

Dr hab. Rafał Różycki, prof. PP
Instytut Informatyki
Politechnika Poznańska
ul. Piotrowo 2, 60-965 Poznań
rafal.rozycki@put.poznan.pl



Poznań, 19.04.2023

Recenzja rozprawy doktorskiej

mgr. inż. Leszka Adama Ciopińskiego

p.t.: *Synteza samoadaptacyjnych systemów czasu rzeczywistego
za pomocą rozwojowego programowania genetycznego*

będącej podstawą o ubieganie się o nadanie stopnia naukowego
doktora w dyscyplinie naukowej informatyka techniczna i telekomunikacja.

Niniejsza recenzja została przygotowana w odpowiedzi na uchwałę nr 402/2022 Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej z dnia 13.12.2022 r. w sprawie powołania komisji doktorskiej, określenia zakresu egzaminów doktorskich oraz powołania komisji przeprowadzających egzaminy doktorskie w przewodzie doktorskim mgr. inż. Leszka Ciopińskiego.

Postępowanie odbywa się w trybie przewidzianym w Ustawie z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz.U. poz.1668, z późn. zm.).

1. Ogólna charakterystyka rozprawy

Do recenzji przedstawiono rozprawę doktorską wydaną w roku 2022 nakładem wydawnictwa Politechniki Warszawskiej. Jedyńm autorem pracy jest Doktorant – mgr inż. Leszek Adam Ciopiński. Strona tytułowa wskazuje, że treść pracy zawiera wyniki badań prowadzonych w dziedzinie nauk inżynieryjno-technicznych, w dyscyplinie naukowej informatyka techniczna i telekomunikacja.

Praca napisana jest po polsku, zawiera też obszerne i wyczerpujące streszczenie w języku angielskim.

2. Cel i zakres rozprawy

Przedstawione do oceny materiał obejmuje wyniki badań nad opracowaniem innowacyjnej metody syntezy oprogramowania systemów czasu rzeczywistego. W systemach czasu rzeczywistego nakłada się na uruchamianie w nich oprogramowanie silne ograniczenia czasowe tak, by uzyskać gwarancję terminowej realizacji funkcji spełnianych przez taki system. W najprostszej sytuacji zakłada się tzw. linię krytyczną, czyli graniczny czas działania wszystkich komponentów oprogramowania, którego przekroczenie jest zabronione.

W pracy jako warstwę sprzętową systemu czasu rzeczywistego rozważa się nowoczesne architektury komputerowe, w których podstawowym elementem wykonawczym jest pojedynczy procesor wielordzeniowy lub zbiór równoległych procesorów. Istotną cechą takich systemów jest to, że podczas wykonywania tego samego zadania obliczeniowego poszczególne rdzenie procesora/processory mogą mieć różną wydajność obliczeniową oraz zużywać różne ilości energii.

W pracy przyjmuje się, że na uruchamianie w systemie oprogramowanie składa się skończony zbiór niepodzielnych zadań obliczeniowych o znanych typowych, minimalnych i maksymalnych wartościach czasów wykonywania na poszczególnych procesorach (rdzeniach). Każdemu czasowi wykonania zadania odpowiada pewne zużycie energii, którego wartość zależy zarówno od zadania jak i procesora. Dodatkowo zakłada się, że zadania nie mają określonego czasu gotowości do wykonania, ale między poszczególnymi zadaniami mogą występować relacje kolejnościowe.

Podstawową zaletą proponowanego w pracy systemu ma być zdolność uruchamianego w nim oprogramowania do adaptacji do zmiennych warunków zewnętrznych wpływających na działanie oprogramowania. Adaptacja ta uwzględniać ma przy tym konieczność dotrzymania przyjętej w systemie czasu rzeczywistego linii krytycznej, a zarazem – w razie możliwości - minimalizować zużycie energii.

Tak określony zakres badań uważam za bardzo interesujący i wychodzący naprzeciw potrzebom opracowywania wydajnych, odpornych na zakłócenia i oszczędnych energetycznie systemów czasu rzeczywistego. Systemy o takich cechach są szczególnie pożądane w zyskujących na popularności koncepcjach Internetu rzeczy i przetwarzania brzegowego.

W szczególności autor postawił sobie za cel zabudowanie systemu, którego działanie nie jest projektowane na bazie najgorszych (pesymistycznych) scenariuszy wykonania komponentów oprogramowania, lecz opiera się ono na scenariuszach typowych – co w zamyśle autora ma pozwolić na bardziej oszczędne wykorzystanie energii.

3. Teza rozprawa

Auto stawia w swojej pracy następującą tezę:

„Synteza oprogramowania systemów wbudowanych czasu rzeczywistego metodą RPG pozwala na uzyskanie większych możliwości samoadaptacyjnych systemu niż w przypadku stosowania istniejących metod przeszeregowania zadań”

4. Układ rozprawy

Oprócz streszczenia praca zawiera siedem rozdziałów, spis wykorzystanych oznaczeń oraz bibliografię.

We Wstępie (Rozdział 1) zarysowany jest kontekst prowadzonych badań, na wielu przykładach systemów rzeczywistych wskazane są różne aspekty samoadaptacyjności.

W Rozdziale 2 autor przedstawia pojęcie samoadaptacyjności w kontekście systemów informatycznych, opisuje klasę samoadaptacyjnych zorientowanych obiektowo systemów czasu rzeczywistego, wskazuje na analogie do ogólnego problemu rozdziału ograniczonych zasobów systemu komputerowego.

Rozdział 3 dedykowany jest przedstawieniu celu i przyjętej tezy rozprawy. Wymienia się w nim przyjęte założenia i zakres pracy badawczej.

Oryginalną wersję metody Rozwojowego Programowania Genetycznego (RPG) w zastosowaniu do rozważanego problemu optymalizacyjnego zaprezentowano w Rozdziale 4.

Aspekt zdolności do samoadaptacji proponowanego systemu do warunków zewnętrznych rozważany jest szczegółowo w Rozdziale 5. Wymieniono tu założenia dotyczące parametrów zadań, opisano algorytmy generowania harmonogramów ich wykonywania na podstawie danych zaszytych w genotypie rozwiązania, zaproponowano też sposoby oceniania uzyskanych rozwiązań.

Opisane wcześniej miary oceny uzyskiwanych rozwiązań stanowią ważny komponent procesu weryfikacji efektywności zaproponowanego podejścia. Dokonano jej na podstawie eksperymentów obliczeniowych opisanych w Rozdziale 6.

Ostatni rozdział podsumowuje wyniki przeprowadzonych badań.

Umieszczony na końcu rozprawy skorowidz użytych symboli i nazw poprawia wygodę zapoznawania się z jej treścią.

5. Dyskusja merytoryczna

A) Teza pracy i sposób jej dowodzenia

Autor postawił przed sobą zadanie udowodnienia postawionej tezy (str.27 – „Teza ta zostanie udowodniona...”), która jest dość śmiała. Twierdzi się w niej, że zaproponowane rozwiązanie jest w przypadku ogólnym obiektywnie lepsze (w sensie przyjętej „miary samoadaptacyjności”) od każdej innej metody, w której realizowane jest przeszeregowywanie zadań. W mojej ocenie przedstawiony w pracy sposób dowodzenia – m.in. na podstawie bardzo ograniczonego eksperymentu obliczeniowego, z formalnego punktu widzenia jest niewystarczający.

Ponadto wątpliwość budzi sama teza, w której uwypukla się znaczenie metody RPG. Wydaje się jednak, że jeżeli poszukiwać źródła raportowanej efektywności proponowanego podejścia, to tkwi ona w sposobie reprezentowania rozwiązania (budowa chromosomu) i przyjętej metodzie jego dekodowania (algorytmy mapowania). Podobną efektywność prawdopodobnie dałoby się uzyskać wykorzystując inne podejścia metaheurystyczne operujące na tej samej reprezentacji i metodzie dekodowania rozwiązania.

B) Sposób funkcjonowania systemu

Główną wątpliwość budzi sposób funkcjonowania projektowanego systemu. Jak rozumiem, procedura RPG jest wykorzystywana do znalezienia metody konstruowania energooszczędnego uszeregowania zadań, która byłaby odporna na (nieprzewidywalne) zakłócenia występujące podczas realizacji tych zadań. Odrębna metoda zaszyta jest w każdym osobniku populacji algorytmu RPG i przyjętych algorytmach dekodowania osobnika do finalnego uszeregowania zadań na procesorach (rdzeniach procesorów). Załóżmy, że RPG dla danej instancji problemu (danego oprogramowania) znalazł najlepszego (w sensie przyjętych kryteriów) osobnika. Autor na str.22 wspomina, że „w systemach czasu rzeczywistego jedynie szybkie metody szeregowania zadań mogą być używane do harmonogramowania lub przeszeregowania zadań w trakcie działania systemu”. W tym kontekście pojawiają się następujące więc pytania:

- czy algorytmy dekodujące osobnika realizowane są na tych samych zasobach obliczeniowych (procesorach/rdzeniach procesorów) co docelowe oprogramowanie uruchamiane w systemie czasu rzeczywistego (w streszczeniu deklaruje się prace nad systemem z wbudowanym mechanizmem przeszeregowania)?
- jeśli odpowiedź na powyższe pytanie jest twierdząca, to czy dokonano oceny czasowej złożoności obliczeniowej algorytmów dekodujących oraz ilości zużywanej przez te algorytmy energii?
- jaki proces i w jaki sposób kontroluje odstępstwa od przyjętego harmonogramu realizacji zadań obliczeniowych składających się na oprogramowanie systemu czasu rzeczywistego?

- czy ewentualna akcja przeszerogowania zadań podejmowana jest natychmiast po wykryciu odstępstwa, czy np. dopiero po zakończeniu zadania, którego wykonywanie trwa dłużej niż zaplanowano?
- jak poradzi sobie system z sytuacją, gdy opóźnione jest (ponad bezpieczną wartość buforową) zakończenie ostatniego zadania w zaplanowanym harmonogramie?
- a może akcja przeszerogowania zadań podejmowana jest dopiero na początku następnego cyklu uruchomienia oprogramowania w systemie? Jak długo ona trwa i jakie parametry zadań są brane pod uwagę – zaktualizowane na podstawie danych z poprzedniego cyklu?

Precyzyjnej odpowiedzi na powyższe przykładowe pytania trudno znaleźć w pracy (poza enigmatycznymi wyjaśnieniami na str. 19, 26), a są one kluczowe dla zrozumienia podstawowych zasad działania finalnej wersji całego systemu.

We Wstępie (str.12-13) wymienione i opisane są etapy działania systemu samoadaptacyjnego (Monitorowanie, Analiza, Planowanie i Wykonanie) – niestety w dalszym tekście nie pojawia się już odniesienie do tych terminów w kontekście prezentowanych wyników badań.

C) Sformułowanie problemu

W pracy brakuje rozdziału, który zawierałby pełny model projektowanego systemu, wszystkie przyjęte kryteria optymalizacji, występujące w nim ograniczenia, parametry oraz zmienne decyzyjne. Taki formalny zapis ułatwiłby śledzenie prezentowanych wyników badań.

Dobrą praktyką jest sformułowanie problemu w postaci problemu programowania matematycznego – tu tego zabrakło.

D) Klasyfikacja rozważanego problemu szeregowania

Rozważany w pracy problem szeregowania autor klasyfikuje jako *Resource Constrained Project Scheduling Problem* (RCPSP), co w mojej ocenie jest założeniem błędnym. W RCPSP zakłada się występowanie wielu różnych typów i kategorii zasobów, o które współubiegają się zadania (czynności) wchodzące w skład całego projektu. W klasycznym RCPSP każde zadanie charakteryzuje zapotrzebowanie (być może zerowe) na każdy typ zasobu występujący w systemie i nie ma możliwości wykonania zadania w alternatywny sposób (czyli przy zapotrzebowaniu na inne zasoby).

W recenzowanej pracy z zasobami utożsamia się procesory (rdzenie procesorów), co do których zakłada się w ogólności, że nie są identyczne i dlatego nie mogą być traktowane jako zasób tego samego typu. Przydzielając zadanie do procesora o innej charakterystyce, de facto realizujemy inny (alternatywny) sposób jego wykonywania.

W celu zamodelowania takiej sytuacji w kontekście problemów rozdziału zasobów (ang. *Project Scheduling*) należałoby przyjąć wariant ogólniejszy tej klasy problemów, czyli tzw. *Multi-Mode Project Scheduling Problem* (MM RCPSP). W mojej ocenie jest to jednak zbędne, gdyż przy przyjętych w pracy założeniach rozważany problem optymalizacyjny jest pewnym szczególnym przypadkiem klasycznego problemu szeregowania zadań zależnych, niepodzielnych na maszynach (procesorach) równoległych dowolnych, bez żadnych dodatkowych zasobów. Specyfika rozważanego maszynowego problemu szeregowania wynika jedynie z charakterystyki maszyn oraz przyjętego kryterium (-ów) optymalizacji. Sugestię tę potwierdza w pewnym sensie autor wskazując - jako referencyjne metody szeregowania - algorytmy LLF i EDF, czyli heurystyki dedykowane do rozwiązywania problemów szeregowania zadań na maszynach.

E) Używana terminologia

W pracy brakuje elementarnego rygoryzmu w używaniu przyjętych terminów, który winien charakteryzować pracę naukową. Ze zbyt dużą swobodą, zamiennie używa się na przykład pojęć: program – oprogramowanie – aplikacja; uszeregowanie – harmonogram; generacja – pokolenie – iteracja.

Chociaż w pracy energia nie jest traktowana jako zasób podwójnie ograniczony, jej zużycie jest uwzględniane w przyjętym kryterium optymalizacji. Na pewną krytykę zasługuje więc fakt iż w pracy nie rozróżnia się pojęć: moc vs. energia (lub pobór mocy vs. zużycie energii), zamiast tego używa się na przykład dość sztucznego pojęcia „jednostkowy koszt energetyczny energii” (patrz Definicja 5.7).

F) Problem numerowania zadań (węzłów) w grafie

W przedstawionej w pracy metodzie dekodowania genotypu w fenotyp, podstawowe znaczenie ma numer zadania, który utożsamiany jest z jego priorytetem. W pracy proponuje się prosty algorytm numerowania zadań (pkt. 4.2.4). W ogóle nie rozważa się jednak możliwości, że poprawna numeracja (uwzględniająca porządek zadań wynikający z relacji kolejnościowych między nimi) nie musi być jedyna. Konsekwencje tego przeoczenia są bardzo poważne. Najlepsze uszeregowanie może być bowiem w ogóle niesosiągalne, jeśli nie uwzględni się wszystkich dopuszczalnych porządków numerów wierzchołków w grafie zadań.

G) Założenia eksperymentu obliczeniowego

Jeśli przyjęte w eksperymentach założenie, że strojenie algorytmu RPG wykonywane jest dla każdej rozwiązywanej instancji problemu, należy przypuszczać, że podobną efektywność osiągnęłyby inne podejścia metaheurystyczne strojone w podobny sposób.

Cytowany w rozprawie algorytm GA autorstwa S. Hartmanna [61] podczas procesu ewolucji również posługuje się genotypem, który w finalne uszeregowanie (harmonogram) przekształcany jest za pomocą dedykowanych algorytmów, tzw. SGS (ang. *Schedule Generation Scheme*). Wykorzystanie w eksperymentach GA Hartmanna (lub innej metody metaheurystycznej), przy założeniu podobnych parametrów zadań (uwzględniających bufora czasowe), przyniosłoby ciekawy materiał porównawczy (abstrahując od faktu, że GA Hartmanna w wersji podstawowej nie nadaje się do maszynowych problemów szeregowania). Porównywanie rozwiązania bazującego na RPG, który jest strojony do rozwiązywania konkretnej instancji (!), z wynikiem działania prostej heurystyki konstrukcyjnej (LPLLF) nie pozwala na sformułowanie rzetelnych wniosków.

H) Uproszczony sposób oceny rozwiązania

Autor sam sygnalizuje problem z rzetelną oceną generowanych przez proponowaną metodę rozwiązań. Wymagałoby to złożonej analizy statystycznej uwzględniającej różne rozkłady prawdopodobieństwa zmiany parametrów zadań obliczeniowych wchodzących w skład oprogramowania. Ze względu na konieczność częstej oceny rozwiązań – podejście takie jest nierealne ze względu na niezbędny czas obliczeń. Autor proponuje więc uproszczony sposób oceny rozwiązań, zakładając w eksperymentach możliwość wystąpienia tylko bardzo ograniczonego zbioru zakłóceń realizacji harmonogramu. Chociaż intencje autora wydają się zrozumiałe, takie podejście podważa jednak sens wyciągania wniosków ogólnych z eksperymentu, w którym używa się procedury szacowania jakości rozwiązania na bazie niereprezentatywnego podzbioru zakłóceń realizacji harmonogramu.

I) Inne uwagi ogólne

- Nie do końca jest jasne dla jakiego systemu zaproponowano opisane podejście. Jeśli jest to system operacyjny (w szczególności – system operacyjny czasu rzeczywistego), w pracy przydałby się rozdział dotyczący systemów operacyjnych i stosowanych w nich mechanizmach zarządzania zasobami.
- Chociaż w pracy jawnie wykorzystywane są różne kryteria oceny poszukiwanego rozwiązania, trudno odnaleźć w niej jakiegokolwiek odniesienie do prac z zakresu optymalizacji wielokryterialnej. Prezentacja własnych wyników badań w odniesieniu do znanych podejść rozwiązywania problemów wielokryterialnych, pozwoliłaby wypełnić tę lukę.
- Nigdzie w pracy nie można znaleźć opisu działania operatora reprodukcji w proponowanym algorytmie RPG.
- Na energię nie są nałożone żadne ograniczenia, dlatego zarówno jej chwilowa dostępność (czyli moc) jak i zużycie są nieograniczone, co wydaje się założeniem mało praktycznym.
- Autor często posługuje się terminem „poprawny harmonogram” nie definiując wcześniej formalnie, co to oznacza. Domyślać się należy, że chodzi tu o uszeregowanie, w którym ostatnie zadanie kończy się przed zadaną linią krytyczną, czyli dopuszczalne.

J) Uwagi szczegółowe

- Na str. 13 autor wymienia technologię DVFS jako alternatywny sposób na osiągnięcie założonych w pracy celów. Jedynym przytoczonym argumentem za projektowaniem systemów samoadaptacyjnych jest stwierdzenie, że „(DVFS) ma jednak różne ograniczenia”. Jest to oczywiście stwierdzenie zbyt ogólne i wymagające przynajmniej skrótowego rozwinięcia.
- Na str. 19 pada stwierdzenie, że problem SZODZ (ang. RCPSP) „jest problemem NP-zupełnym”, choć wcześniej problem ten przedstawiany jest jako problem optymalizacyjny. W teorii złożoności obliczeniowej klasę problemów NP-zupełnych definiuje się jednak w kontekście problemów decyzyjnych, a nie optymalizacyjnych.
- Pierwsze zdanie na str.20 („Podejmowane wiele...”) demonstruje pewną niefrasobliwość autora w precyzyjnym formułowaniu przekazu. Niezgodnie z intencją autora (tak się domyślam!), zdanie to bowiem sugeruje, że podejmowane są próby rozwiązania problemu RCPSP jako takiego, a nie opracowania efektywnych metod poszukiwania rozwiązań różnych klas jego instancji.
- Na str. 20 pojawia się również termin „wielozadaniowe SZODZ” – bez dodatkowego wyjaśnienia, trudno dociec, co on oznacza.
- Na stronie 20 autor wspomina o klasie problemów SO SZODZ, czyli takich, w których występują „silne ograniczenia”. Nie sprecyzowano jednak kiedy ograniczenia mogą być interpretowane jako „silne” i czy jest to własność samej instancji problemu, czy też jest ona ściśle powiązana z algorytmem zastosowanym do rozwiązywania problemu.
- W punkcie 4.3.2 znajduje się opis reprezentacji rozwiązania (genotypu). W kontekście opisu ze str.53 (wprowadza się tam dodatkowy chromosom) okazuje się on jednak niepełny.

- Na str. 40 znajduje się odnośnik do Rys.4.4, którego interpretacja na tym etapie przedstawiania metody dekodowania genotypu w fenotyp (bez referencji od odpowiednich algorytmów) jest niemożliwa.
- Na str. 52 autor zauważa, że „im większe rezerwy czasowe, tym (...) zwykle większa energochłonność całego układu”. Stwierdzenie to stoi w sprzeczności z Definicją 5.6.
- Na str. 75 (pkt. 6.1.2) pada stwierdzenie, że „(...)wszystkie algorytmy zostaną dostrojone zgodnie z opisem z podrozdziału 6.1.1.”. Jako, że podrozdział 6.1.1 dotyczy strojenia metody RPG, zacytowane zdanie sugeruje, że w opisywanym eksperymencie obliczeniowy, algorytm RPG był ręcznie dostrajany do rozwiązywanej instancji problemu. W takiej sytuacji trudno utrzymać argument, że proponowany w pracy system ma własność samoadaptacji.

7. Strona redakcyjna rozprawy

Niektóre fragmenty pracy zmuszają czytelnika do uciążliwego wertowania kart – na przykład interpretacja Rys. 4.4 wymaga jednoczesnego dostępu do stron: 37, 40, 41, 46. W pracy występuje też dużo drobnych błędów redakcyjnych, błędy zawierają też niektóre rysunki (np. Rys. 6.7). Zdarzają się niedokończone zdania (np. koniec str. 93). Używane są akronimy, których znaczenie nie jest nigdzie wyjaśnione (DVFS – str.13, FPGA, LUT – str.22,...).

Zaletą pracy jest bez wątpienia poprawność stylistyczna prezentowanego materiału.

8. Wniosek końcowy

Jak wykazano powyżej, recenzowana rozprawa doktorska ma wiele mankamentów, zawiera sporo niejasności – raportowane w niej wyniki są trudne do interpretowania. Rzeczowa odpowiedź na większość (jeśli nie wszystkie) z uwag krytycznych w pełni potwierdziłaby predyspozycje Doktoranta do prowadzenia badań naukowych.

Mimo wielu uwag krytycznych rozprawa doktorska potwierdza ogólną wiedzę teoretyczną kandydata w dyscyplinie informatyka techniczna i telekomunikacja oraz podstawowe umiejętności samodzielnego prowadzenia pracy naukowej. Jej przedmiotem jest oryginalne rozwiązanie problemu naukowego, mającego potencjał zastosowanie praktyczne.

Na podstawie materiału zgromadzonego w recenzowanej pracy, za największe osiągnięcia Doktoranta uważam:

- zaproponowanie ogólnego podejścia do syntezy energooszczędnego oprogramowania spełniającego twarde wymagania czasowe nowoczesnego systemu czasu rzeczywistego, bazującego na reaktywnej procedurze korygowania proaktywnego harmonogramu wykonywania poszczególnych zadań wchodzących w skład oprogramowania;
- opracowanie oryginalnego sposobu reprezentowania i dekodowania metody konstruowania harmonogramu zadań obliczeniowych;
- dokonanie analizy efektywności zaproponowanego rozwiązania na podstawie wyników eksperymentu obliczeniowego.

Doceniając duży wkład Doktoranta w przygotowanie proponowanego podejścia do syntezy oprogramowania systemów czasu rzeczywistego, eksperymentów obliczeniowych i samej rozprawy, uwzględniając, że jest On dopiero na początku swojej pracy naukowej, składam się ku pozytywnej ocenie przedstawionej do recenzji rozprawy.

W związku z powyższym uznaję recenzowaną rozprawę za osiągnięcie wystarczające do skierowania wniosku mgr. inż. Leszka Adama Ciopińskiego, o nadanie stopnia naukowego doktora w dziedzinie nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja do kolejnych etapów postępowania.

Dr hab. inż. Rafał Różycki, prof. PP