsze.

## 2 *Jaka jest konstrukcja rozprawy?*

Rozprawa została przedstawiona w siedmiu numerowanych rozdziałach z dodanymi zwyczajowymi spisami oraz literaturą.

We wstępie autor wprowadza w tematykę pracy, definiuje cel i zakres rozważań oraz prezentuje tezę rozprawy,

Opisanie fuzzingu znajduje się w rozdziale drugim, w którym obok prezentacji stanu wiedzy, autor wprowadza podstawowe terminy i pojęcia związane z tą metodą. Autor porusza tutaj też ciekawy – i właściwie nadający się na odrębną rozprawę – temat przeciwdziałania fuzzingowi, stosowanymi jako jeden ze sposobów na niepożądaną penetrację kodu.

Rozdziały trzeci i czwarty są jest poświęcony metodom poszukiwania rozwiązania skutecznego opartych na mechanizmach dziedziczenia: mamy więc opis algorytmów ewolucyjnych oraz zjawisk epigenetycznych.

Problemy badawcze pracy są omówione w rozdziale piątym. Wśród tych problemów główne miejsce zajmuje destylacja korpusu traktowana jako problem optymalizacji wielokryterialnej oraz uzupełnienie algorytmu VEGA o operator epigenetyczny w celu sterowania różnorodnością populacji.

Etap eksperymentalny szczegółowo opisano w rozdziale szóstym. Pracę kończy podsumowanie i wnioski.

## 3 *Czy tematyka rozprawy jest aktualna lub dostatecznie ważna?*

U podstaw sformułowania problemu badawczego leży, oczywiście, zadanie zbudowania poprawnie działającego programu bądź szerzej – poprawnie działającego oprogramowania.

Termin „poprawne działanie" nie jest – moim zdaniem – zdefiniowany w informatyce, chodź w pewnych innych dyscyplinach nauki i techniki termin taki funkcjonuje.

Najlepszym przykładem jest tutaj kompatybilność elektromagnetyczna, w której zasada poprawnego działania przyrządów, urządzeń i systemów działa od dawna. Trawersując po znaczeniu tego pojęcia, w interesującym nas przypadku oprogramowania możemy powiedzieć, że program działa poprawnie, jeśli akceptujemy działanie tego programu. Akceptacja zaś oznacza zachowanie podstawowych funkcjonalności przy braku wpływu jakichkolwiek działań niepożądanych.

Brak podatności oprogramowania jest w zasadzie cechą projektową co powinno być odzwierciedlone w założeniach projektowych. Jednocześnie, przy stopniu złożoności oprogramowania podatności są nieuniknione. Dlatego tak ważną rzeczą jest procedura sprawdzania szczelności gotowego oprogramowania.

Jeśli oprogramowanie jest tworzone zgodnie z kanonami sztuki, to nie powinniśmy się spodziewać istnienia oczywistych podatności, a raczej oczekujemy niepożądanych reakcji na szczególne i rzadko spotykane sytuacje. I stajemy przed problemem podobnym do tego, jakie mieliśmy z pomiarem przekłamań podczas transmisji. Póki elementowa stopa błędów jest rozsądnie duża, to w pewnym rozsądnym interwale czasu możemy wykryć błędy i dokonać pomiaru. Jednak w miarę wzrostu jakości transmisji liczba błędów jest tak mała, że w rozsądnym czasie pomiaru może się nie pojawić żaden błąd transmisji; nie oznacza to, oczywiście, że mamy do czynienia z transmisją bezbłędną. Należało więc stworzyć – i zrobiono to – metody pomiaru maleńkich elementowych stóp błędów w rozsądnym czasie.

Z podobną sytuacją mamy do czynienia podczas wykrywania podatności oprogramowania na szczególne działania lub szczególne sytuacje. Chcemy wykryć możliwie wiele błędów w rozsądnie długim czasie. Wszelkie metody wykrywające podatności oprogramowania w akceptowalnym czasie są niezbędnym elementem procesu produkcji oprogramowania.

Jak wynika z powyższego wywodu, tematyka rozprawy jest aktualna, niezwykle ważna i dotyczy niepomijalnego elementu powstawania oprogramowania.

## 4 *Czy rozprawa prezentuje ogólną wiedzę teoretyczną doktoranta?*

Oczywiście, od doktora nauk inżynieryjno-technicznych powinno się wymagać pewnej ogólnej wiedzy inżynierskiej i technicznej, jednakże wyłącznie na podstawie przedstawionej rozprawy nie można stanu takiej wiedzy potwierdzić.

Jeśli zaś chodzi o wiedzę właściwą dla rozpatrywanej dyscypliny naukowej, to w tym obrębie Doktorant porusza się swobodnie. Nie znalazłem też żadnej poważnej teoretycznej nieprawdy.

Zakres wiedzy autora rozprawy jest szeroki. A więc na początek mamy sposoby i techniki testowania oprogramowania na obecność podatności. Autor sprawnie porusza się także wśród algorytmów optymalizacyjnych i to nie tych najprostszych. Oczywiście, potrafi także projektować nowe i modyfikować istniejące algorytmy. Umie też przekładać algorytmy na kod. Porusza się dobrze w środowisku systemów linuksowych i oprogramowania dla takich systemów, co uważam za cenną, choć coraz rzadszą umiejętność młodych informatyków.

Autor ma szeroką wiedzę o algorytmach ewolucyjnych. Proces dziedziczenia genowego oraz zjawiska epigenetyczne, czyli dziedziczenie pozagenowe, zna doskonale i to z wszelkimi szczegółami i niuansami, a porusza się w tym materiale jak biolog-genetyk.

Tak więc treść rozprawy pokazuje, iż doktorant sprawnie posługuje się wiedzą teoretyczną w zakresie co najmniej reprezentowanej dyscypliny.

5 *Czy rozprawa wykazuje umiejętność samodzielnego prowadzenia pracy naukowej?*

To złożone pytanie, na które zbiorcza i jednoznaczna odpowiedź brzmi: tak. Cała praca, zarówno w swojej warstwie teoretycznej, koncepcji algorytmu, stosowanych metod matematycznych, programistycznych oraz umiejętności pomiarowych i poprawnej interpretacji wyników pomiarów świadczą o dużej i zaawansowanej wiedzy autora, i o jego dobrym przygotowaniu do dalszej pracy naukowej.

Opiniowana rozprawa jest napisana bardzo ładnym polskim językiem, poza kilkoma slangowymi określeniami. Sprawne posługiwanie się językiem jest umiejętnością ciężko i długotrwale nabywaną przez naukowców. Tutaj mamy do czynienia z dobrą jakością, co więcej łatwą do doskonalenia.

Również od strony edycyjnej trudno mi znaleźć większe uchybienia. Rozprawa jest dopracowana w wielu szczegółach edycyjnych, co oczywiście nie oznacza, że jest idealna, o czym piszę dalej.

Wbrew przekonaniom wielu młodych (i niestety - licznych starszych) naukowców, podstawą badań naukowych jest właściwa identyfikacja problemu i szczegółowe opisanie przeszkód do pokonania wraz z możliwie dokładną systematyką ilościową, a przynajmniej jakościową tychże przeszkód. Doktorant właściwie zidentyfikował problem.

Jest dla mnie jasne, że Doktorant wykazała dużą znajomość metod i zasad naukowego podejścia do badanych zagadnień. Wspomniane zagadnienia są przedstawione jasno i precyzyjnie, przy zastosowaniu właściwego formalizmu matematycznego, typowego dla rozpraw naukowych z dziedziny nauk technicznych. Autor rozprawy dobrze sobie radzi z większością opisywanych zagadnień matematycznych, które – swoją drogą – nie są tu bardzo skomplikowane. Autor wykazuje umiejętność przeprowadzania szerokiej i dogłębnej analizy rozpatrywanych zagadnień. Widać wyraźnie, że dobrze przyswoił metodykę badań naukowych.

6 *Czy rozprawa stanowi oryginalne rozwiązanie problemu naukowego albo oryginalne rozwiązanie w zakresie zastosowania wyników własnych badań naukowych w sferze gospodarczej lub społecznej?*

Autor w podsumowaniu rozprawy wymienia swoje oryginalne liczne osiągnięcia. Z zmieszczonych tam na uwagę wydają mi się następujące:

- zastosowanie pojęcia entropii uproszczonej jako dodatkowego kryterium destylacji korpusu danych testowych;
- wykorzystanie operatora epigenetycznego; operator ten jest stosowany zarówno jako rozszerzenie algorytmu VEGA, jak i do uwzględnienia eliminacji osobnika w procesie selekcji (jako czynnik stresowy);
- rozszerzenie algorytmu VEGA o sterowanie zbieżnością;
- eksperymentalne wyznaczenie minimum skutecznego prawdopodobieństwa wystąpienia operatora epigenetycznego dla poszczególnych formatów plików podczas destylacji korpusu, a przede wszystkim:
- potraktowanie destylacji korpusu danych testowych jako problemu wielokryterialnego pokrycia zbioru.

Za szczególne osiągnięcie uważam także wybór systemu operacyjnego oraz środowiska programistycznego do badań. Co prawda Doktorant nie uzasadnił wyboru tej

szczególnej wersji Linuksa, ale samo odejście od schematycznego Windows jest warte podkreślenia. Przypomnę, że większość serwerów sieciowych pracuje pod kontrolą systemów linuksowych; jest to także system pierwszego wyboru dla urządzeń z systemami wbudowanymi.

Autor wskazuje także na kierunki dalszych badań nad rozpatrywanymi w rozprawie zagadnieniami. Nie będę tych sugestii przytaczał, stwierdzę tylko, że świadomość dalszych możliwości rozwoju świadczy to dobrze o procesie pracy badacza.

## 7 *Jakie są słabe strony rozprawy i jej główne wady?*

Nie znajduję istotnych merytorycznie słabych stron rozprawy, ani też wad obniżających istotnie ocenę pracy. Oczywiście, niektóre wątki badań poprowadziłbym inaczej, zgodnie z prawem autora do nieskrępowanej twórczości. Przedstawione dalej uwagi mają więc raczej charakter moich osobistych opinii lub też są stwierdzeniem oczywistych przeoczeń doktoranta.

Spis treści wydaje mi się zbyt szczegółowy i rozdrobniony. Zdarzają się dwa, trzy krótkie fragmenty na jednej stronie. Umieszczanie tego wszystkiego w spisie treści rozdmuchuje ów spis do czterech stron, co czyni ten spis mało poręcznym.

Nie sposób nie wspomnieć o terminologii oraz o używaniu terminów obcojęzycznych. Podkreślam tu stanowczo, że wyjaśnień w rodzaju: „My tak tego terminu używamy" nie przyjmuję, jako że po to są zasady języka polskiego, aby zasady te stosować. Posiadanie stopnia doktora (co z łaciny: nauczyciel, mistrz) zobowiązuje do odpowiednio mistrzowskiego używania rodzimego języka. A więc wychodząc z tego stwierdzenia przypominam o kilku zasadach naszego języka.

W języku polskim przyjmujemy, iż słowa obcojęzyczne używane są nieodmiennie. Jeśli jednak chcemy słowo taki odmieniać po polsku, piszemy owe słowo po polsku; tak więc: Linux ale już Linuksa, Tex ale już Techem itd. W naszym przypadku *locus* ale już lokusu.

Pojęcia „czarna (biała, szara) skrzynka" są pojęciami szeroko stosowanymi, pierwotnie prawdopodobnie w teorii systemów i teorii obwodów elektrycznych. Obecnie terminy te są powszechnie używane w wielu dziedzinach matematyki, statystyki, teorii systemów, informatyki itd. itd. Nie są to więc ani terminy specyficznie informatyczne, ani nawet stworzone specjalnie na potrzeby informatyki.

No i na końcu sam termin 'fuzzy'. Doktorant zna z pewnością termin 'fuzzy logic' i ładny polski odpowiednik tego terminu. Być może to zrozumienie pomoże uczynić dziwny polski język informatyki językiem zrozumiałym dla specjalistów innych dziedzin.

Często osoby stosujące słowa angielskie w miejsce słów polskich uzasadniają to tym, że dane słowo nie ma polskiego odpowiednika. Tłumacze doskonale wiedzą, że żadne słowo obce nie ma dokładnego polskiego odpowiednika. A tłumaczyć jednak trzeba.

Jak wspomniałem, praca ma poprawną stronę edytorską. Jednakże zawsze mnie dziwiło i dziwi w tym przypadku, iż po wielu latach pracy z edytorami tekstu – czy wręcz programami DTP – prace na stopień nadal bardziej przypominają zgrzebne maszynopisy sprzed lat, niż nowoczesne wydawnictwa. Praca doktorska może i powinna – moim zdaniem – wyglądać jak książka dobrego wydawnictwa. Nie wymagam od autora znajomości systemu TeX czy LaTeX, choć w instalacjach Linuksa te systemy zawsze się

znajdują; ale już sam MS Word daje potężne możliwości napisania pracy ładnej dla oka. Myślę, że jeśli Doktorant myśli o dalszym życiu naukowym, to z wymienionymi systemami powinien się zapoznać.

Odrębną sprawą jest niestaranna edycja wzorów matematycznych. Przy zapisywaniu wzorów nadal stosujemy polską interpunkcję, przecinki i kropki. Wzory wyróżnione w oddzielnym wierszu muszą mieć jednakowej wielkości czcionkę. No i znakiem mnożenia jest kropka. Gwiazdka oznacza mnożenie splotowe, o które z pewnością nie chodzi doktorantowi. Rozumiem, że gwiazdka jest błędnym zapożyczeniem z programowania.

Na koniec chciałbym przedstawić moją opinię dotyczącą wykorzystania algorytmów genetycznych. Opinia ta jest wynikiem wieloletniego doświadczenia wspartego lekturą znanej książki Rogera Penrose *Moda, wiara i fantazja*. Otóż praca naukowa, tak jak i wszystkie inne formy życia intelektualnego, podlega modzie. Teraz panującą modą w optymalizacji są różnego rodzaju algorytmy będą ce odbiciem zjawisk naturalnych. Autorów prac o takich algorytmach napędza mylne przekonanie, że zjawiska przyrodnicze są naturalnie zoptymalizowane. Nic bardziej mylącego. Zjawiska naturalne nie są optymalne, ale jedynie sprawne. Nikt przecież nie będzie twierdził, że człowiek jest obiektem optymalnym. Jest jedynie obiektem sprawnie działającym, jednak z wieloma ograniczeniami. Można powiedzieć, że wszystkie realizacje przyrody są obiektami optymalizowanymi, czyli w trakcie optymalizacji.

Rozumie to Doktorant, umieszczając w tezie pracy przymiotnik *efektywny*, a to znaczy jednak, że nie optymalny. Umieszczenie, jednakże, modnego algorytmu w swojej pracy gwarantuje właściwie przyjęcie pracy do druku w wielu wydawnictwach. Nie dziwię się autorowi, że taką drogę obrał.

## 9 PODSUMOWANIE

Reasumując – uważam, iż recenzowana rozprawa doktorska jest interesującą i oryginalną pracą badawczą. Autor badał problem, który ma istotne znaczenie z punktu widzenia poznawczego i praktycznego, a wyniki ocenianej rozprawy doktorskiej mają duże potencjalne znaczenie praktyczne.

Przeprowadzając swoje wywody, autor wykazał się wszechstronnością, dobrą znajomością wiedzy teoretycznej i praktycznej w reprezentowanej przez niego dyscyplinie, a także umiejętnością samodzielnego prowadzenia pracy naukowej. Doktorant osiągnął założone w pracy cele i wykazał trafność postawionej tezy naukowej.

Jestem przekonany, że praca spełnia wymogi stawiane rozprawom doktorskim w obowiązujących przepisach i wnoszę o dopuszczenie pracy do obrony. W trakcie postępowania będę głosował za przyznaniem Marcinowi Pachnikowi stopnia doktora nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja.