

dr hab. inż. Jędrzej Musiał
Instytut Informatyki
Politechnika Poznańska
ul. Piotrowo 2, 60-965 Poznań
Jedrzej.Musial@cs.put.poznan.pl

Poznań, dn. 30.06.2020 r.

Recenzja rozprawy doktorskiej

mgra inż. Andrzeja Wilczyńskiego

zatytułowanej:

Blockchain-based task scheduling in computational clouds

1. Problem badawczy i jego znaczenie

Bitcoin jest pierwszą kryptowalutą, wprowadzoną do obiegu w 2009 r. De facto jest to system (finansowy/księgowy), który bazuje na obliczeniach i metodach kryptograficznych. Odpowiednie obliczenia i ich wyniki, które są przechowywane w zdecentralizowanym systemie pozwalają na obrót umownymi jednostkami walutowymi. Wraz z szybkim (w pewnym momencie wręcz lawinowym) wzrostem zainteresowania bitcoinem coraz większą popularność zyskiwała technologia blockchain, która umożliwiała funkcjonowanie kryptowaluty. Blockchain można opisać jako zdecentralizowaną bazę danych (nie posiada jednego centrum emisji informacji, czy tzw. urzędu certyfikacji, najważniejszej strony zaufanej). Baza funkcjonuje najczęściej w modelu open source i jest zbudowana w oparciu o architekturę peer-to-peer. Obecnie technologia blockchain zyskuje ogromną popularność (niesłabnącą od kilku lat), a to głównie za sprawą ogromnego potencjału użytkowego. Poza oryginalnym wykorzystaniem technologii (bitcoin) blockchain znajduje zastosowanie w takich sektorach jak choćby IoT (Internet of Things), transakcje giełdowe, obsługa ksiąg wieczystych, rachunkowych, akcyjnych, handel energią elektryczną, rejestr handlowy samochodów, rejestr danych medycznych i wiele innych.

W ramach swojej rozprawy doktorskiej mgr inż. Andrzej Wilczyński poruszył ciekawy i nieeksploatowany do tej pory sektor wiedzy związany z szeregowaniem zadań w chmurach obliczeniowych (harmonogramowanie). Szczególny nacisk został położony na dodatkowe aspekty związane z bezpieczeństwem (w związku z tym zdefiniowany został współczynnik *security level*). Opracowane zostało rozwiązanie nazwane *Secure Blockchain Scheduler*. Jest ono oparte na technologii blockchain. P. Wilczyński stworzył również oryginalny algorytm *Proof of Schedule*. Jest to rozwiązanie potwierdzające kto może publikować w sieci nowe treści i jest ono oparte na zasadach modelu konsensusu. W celu przeprowadzenia eksperymentów obliczeniowych zaimplementowany został autorski symulator nazwany *Blockchain Secure Cloud Scheduler*. Jako, że podstawa problemu badawczego jest dynamicznie zmieniającym się środowiskiem i bardzo atrakcyjną gałęzią nauki należy stwierdzić, że tezy postawione

przez mgr. Wilczyńskiego są istotne z punktu widzenia badawczego. Ich znaczenie, w zakresie modelowania uszeregowania w cloud computingu, jest istotne.

2. Wkład autora

Doktorant w ramach swojej rozprawy dotyczącej szeregowania zadań w środowisku chmurowym zawarł szereg oryginalnych i istotnych, z punktu widzenia obecnej wiedzy naukowej, wyników pracy badawczej. Jako najistotniejszy wkład należy wymienić:

- opracowanie nowej taksonomii dla problemu szeregowania zasobów w chmurze (opisane w sekcji 2.1),
- definicja nowego kryterium związanego z bezpieczeństwem w ramach szeregowaniem zasobów – tzw. *security level* (sekcja 3.3),
- opracowanie nowego mechanizmu szeregowania zasobów – blockchain-based scheduler (sekcja 6.1),
- zamodelowanie i zaimplementowanie nowego algorytmu nazwanego *Proof of Schedule* (sekcja 6.3),
- opracowanie i zaimplementowanie nowego narzędzia informatycznego służącego do przeprowadzania symulacji działania modułów szeregowania zadań w środowisku chmurowym. Narzędzie zostało nazwane *Blockchain Secure Cloud Scheduler Simulator*.

Dodatkowym atutem rozprawy i prowadzonych prac jest opracowanie i zaprezentowanie przeglądu istniejących technik i metod szeregowania (ang. mini survey). Uzyskane wyniki mają oryginalny charakter, który wnosi nową wiedzę w obszarach badań dotyczących szeregowania zadań w środowisku chmurowym, sieci blockchain, czy w ogólności zagadnienia bezpieczeństwa informacji w środowisku rozproszonym. Prace zostały już dostrzeżone przez społeczność międzynarodową, czego potwierdzeniem są pierwsze publikacje naukowe. Doktorant opublikował 8 prac, z czego wszystkie dotyczą tematyki poruszanej w ramach rozprawy doktorskiej. Obrazuje to prawidłowe skupienie doktoranta na obranym problemie badawczym. Źródłami publikacji są zarówno publikacje w czasopiśmie naukowych jak i konferencyjne. Najnowsza publikacja to *Modelling and simulation of security-aware task scheduling in cloud computing based on Blockchain technology* w czasopiśmie *Simulation Modelling Practice and Theory* (100 pkt. MNiSW), 2020 r.

Należy zwrócić uwagę, że wszystkie artykuły naukowe, będące bazą do rozprawy doktorskiej, to prace wieloautorskie. Określenie wkładu doktoranta w powyżej opisane kluczowe elementy (5 wymienionych powyżej) pozostaje nie w pełni jasne. Należy natomiast podkreślić, że wiele prac to publikacje pisane wspólnie z promotorem (p.prof. Joanną Kołodziej), co jednoznacznie wskazuje na znaczny udział prac doktoranta w finalnym wyniku prowadzonych badań. Ponadto treść i materiały zawarte w rozprawie stanowią rozszerzenie już opublikowanych wyników badań. Podczas czytania rozprawy moją uwagę zwróciło kilka kwestii, które warte byłoby doprecyzowania:

- W sekcji 2.1 zabrakło szerszych opisów wyjaśniających naturę i charakterystyki prezentowanych taksonomii. Opisy warto byłoby rozwinąć o podkreślenie motywacji, czy wyjaśnienie doboru nowych, innych parametrów, a także zwrócenie uwagi na kolejno pojawiające się aspekty związane z szeregowaniem zadań w chmurze;

- W ramach przygotowania algorytmów wybranych do testów eksperymentalnych wskazany byłby szerszy komentarz wyjaśniający słuszność wybranych rozwiązań;
- Zaprezentowanie aktualnej wiedzy dotyczącej algorytmów opartych o acykliczne grafy skierowane (DAG) uzupełniłoby pracę i pozwoliło na uzupełnienie badań (a może i wykorzystanie innych algorytmów). W szczególności dotyczy to modeli, w których w ramach grafów, analizuje się aspekt komunikacji (communication-aware directed acyclic graph);
- Czy oprogramowanie przygotowane przez doktoranta (symulator) jest szerzej udostępnione i wykorzystywane przez inne jednostki naukowe;
- Naturą działania systemów szeregowania w chmurze jest ich swobodna praca przez dłuższy czas. Podczas pracy systemu (co słusznie zauważył doktorant) kryteria, ich wagi i atrybuty systemu mogą ulegać zmianie. Przeprowadzenie badań wychodzących poza symulator uruchomiony na jednym urządzeniu z pewnością wzbogaciłoby pracę. Brak pierwiastka związany z testowaniem w warunkach operacyjnych (choćby jego namiastką) pozostawia pewien niedosyt, ale jednocześnie jest bramą do kontynuacji ciekawych badań obliczeniowych;
- Badania testowe (rozdział 7) zostały podzielone na szereg pojedynczych eksperymentów, gdzie dla każdego z nich wskazane zostały parametry i określono, kolejno, różne badane główne kryteria (funkcje celu?) takie jak: makespan (tutaj badania dotyczyły również zastosowania opracowanego wskaźnika *security level*), flowtime, economic cost, resource utilization. Wyniki poszczególnych eksperymentów zostały skomentowane jako odrębne testy. Przeprowadzenie zbiorczych eksperymentów, które badają wykorzystanie kilku kryteriów, z pewnością wzbogaciłoby pracę. Ponadto umożliwiłoby doktorantowi dopasowanie parametrów pracy algorytmu BS (lub opracowanie nowego na bazie istniejącego) w celu osiągnięcia lepszych rezultatów w szeregowaniu wielokryterialnym;
- Do głównych eksperymentów wykorzystane zostały 4 zestawy danych w których dobrano 2 liczby maszyn wirtualnych (32, 128), a także 2 liczby zadań (1024, 4096). Uważam, że mimo powołania na publikację temat powinien zostać rozwinięty w ramach rozprawy. Ponadto przeprowadzenie badań (a może właśnie jednego agregującego przy wykorzystaniu wielokryterialnej oceny) dla szerszego spektrum zadań i maszyn wirtualnych, pozwoliłoby zobrazować zachowanie kolejnych algorytmów dla zmieniających się wolumenów struktury sieci.

Pomimo powyższych komentarzy uważam, że wkład autora w obecnej formie jest znaczny i opiera się na weryfikacji obecnych modeli, zaprezentowaniu (również swoich) taksonomii, opracowaniu algorytmu szeregowania BS, przygotowaniu narzędzia symulacyjnego i przeprowadzeniu istotnych badań eksperymentalnych.

3. Poprawność

Rozprawa doktorska ma charakter teoretyczno-eksperymentalny. Jest napisana na 120 numerowanych stronach (poprzedzona 7 stronami nienumerowanymi zawierającymi stronę tytułową, streszczenia w języku angielskim i polskim, a także spis treści) w języku angielskim. Główna treść pracy zawiera się na stronach 1-95, po których umieszczono zestawienie rysunków, tablic, list, akronimów. Bibliografia jest zawarta na stronach 107-116 i obejmuje 115 pozycji literaturowych. Na końcu załączono dodatek A, podsumowujący

działalność naukowo-dydaktyczną doktoranta. Pierwsze pięć rozdziałów ma charakter pracy informacyjnej, wprowadzającej do tematu. Jednocześnie zawarto swoisty przegląd wykorzystywanych technologii, modeli, algorytmów. Można stwierdzić, że fragmenty zawierają materiały przeglądowe (survey), co osobiście odbieram jako wartościowe przeglądy naukowe. W rozdziale 1. zaprezentowane zostały podstawowe informacje dotyczące szeregowania zadań w chmurze, prezentacji problemu badawczego, czy obranych celów naukowych doktoranta. W rozdziale 2. zaprezentowano podstawową wiedzę dotyczącą szeregowania zadań w chmurze. W ramach wspomnianych wstępnych informacji brakuje mi rozdziału opisującego ogólne podstawy cloud computingu, a jeszcze „niżej”, teoretycznych podstaw prac informatycznych – modelownie matematyczne, klasy złożoności problemów kombinatorycznych (autor wielokrotnie wspomina o optymalizacji), czy szerokiej sekcji dotyczącej podstaw algorytmiki i struktur danych. Taki rozdział stanowiłby świetne wprowadzenie, szczególnie dla osób, które na co dzień zajmują się inną gałęzią informatyki. W 3. rozdziale opisany został problem szeregowania zadań w chmurach obliczeniowych. W tym wartościowym materiale zabrakło analizy konkretnych rozwiązań stosowanych w ramach szeregowania zadań (nie tylko w chmurze) – wiadomości zostały zasygnalizowane w sekcjach 1.1 i 1.2, ale tutaj (w rozdziale 3.) zdecydowanie warto byłoby rozwinąć opisy konkretnych przykładów. W szczególności klasycznych, dobrze sprawdzonych algorytmów szeregowania (wtedy stanowią świetne odniesienie przy testach dla problemów nowej natury), jak również podejść będących na granicy klasycznego szeregowania i szeregowania w środowisku rozproszonym (jak choćby wspomniany na stronie 4. DAG (tutaj w szczególności ciekawe wydaje się wykorzystanie CA-DAG (Communication-Aware Directed Ayclic Craph)). Rozdział 4. wprowadza czytelnika w arkany technologii blockchain. Istotne jest tutaj (choć zaprezentowane bardzo skrótowo) opisanie aspektów bezpieczeństwa, szczególnie ze względu na spory nacisk, położony przez autora na kwestie wykorzystania nowego wskaźnika *security level*. Bardzo ważnym elementem, dla ciągłości logicznej przedstawionej rozprawy doktorskiej, jest rozdział 5. dotyczący gier i modelu Stackelberga. Zaprezentowane zostały podstawowe informacje związane z teorią gier i modelem Stackelberga, na podstawie którego doktorant opracował swój autorski algorytm *Proof of Schedule*.

Za kluczowe dla pracy uważam rozdziały 6 i 7. W ramach rozdziału 6. doktorant przedstawia i opisuje zasady działania swojego autorskiego rozwiązania służącego do szeregowania zadań przy wykorzystaniu technologii blockchain – *Secure Blockchain Scheduler*. Warto zauważyć, że model został tak przygotowany, aby podczas generowania optymalnego uszeregowania, brać pod uwagę wymagania klientów, jak i dodatkowe wymagania związane z bezpieczeństwem. Uzupełnieniem modelu jest przygotowany przez doktoranta *Blockchain Secure Cloud Scheduler Simulator* – symulator zaimplementowany w środowisku Java. Rozdział 7. zawiera opisy eksperymentów obliczeniowych, które miały zweryfikować skuteczność działania nowego algorytmu (BS). Badania eksperymentalne zostały podzielone na szereg pojedynczych eksperymentów, gdzie dla każdego z nich wskazane zostały parametry i określono, kolejno, różne główne kryteria (funkcje celu?) takie jak: makespan (tutaj badania dotyczyły również zastosowania opracowanego wskaźnika *security level*), flowtime, economic cost, resource utilization. Wyniki poszczególnych eksperymentów zostały skomentowane jako odrębne testy. Przeprowadzenie zbiorczych testów, które badają wykorzystanie kilku kryteriów, z pewnością wzbogaciłoby pracę. Ponadto umożliwiłoby doktorantowi dopasowanie parametrów pracy algorytmu BS (lub opracowanie nowego na bazie istniejącego) do osiągnięcia lepszych rezultatów w szeregowaniu wielokryterialnym. Do głównych

eksperymentów wykorzystane zostały 4 zestawy danych w których dobrano 2 liczby maszyn wirtualnych (32, 128), a także 2 liczby zadań (1024, 4096). Pomimo powołania na publikację temat powinien zostać rozwinięty. Ponadto przeprowadzenie badań (a może właśnie jednego agregującego przy wykorzystaniu wielokryterialnej oceny) dla szerszego spektrum zadań i maszyn wirtualnych, pozwoliłoby zobrazować zachowanie kolejnych algorytmów dla zmieniających się rozmiarów środowisk testowych. Dodatkowo uzupełnienie eksperymentów o jeszcze jedną (poza Hybrid Heuristic Method based on Genetic Algorithm, HSGA) implementację, któregoś z bardziej zaawansowanych algorytmów szeregowania zadań, z pewnością wzbogaciłoby wyniki analizy. Zastosowanie trywialnych rozwiązań jak First Come First Served, Shortest Job First, czy Round Robin stanowią świetne odwołanie dla nowych, wyrafinowanych propozycji, ale raczej nie będą dla nich konkurencją. Szczegóły dotyczące algorytmu HSGA powinny być zamieszczone w ramach pracy, a nie tylko jako odwołanie do literatury.

Podsumowując uważam, że doktorant zachował poprawność w realizacji swoich kolejnych prac, a agregująca dorobek naukowy treść rozprawy doktorskiej jest poprawna od strony dedukcji naukowej i struktury pracy. Wymienione braki wskazują miejsca, które mogłyby być przygotowane lepiej, ale nie są zarzutem obniżającym moją ocenę w ramach niniejszego kryterium.

4. Wiedza kandydata

Doktorant w swojej rozprawie wielokrotnie potwierdza głęboką wiedzę w zakresie informatyki, a w szczególności w ramach cloud computingu, szeregowania zadań, czy tematyki blockchain. Cytowania obejmujące 115 pozycji stanowią, że bibliografia jest obszerna i wyczerpująca. Zauważam, że doktorant odwołuje się również do wielu najnowszych źródeł pochodzących z ostatnich lat. To bardzo ważne przy tak dynamicznie rozwijającym się zagadnieniu. Ponadto jak słusznie zauważa tematyka jest tak szybko ewoluująca, że nie sposób zaprezentować dorobek kilku lat pracy i stwierdzić, że wszystkie proponowane rozwiązania są idealne i nic nie można lepiej zrobić. Świadomość tego aspektu, który zauważa sam doktorant jest bardzo istotna. Może również być doskonałą wskazówką do przyszłych prac badawczych.

Wiedza zaprezentowana przez doktoranta jest obszerna i obejmuje wiele aspektów. W ramach badań dokonano szerokiego przeglądu klasycznych rozwiązań i, co bardzo istotne, zaproponowano swoje autorskie rozwiązania w postaci taksonomii, algorytmu, symulatora, czy w końcu wartościowych badań eksperymentalnych.

5. Inne uwagi

Mgr inż. Andrzej Wilczyński jest współautorem 8 publikacji naukowych (jedna z nich to publikacja w tzw. czasopiśmie z listy filadelfijskiej). Kolejna publikacja w prestiżowym czasopiśmie *Sensors* znajduje się w recenzji. Uczestniczył w realizacji jednego projektu badawczego (ICT COST). Przygotował recenzje artykułów naukowych (3), uczestniczył w konferencjach naukowych (6). Brał udział w organizacji konferencji i wydarzeń naukowych (7). Prowadzi (-ł) zajęcia dydaktyczne z 6 „przedmiotów”. Otrzymał 2 Nagrody Rektora za osiągnięcia organizacyjne (2018 r., 2019 r.).

Rozprawa posiada logiczną strukturę rozdziałów oraz sekcji. Jakość wykorzystanego języka angielskiego oceniam wysoko. Pracę czyta się dobrze i płynnie. Wykorzystane rysunki i tablice ułatwiają

zrozumienie prezentowanych treści, a także poprawiają ogólną atrakcyjność publikacji. Poza uwagami sygnalizowanym w powyższych punktach niniejszej recenzji dostrzegam kilka uwag edytorskich:

- w tytule każde słowo powinno się zaczynać z wielkiej litery (podobnie jak to jest stosowane w nazwach rozdziałów),
- sugerowałbym nie korzystanie z liter diakrytycznych, szczególnie w nazwisku autora, czy mieście. Może to powodować problemy przy cytowaniu pracy i poprawnym indeksowaniu w międzynarodowych bazach,
- wykorzystanie zarówno kursywy jak i apostrofów (np. *'cloud computing'*) jest nieczytelne. Ponadto nie jest to apostrof otwierający i zamykający, a dwa zamykające,
- szereg bardzo drobnych błędów edytorskich (spacja przed kropką, brak zamknięcia nawiasu, brak spacji przed myślnikiem, podwójna kropka, itp.),
- rysunki są bardzo niespójne – nie zachowują podobnej czcionki (rozmiar), nie są podobnej szerokości (mimo np. ułożenia poziomego), nie zachowują tonacji kolorystycznej (raz „chmura” jest szara, stronę dalej już czarna,
- tablica 2.1 wyglądałaby lepiej i czytelnie gdyby była zawarta na jednej stronie (np. mniejsza szerokość niektórych kolumn czy mniejsze marginesy),
- drobne pomyłki jak np. „cpus” -> „CPUs”,
- bibliografia jest niespójna w kilku aspektach. Po pierwsze autor nie używa jednej formy podawania imion i nazwisk. Czasami wykorzystuje inicjał imienia i pełne nazwisko (np. poz. 3), a za chwilę wpisane jest pełne imię (poz. 5). Co więcej taki niespójny zapis zdarza się przy tym samym autorze (np. poz. 4 i 5). W innym przypadku zamieniona jest kolejność imię, nazwisko – nazwisko, imię. Występuje to nawet w przypadku tych samych pozycji literaturowych (np. poz. 4). Niekiedy nazwy źródeł wpisane są wielkimi literami. Schemat wpisywania DOI jest niespójny – czasem jest to numer, innym razem odwołanie do adresu strony internetowej. Ponadto uważam, że przy publikacjach wydanych już w czasopiśmie (gdzie wskazany jest wolumen, wydanie, strony, itd.) wpisywanie DOI jest zbędne i tylko zaburza przejrzystość tekstu. Nazwiska autorów powinny być wyraźnie oddzielone (np. w poz. 50 tego zabrakło). Ponadto (odwołując się do tej samej pozycji literaturowej) powinno się wpisywać autorów wedle kolejności podanej w źródłowej pracy,
- zdania bywają za długie, co nieznacznie utrudnia zrozumienie pewnych fragmentów tekstu,
- miejscami brakuje przecinków,
- autor korzysta, w pewnych momentach, ze zbyt pospolitego słownictwa,
- występują (w niewielkiej ilości) drobne błędy gramatyczne,
- na stronie 103 zawarto „List of Abbreviations”. Poprawnym nazewnictwem byłoby „List of Acronyms”, gdyż lista zawiera spis skrótowców, a nie skrótów.

Zaznaczam, że wskazane błędy nie wpływają jednak na merytoryczną wartość pracy.

6. Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez artykuł 13 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym (z późniejszymi zmianami) moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego?



Zdecydowanie
TAK



Raczej TAK



Trudno
powiedzieć



Raczej NIE



Zdecydowanie
NIE

B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka?



Zdecydowanie
TAK



Raczej TAK



Trudno
powiedzieć



Raczej NIE



Zdecydowanie
NIE

C. Czy kandydat posiada umiejętności samodzielnego prowadzenia pracy naukowej?



Zdecydowanie
TAK



Raczej TAK



Trudno
powiedzieć



Raczej NIE



Zdecydowanie
NIE

Jędrzej Musiał