

Politechnika Warszawska
Wydział Elektroniki i Technik Informatycznych

DZIEKAN I RADA WYDZIAŁU ELEKTRONIKI I TECHNIK INFORMACYJNYCH
POLITECHNIKI WARSZAWSKIEJ

zawiadamiają o

PUBLICZNEJ OBRONIE ROZPRAWY DOKTORSKIEJ

mgr. inż. Sławomira Bąka

która odbędzie się w dniu czerwca 18 czerwca 2019 r. o godzinie 11.00

na Wydziale Elektroniki i Technik Informatycznych Politechniki Warszawskiej
Gmach im. Janusza Groszkowskiego, Warszawa ul. Nowowiejska 15/19, w sali nr 116

Tytuł rozprawy doktorskiej:

„Synteza rozproszonych aplikacji czasu rzeczywistego przetwarzanych w chmurze obliczeniowej”

promotor: dr hab. inż. Roman Deniziak, prof. PŚw. Wydział Elektrotechniki, Automatyki i Informatyki Elektroniki i Technik Informatycznych,

promotor pomocniczy: dr inż. Radosław Czarnecki z Katedry Elektrotechniki Teoretycznej i Informatyki Politechniki Krakowskiej,

recenzenci: dr hab. inż. Paweł Czarnul, prof. Politechniki Gdańskiej Wydział Elektroniki, Telekomunikacji i Informatyki,

dr hab. inż. Tadeusz Nowicki, prof. Wojskowej Akademii Technicznej Wydział Cybernetyki.

Na stronie internetowej wydziału www.elka.pw.edu.pl/Wydzial/Rada-Wydzialu/Harmonogram-obron-doktorskich-streszczenia-i-recenzje znajdują się streszczenie rozprawy oraz recenzje, jak również dostęp do tekstu rozprawy umieszczonej w Bazie Wiedzy Politechniki Warszawskiej.

Dziekan



prof. dr hab. inż. Krzysztof Zaremba

Streszczenie

Rola chmur obliczeniowych w IT dynamicznie wzrasta. Chmury wypierają serwery dedykowane, ze względu na swoją elastyczność w rozbudowie i model płatności zależny od bieżącego użycia, który pozwala płacić użytkownikom za to, czego naprawdę potrzebują. Wiele przedsiębiorstw rezygnuje z tworzenia aplikacji desktopowych dla swoich klientów, na rzecz aplikacji uruchamianych w swojej prywatnej chmurze obliczeniowej, co tworzy całkowicie nowy rynek usług informatycznych. Dodatkowo dzięki rozwojowi takich dziedzin jak IoT, S2aaS, HPCaaS (High Performance Computing as a Service) wzrasta potrzeba rozwoju chmur czasu rzeczywistego (RTCC). Rozwój dziedziny RTCC wymaga opracowania efektywnych metod, zarządzania i projektowania aplikacji dla RTCC, wyspecjalizowanych do syntezy rozproszonych aplikacji czasu rzeczywistego przetwarzanych w chmurach obliczeniowych, umożliwiających zmniejszenie kosztów wynajęcia zasobów oraz ich efektywniejsze wykorzystanie.

W niniejszej pracy przedstawiona została nowa metoda syntezy i optymalizacji rozproszonych aplikacji czasu rzeczywistego dedykowanych do wykonywania w środowisku chmury obliczeniowej (ACCR). Głównym zadaniem opracowanego algorytmu RMO^{ACCR} jest obniżenie kosztów korzystania z zasobów chmury obliczeniowej przy równoczesnym zwiększeniu efektywności ich wykorzystania, poprzez współdzielenie zasobów między aplikacjami. Opracowany algorytm charakteryzuje się złożonością obliczeniową umożliwiającą zastosowanie go w praktycznych zastosowaniach, co zostało wykazane poprzez eksperymenty, a także poprzez analizę teoretyczną. Do algorytmu dobrano funkcję zysku, sterującą procesem rafinacji, która uwzględnia nie tylko koszt wykonania ACCR, ale także jej wydajność. Te parametry algorytmu zmniejszają prawdopodobieństwo zatrzymania się algorytmu w lokalnym minimum kosztu, co jest wadą innych metod. Zastosowanie opracowanej metody syntezy pozwoli nie tylko na znaczne obniżenie kosztów po stronie klienta IaaS, ale także obniży koszty dostawcy usług związane z utrzymaniem infrastruktury. Ponadto efektywne wykorzystanie zasobów pozwala na udostępnienie ich dla szerszego grona odbiorców, zwiększając tym samym zyski z wynajmowania infrastruktury.

Słowa kluczowe: przetwarzanie w chmurze, infrastruktura jako usługa, system czasu rzeczywistego, chmura czasu rzeczywistego, IoT.

Abstract

The role of computing clouds in IT is growing dynamically. Clouds replace dedicated servers due to their flexibility in expansion and the payment model depends on the current usage, what allows users to pay for what they really need. Many companies give up the creation of desktop applications for their clients, for applications launched in their private computing cloud, which creates a completely new field of IT services. In addition, due to the development of such areas as IoT, S2aaS, HPCaaS (High Performance Computing as a Service), the need to develop real-time clouds (RTCC) is increasing. The development of the RTCC field requires the development of effective methods, management and design of applications for RTCC, specialized for the synthesis of distributed real-time applications processed in cloud computing, enabling the reduction of resource rental costs and their most efficient usage.

This PhD thesis presents a new method of synthesis and optimization of distributed real-time applications dedicated to perform in the cloud computing environment (ACCR). The main task of the developed RMO^{ACCR} algorithm is to reduce the costs of using cloud computing resources while increasing the efficiency of their use, by sharing resources between applications. The developed algorithm is characterized by computational complexity, enabling its usage in practical applications, which was demonstrated through experiments as well as through theoretical analysis. The algorithm has a profit function that controls the refining process, which not only takes into account the cost of ACCR performance, but also its performance. These algorithm parameters reduce the likelihood of the algorithm stopping at the local minimum cost, which is a disadvantage of the other methods. The use of the developed synthesis method will not only allow for a significant costs reduction on the IaaS customer side, but also reduce the service provider's costs related to maintaining the infrastructure. In addition, the effective usage of resources allows them to be made available for a wider group of recipients, thus increasing profits from renting infrastructure.

Keywords: cloud computing, infrastructure as a service, real-time system, real-time cloud, IoT.

***KWESTIONARIUSZ-RECENZJA ROZPRAWY DOKTORSKIEJ DLA RADY
WYDZIAŁU ELEKTRONIKI I TECHNIK INFORMACYJNYCH
POLITECHNIKI WARSZAWSKIEJ***

Tytuł rozprawy: Synteza rozproszonych aplikacji czasu rzeczywistego przetwarzanych w chmurze obliczeniowej

Autor rozprawy: mgr inż. Sławomir Bąk

1. Jakie zagadnienie naukowe jest rozpatrzone w pracy (teza rozprawy) i czy zostało ono dostatecznie sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?

Rozprawa doktorska mgr. inż. Sławomira Bąka związana jest z aktualną dla informatyki problematyką opracowania modeli i metod formalnych oraz implementacyjnych służących zwiększeniu efektywności obliczeń zestawu zadań realizowanych w chmurze obliczeniowej. Chociaż problematyka ta jest już od wielu lat rozpatrywana w literaturze, to jednak możliwość konstruowania i wykorzystania modeli oraz metod efektywnego wykonywania zadań w systemach komputerowych zwanych chmurą obliczeniową spowodowała rozszerzenie możliwości w tym względzie. Chmury obliczeniowe są swego rodzaju zbiorem połączonych i wynajmowanych systemów komputerowych o zwiększonej, sumarycznej mocy obliczeniowej oraz pamięci potrzebnej do realizacji zadań. W systemach tego typu oferowane są zarówno moce obliczeniowe, jak i pamięć, które w zależności od lokalizacji obliczeń są tańsze lub droższe. Zwykle jednak zwiększona moc obliczeniowa niesie za sobą zwiększone koszty obliczeń, to samo jest z oferowaną pamięcią komputerową. Istnieje możliwość wyboru środowisk obliczeniowych dla zadań realizowanych w chmurze. Czasem wybór ten jest po stronie tego, który obliczenia te zleca, a czasem pozostaje to w gestii administratora chmury obliczeniowej. Patrząc z dwóch różnych perspektyw, zlecającego obliczenia lub realizującego te obliczenia, pojawia się ten sam problem. Polega on na odpowiedzi na pytanie: z których zasobów chmury obliczeniowej należy korzystać, aby obliczenia te zrealizować w odpowiednim czasie i przy spełnieniu licznych dodatkowych ograniczeń, a jednocześnie, aby koszt obliczeń był jak najmniejszy. Ponadto, uwzględnia się często przepustowość sieci komputerowej, jaka wiąże ze sobą poszczególne centra obliczeniowe chmury obliczeniowej. Tak zdefiniowany problem wyznaczania strategii obliczeń w chmurze jest zarówno nietrywialny, jak też coraz częściej formułowany tym bardziej, że mamy do czynienia z coraz bardziej złożonymi obliczeniami, które muszą być zrealizowane w stosunkowo krótkim czasie. Mowa tu o obliczeniach dedykowanych dla systemów czasu rzeczywistego. Ten problem jest właśnie rozpatrywany w przedstawił do recenzji rozprawie.

Rozdział pierwszy jest wstępem do rozprawy. W pierwszym rozdziale Autor informuje, że praca związana jest z realizacją zadań stanowiących rozproszoną aplikację czasu rzeczywistego. Dokonuje klasyfikacji wieloprocessorowych systemów przetwarzania zadań z

punktu widzenia szeregowania wielu zadań na wielu procesorach. Pokazuje wyniki prac naukowych w tym zakresie uwzględniając różne typy problemy szeregowania, w tym takie z heterogenicznymi lub homogenicznymi procesorami, ze statycznym i dynamicznym priorytetowaniem zadań, z określonym zapasem czasu dla realizacji zadań, z zadaniami niecierpliwymi, itp. Dla wszystkich typów problemów charakteryzowane są poznane w literaturze metody rozwiązywania dokładnego i heurystycznego sformułowanych problemów szeregowania zadań. Pokazywane są przy tym wady i zalety różnych modeli szeregowania zadań oraz metod rozwiązywania sformułowanych na ich podstawie zadań optymalizacji. Pokazano również nowe własności problemów szeregowania zadań wynikające z tego, że zadania są realizowane w chmurze obliczeniowej. Chodzi tu przede wszystkim o to, że w szeregowaniu zadań w chmurze obliczeniowej dochodzi ciekawy aspekt sterowania obliczeniami zadań uwzględniający możliwość ponoszenia różnych kosztów ich realizacji oraz w mniejszym stopniu uwzględnienie różnych przepustowości w podsystemie komunikacji. W rozważanych problemach sterowania realizacją zadań poszukiwane jest takie rozwiązanie, które dąży do uzyskania jak najkrótszego czasu zakończenia realizacji zadań przy jak najmniejszym ponoszonym przy tym koszcie. Oczywiście, należy również uwzględnić szereg ograniczeń nakładach na proces realizacji zadań. Przegląd metod nie ogranicza się tu do klasycznych sposobów definiowania i rozwiązywania problemów szeregowania zadań w systemach wieloprocessorowych z różnego typu dodatkowymi ograniczeniami.

Autor odnosi się również do poznanych w literaturze problemów tego typu związanych z wykorzystaniem chmur obliczeniowych. Omawia liczne metody rozwiązywania problemów szeregowania zadań zakładających wynajmowanie zasobów chmury obliczeniowej. Podawana jest charakterystyka proponowanych metod rozwiązywania problemów szeregowania zadań wykorzystujących między innymi: modele prognozowania przy wyborze najlepszych zasobów chmury w celu optymalizacji wydajności aplikacji, wykorzystanie mniejszej od maksymalnej liczby zasobów w chmurze przy jednoczesnym spełnieniu wymagań czasowych na realizację zadań, wynajmowanie zasobów chmury obliczeniowej uwzględniające gwarancję odpowiedniej przepustowości łączy danych, aplikacje skracające czas wykonania aplikacji przy uwzględnieniu zużycia energii systemu, itp. jednak szczególną uwagę skupia Autor na problemach szeregowania zadań z minimalizacją kosztów wynajmowanych zasobów. Pokazuje przy tym, że najefektywniejsze metody rozwiązywania problemów tego typu bazują na idei algorytmów genetycznych. Wyznacza się wtedy pewne rozwiązanie lub zbiór rozwiązań dopuszczalnych kierując się pewnymi zdroworozsądkowymi zasadami i w kolejnych krokach algorytmu poprawiane są uzyskiwane wcześniej rozwiązania. Dbą się przy tym o to, aby w kolejnych rozwiązaniach spełnione były założenia dodatkowe, charakterystyczne dla konstruowanego modelu szeregowania zadań. Prowadzi to do uzyskania rozwiązania suboptymalnego. Muszę przyznać, że analiza dostępnych w literaturze metod formułowania i rozwiązywania problemów szeregowania zadań w systemach wieloprocessorowych, w tym w chmurach obliczeniowych, została dokonana profesjonalnie, chociaż na pewnym poziomie ogólności rozważań. Brak tu pewnych zapisów formalnych, które mogłyby wzbogacić istotnie wiedzę czytelnika w tym zakresie.

Rozdziały trzeci i czwarty, to prezentacja motywacji powstania rozprawy i sformułowania jej tezy. Teza badawcza sugeruje to, że zastosowanie specjalizowanego algorytmu do syntezy rozproszonych aplikacji czasu rzeczywistego przetwarzanych w chmurze obliczeniowej pozwoli na istotne obniżenie kosztów wynajmowania i stopnia wykorzystania zasobów chmury obliczeniowej.

Rozdział piąty poświęcony został w całości modelowaniu systemu chmurowego czasu rzeczywistego. Przedstawione zostały tu, istotne dla opracowanej metody, założenia odnoszące się do rozwiązywanego problemu. Zakłada się, że rozważany jest problem

szeregowania pewnej grupy zadań wykonywanych cyklicznie. Ta grupa zadań nazywana jest aplikacją rozproszoną. Każde zadanie stanowi skończony zbiór obliczeń, przy czym znane są złożoności obliczeniowe poszczególnych zadań. Występuje przy tym zależność w realizacji poszczególnych zadań. Niektóre z nich muszą być poprzedzone zakończeniem niektórych innych zadań. Zależności te zdefiniowane zostały w postaci grafu, gdzie wierzchołkami są zadania z rozpatrywanej grupy, a łuki definiują relację następstwa w ich realizacji. W pewnym sensie następstwa te wynikają z konieczności komunikacji pomiędzy zadaniami. Autor nazywa ten graf grafem zadań.

Dalej zdefiniowana została aplikacja rozproszona. W mniejszym stopniu przedstawiany jest tu model matematyczny aplikacji, raczej prezentowana jest ona przez podanie pewnych jej charakterystyk i procedur wykonywania. Omawia się też architekturę systemu chmurowego czasu rzeczywistego. Definiowane są kanały komunikacyjne z ich przepustowościami i kosztami wynajęcia, serwery obliczeniowe z podaniem ich mocy obliczeniowej i kosztami wynajęcia tych mocy oraz kanały transmisji. Przyjmuje się, że czasy transmisji wynikają z jego rozmiaru, natomiast czasy wykonania zadań wynikają z ich złożoności obliczeniowej. Moim zdaniem zabrakło tu podania metodyki szacowania zarówno rozmiaru transmisji, jak też złożoności obliczeniowej zadań. W dalszej części rozdziału prezentowany jest problem szeregowania zadań. Definiowana jest ścieżka krytyczna w grafie zadań, będąca później podstawą do konstrukcji algorytmu szeregowania. Czas wykonania zadań ścieżki krytycznej zależy od przyporządkowania zadań do serwerów, zatem od czasów ich wykonania, oraz czasów transmisji zadań do tych serwerów. Wprowadza się pojęcie ograniczeń czasowych na zakończenie realizacji poszczególnych zadań, co jest charakterystyczne dla systemów czasu rzeczywistego. Mówi się też o jakości usług w systemie chmurowym, rozumianej jako aplikacja rozproszonych uruchomionych w chmurze obliczeniowej. Deklarowana jest taka strategia szeregowania zadań w chmurze obliczeniowej, która znajdowałaby kompromis między minimalizacją kosztu realizacji zadań i maksymalizacją jakości usług, gwarantującej spełnienie ograniczeń czasowych dla poszczególnych zadań.

W rozdziale szóstym przedstawiony został autorski algorytm wyznaczania rozwiązania szeregowania zbioru zadań definiowanych grafem zadań przy znajomości charakterystyk chmury obliczeniowej. Prowadzić ma on do wyznaczenia takiego przydziału zadań do serwerów obliczeniowych, aby uwzględnić następstwo realizacji zadań, minimalizując koszt wynajęcia zasobów chmury obliczeniowej przy jednoczesnym zachowaniu ograniczeń czasowych dla aplikacji rozproszonej. Konstrukcja prezentowanej metody polega na wykorzystaniu idei tak zwanego algorytmu rafinacji. Generowane jest tu rozwiązanie początkowe, gwarantujące najkrótsze czasy realizacji grupy zadań przy wykorzystaniu wszystkich dostępnych i jakościowo najlepszych serwerów chmury obliczeniowej. W kolejnych krokach modyfikowane jest to rozwiązanie uwzględniając możliwe osłabienie kryterium minimalizacji czasów realizacji zadań przy jednoczesnym obniżeniu kosztów tego procesu. Autor proponuje wykorzystanie kilku operatorów mutacji (modyfikacji) rozwiązań dotąd uzyskiwanych. Modyfikacje te polegają między innymi na zmianie kanału transmisji na tańszy, zmianie serwera obliczeniowego na tańszy, zmniejszenie zaalokowanych serwerów do realizacji zadań oraz zmianie serwera tańszego na węzeł efektywniejszy (droższy) ale jednocześnie przy zmianie liczby zaalokowanych serwerów, co prowadzi do zwiększenia jakości usług. Używam pojęcia operatora mutacji, ponieważ, chociaż wprost tego się nie mówi, że proponowana metoda jest faktycznie algorytmem genetycznym. Pomimo tego, że startuje się tu od pojedynczego rozwiązania, a w algorytmach genetycznych rozpatruje się populację początkową, to zastosowanie modyfikatorów istniejącego rozwiązania przez pewne zmiany prowadzące do uzyskania rozwiązania tańszego, przy jednoczesnym spełnieniu ograniczeń czasowych na realizację zadań jest charakterystyczne dla algorytmów genetycznych. Poprawa rozwiązania wynika stąd, że początkowe rozwiązanie gwarantuje w

większym lub mniejszym stopniu minimalizację czasów realizacji zadań. Autor definiuje tak zwany przez niego zapas czasu (dla zadań), dający możliwość osłabienia kryterium minimalizacji czasu łącznego (nazywane jest to swobodą rafinacji), co z kolei pozwala na zmianę przyporządkowania zadań do serwerów obniżając koszt realizacji zadań. Możliwości modyfikacji rozwiązania są różne i tu Autor kieruje się zdefiniowanym wskaźnikiem nazywanym zyskiem z rafinacji. Dodatkowym elementem metody jest wprowadzenie grafu zadań, czyli zależności pomiędzy realizacją zadań, do wykorzystywanego z literatury algorytmu „echo”.

Idea opracowanej metody jest klarowna i poprawnie przedstawiona. Prezentowane są też pewne jej własności, jak wykazanie zbieżności metody, w tym osiągnięcie warunku końca obliczeń oraz szacowanie złożoności obliczeniowej.

W rozdziale siódmym przedstawione zostały wyniki eksperymentów obliczeniowych pokazujące korzyść ze stosowania opracowanej metody w stosunku do pewnych innych algorytmów, w tym ilustracja wykorzystania metody do adaptacyjnego systemu nawigacji samochodowej. Całość kończy się podsumowaniem, wykazem ilustracji i tabel oraz spisem literatury.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle, świadczący o dostatecznej wiedzy autora? Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

Przegląd literatury zajmuje sporą część rozprawy i jest dokonany dość szczegółowo. Autor odnosi się tu w większości przypadków do wyników uzyskiwanych przy wykorzystaniu algorytmów ewolucyjnych, co wynika z tego, że taką koncepcję rozwiązania przyjął w swojej pracy. W tym sensie nie zawsze można odwoływać się do wszystkich wyników napotkanych w dostępnej literaturze i odnoszących się do całości rozprawy. Niemniej jednak, w licznych miejscach odwołuje się Autor do różnych pozycji literatury odnoszących się do zagadnień szczegółowych. Odwołania te są trafne i po każdym odwołaniu następuje ocena przydatności istniejących rozwiązań dla opracowywanej i oryginalnej metodyki.

Ocena przeglądu literatury jest zatem dokonana w stopniu dobrym. Nie jest ona być może zbyt rozbudowana, ale przez to nie powoduje powstania niepotrzebnych nadmiarów, natomiast dla czytelnika zaawansowanego w zagadnieniach poruszanych w pracy stanowi wystarczającą informację dla uzasadnienia tezy rozprawy i umiejscowienia jej w dotychczasowych dokonaniach w tym zakresie.

3. Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

Autor rozwiązał postawione zadanie badawcze, a użyta w tym celu metoda ewolucyjna, właściwie metodyka składająca się z szeregu transformacji uzyskanego na początku rozwiązania początkowego, jest właściwa i dobrze opisana. Świadczy też o dużej jego wiedzy w zakresie obejmującym tezę rozprawy. Założenia przyjęte w rozprawie są jasno i dobrze sprecyzowane. Założenia te można odnieść do licznych w praktyce aplikacji rozproszonych czasu rzeczywistego. Z tego punktu widzenia przyjęte w rozprawie założenia są uzasadnione. Warto przy tym podkreślić spójność, kompletność i zwięzłość sposobu, w jaki przedstawiona została opracowana metoda.

4. Na czym polega oryginalność rozprawy, co stanowi samodzielny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Autor opracował oryginalną metodę wyznaczania rozwiązania przydziału grupy zadań realizowanych do serwerów chmury obliczeniowej przy uwzględnieniu pewnych ograniczeniach czasowych i następstwach w ich realizacji. Metoda, chociaż jest pewnym standardem w rozwiązywaniu zadań tego typu, jest oryginalna. Pokazane zostały jej dobre własności w stosunku do innych metod znanych w literaturze. W tym sensie spełnia wymagania stawiane rozprawom doktorskim w odniesieniu do poziomu techniki reprezentowanej przez literaturę światową.

W ten sposób opracowana została oryginalna metodyka wyznaczania rozwiązania szeregowania grupy zadań na wielu procesorach uwzględniająca specyficzne własności chmur obliczeniowych.

5. Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?

Praca jest napisana w sposób zwięzły, bez niepotrzebnych przerostów. Jednocześnie napisana jest w sposób przejrzysty. Pewnym mankamentem rozprawy jest pewna liczba błędów i niezręczności redakcyjnych. Nie wpływają jednak one istotnie na moją ocenę pracy, jako całości. Wszelkie tezy są prezentowane w sposób nie budzący wątpliwości i jednocześnie są w dostateczny sposób uzasadniane.

6. Jakie są słabe strony rozprawy i jej główne wady?

Do słabszych stron rozprawy zaliczyłbym:

- brak formalnego (matematycznego) zdefiniowania rozwiązania problemu szeregowania zadań w chmurze obliczeniowej, to znaczy formalnego przedstawienia zmiennej decyzyjnej,
- omawia się kryteria optymalizacji, ograniczenia i jakość usług bez odniesienia się do formalnej definicji zmiennej decyzyjnej, chociaż zazwyczaj tak się czyni,
- nie wiadomo, czy wyznaczanie strategii realizacji grupy zadań jest stałe i nie mają na to wpływu zazwyczaj dynamicznie zmieniające się charakterystyki szczegółowe wydajności chmury obliczeniowej,
- nie uwzględnia się zatem istnienia innych obliczeń realizowanych w chmurze obliczeniowej,
- nie wiadomo jakiego rozmiaru są przesyłane w kanałach transmisji dane związane z poszczególnymi zadaniami wpływające istotnie na czasy transmisji,
- nie odniesiono się do tego, czy wyznaczenie serwera do realizacji zadania determinuje też jednoznacznie kanał transmisji danych,
- poprawiane jest pojedyncze rozwiązanie początkowe, oparte o ścieżkę krytyczną i silnie determinujące rozwiązanie finalne, a przecież można byłoby zastosować szereg rozwiązań początkowych, które przy wykorzystaniu operatorów mutacji i operatora selekcji dawałoby kolejne grupy rozważanych rozwiązań,
- metoda oparta jest o złożoność obliczeniową zadań, chociaż można byłoby, tak jak jest to robione w literaturze, szacować po prostu czasy realizacji poszczególnych obliczeń w rozpatrywanych zadaniach,

- nie widać tu związków między zadaniami i możliwością ich realizacji na poszczególnych serwerach, wynikającej między innymi z dostępnych tam usług,
- nie dość dobitnie przeanalizowano to, czy liczne, jednak drobne zmiany uzyskane w transformacji rozwiązania dają istotne zyski w doniesieniu do uzyskiwanego, kolejnego rozwiązania, raczej pokazuje się to, że modyfikacje rozwiązania spełniają ich dopuszczalność i pomniejszają koszty przy spełnieniu ograniczeń czasowych,
- nie wiadomo, czy zastosowanie innych metod do uzyskiwania rozwiązania początkowego nie dałoby lepszych finalnych efektów,
- wybór metod porównywanych z metodą własną nie został właściwie uzasadniony,
- rozpatruje się grupę zadań definiowanych grafem i nie odniesiono się do tego, czy metoda tu opracowana nadawałaby się, po pewnych modyfikacjach, do rozwiązywania problemów ze strumieniami zadań napływających do chmury obliczeniowej.

Powyższe mankamenty nie pomniejszają bardzo dobrego wrażenia, jakie wywołała na mnie praca. Jest napisana z dużym polotem, zawiera wszystkie istotne dla tematu zagadnienia. Opracowanie metody wyznaczania strategii szeregowania zadań w chmurze obliczeniowej daje w efekcie pracę napisaną na dość wysokim poziomie merytorycznym.

7. Jaka jest przydatność rozprawy dla nauk technicznych?

Rozprawa zawiera szereg oryginalnych dokonań Autora i dobrze opracowaną metodę rozwiązywania problemu przydziału zadań do serwerów w chmurze obliczeniowej. Uzyskiwanie tego typu rozwiązań jest ważne w praktyce projektowania systemów obliczeniowych czasu rzeczywistego i potrafi dać znakomite efekty w produkcie finalnym, zatem nie ulega wątpliwości, że rozprawa jest przydatna dla nauk technicznych, w szczególności dla informatyki.

Warto przy tym pokazać szerszy aspekt stosowania prezentowanego tu podejścia. Sądzę, że w wielu przypadkach podejście tego typu jest charakterystyczne w innych zagadnieniach charakterystycznych dla nauk technicznych, w których można zastosować taki właśnie sposób postępowania. Ewolucyjna metoda poprawiania, suboptymalnego w sensie czasu realizacji grupy zadań, rozwiązania początkowego, dająca w efekcie dodatkowe zalety (własności) poszukiwanego rozwiązania, jest przecież typowa dla rozważań o charakterze badawczym. Wyniki rozprawy mają więc, moim zdaniem, szerszy aspekt w sensie ich przydatności dla nauk technicznych, niż wynika to ze sformułowania zadania badawczego.

Zaliczam rozprawę do kategorii: spełniającej wymagania

Konkluzja.

Biorąc pod uwagę postanowienia Ustawy o tytule naukowym i stopniach naukowych oraz o stopniach i tytule w zakresie sztuki uważam, że rozprawa mgr. inż. Sławomira Bąka spełnia wymagania stawiane rozprawom doktorskim w dziedzinie nauk technicznych w dyscyplinie informatyka techniczna i telekomunikacja i wnioskuje o jej dopuszczenie do publicznej obrony.

Nowicki

Katedra Architektury Systemów Komputerowych
Wydział Elektroniki, Telekomunikacji i Informatyki
Politechnika Gdańska

**RECENZJA ROZPRAWY DOKTORSKIEJ DLA RADY
WYDZIAŁU ELEKTRONIKI I TECHNIK INFORMACYJNYCH
POLITECHNIKI WARSZAWSKIEJ**

Tytuł rozprawy: Synteza rozproszonych aplikacji czasu rzeczywistego przetwarzanych w chmurze obliczeniowej

Autor rozprawy: mgr inż. Sławomir Bąk

- 1. Jakie zagadnienie naukowe jest rozpatrzone w pracy /teza rozprawy/ i czy zostało ono dostatecznie jasno sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?**

Teza rozprawy określa, że wykorzystanie algorytmu, który jest wyspecjalizowany do syntezy rozproszonych aplikacji czasu rzeczywistego i wykorzystany w chmurze obliczeniowej, pozwala na znaczne obniżenie kosztów wykorzystania zasobów oraz lepsze wykorzystanie zasobów w zestawieniu z aktualnie stosowanymi rozwiązaniami.

Autor wyjaśnił szczegóły rozpatrywanego zagadnienia definiując precyzyjnie elementy i rozproszoną aplikację ACCR w sekcjach 5.1 oraz 5.2 jak również architekturę SCCR z elementami jak kanały komunikacyjne, serwer, chmura IaaS, czas wykonania przez zasób, koszt wykorzystania architektury w sekcji 5.3, kryteria optymalizacji w sekcji 5.4, chmurę czasu rzeczywistego i kryterium optymalizacji w takiej chmurze w sekcji 5.5. Szczegóły modelu pozwalają na określenie, że zagadnienie naukowe zostało precyzyjnie i jasno opisane przez autora.

Praca ma charakter teoretyczno-eksperymentalny ze względu na to, że autor zarówno określił złożoność obliczeniową zaproponowanego algorytmu wykorzystując znaną notację $O()$ jak i przeprowadził szereg eksperymentów zakładając konkretne modele środowisk oraz aplikacji z zestawami parametrów je opisujących w celu porównania zaproponowanego algorytmu z algorytmem referencyjnym, bazującym na podejściu znanym z literatury.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł / w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle /świadczący dostatecznej wiedzy autora. Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

Analiza źródeł przedstawiona została w rozdziale 2 rozprawy. Ponadto, autor odwołał się także do istniejącej literatury w rozdziale 1 we wprowadzeniu do zagadnień przetwarzania w chmurze oraz zastosowań chmury czasu rzeczywistego. Mgr Bąk wyróżnił różne kategorie systemów rozproszonych ze względu na różnorodność procesorów oraz zależność czasu wykonania zadań od parametrów. Autor zidentyfikował różne ogólne metody szeregowania dla zadań lub grup zadań pojawiających się okresowo, ogólnie dla systemów rozproszonych. W dalszej kolejności przeanalizowane zostały podejścia i algorytmy optymalizacji w systemach chmurowych. W szczególności: podejścia minimalizujące koszt, maksymalizujące wydajność, ale także poszukujące kompromisu pomiędzy wydajnością a kosztem wykorzystania infrastruktury. Autor rozpatruje algorytmy uwzględniające wymagania czasowe, alokację przepustowości łącza. Mgr Bąk przywołuje także sformułowania problemów uwzględniające minimalizację zużycia energii, także w kontekście chmury mobilnej. W szczególności prezentuje szereg algorytmów szeregowania uwzględniających koszt i kryteria czasowe. Następnie przedstawione są rozszerzenia problemu uwzględniające wykonanie usług w czasie rzeczywistym, bądź zbliżonym do rzeczywistego, wciąż w środowisku chmurowym. Przedstawione zostały różne typy algorytmów, przywołane jako rafinacyjne, probabilistyczne, symulowanego wyżarzania. Uważam, że mgr Bąk posiada odpowiednią wiedzę w rozpatrywanej tematyce. Literatura obejmuje 125 pozycji. Dla kompletności, sugerowałbym także odniesienie się do najnowszej literatury dotyczącej terminu „cloud workflow scheduling”, również odnoszącej się do optymalizacji z ograniczeniami czasowymi. Wnioskiem z analizy, wystarczającej z punktu widzenia problemu pracy, jest motywacja dla podjęcia optymalizacji wykorzystania zasobów w chmurach IaaS dla aplikacji z ograniczeniami czasowymi.

3. Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

Przyjęte założenia co do modelu aplikacji nie budzą zastrzeżeń, autor określił jasno definicję zadania, grafu zadań, specyfikację rozproszonej aplikacji ACCR w formie algorytmu echo. Jak podkreślono niżej, można by ściślej określić jakie klasy aplikacji można zdefiniować za pomocą podanej specyfikacji. Od strony sprzętowej, model architektury SCCR obejmuje warstwy sieciową i serwerową. Wraz z parametrami takimi jak moc obliczeniowa węzła, przepustowość kanału komunikacyjnego oraz koszt wykorzystania mocy obliczeniowej węzła wystarczająco modelują chmurę IaaS. Ponadto, autor zdefiniował ograniczenia czasowe dla regularnie uruchamianych ACCR oraz kryterium optymalizacji uwzględniające minimalizację kosztu oraz maksymalizację jakości. Metoda nie budzi zastrzeżeń, autor zaproponował własny algorytm optymalizacji, wykazał jego istotne z punktu widzenia cechy (opisane poniżej w punkcie 4) oraz wykazał przewagę zaproponowanego algorytmu nad algorytmem referencyjnym dla aplikacji o losowych parametrach. Uważam, że można to uznać za dowód tezy rozprawy. Z drugiej strony, można by oczekiwać bardziej precyzyjnego określenia zestawów parametrów aplikacji będących przedmiotem porównania w kontekście aplikacji rzeczywistych (uwaga w punkcie 6). Bezpośredni dowód tezy rozprawy (dla przyjętych założeń co do parametrów aplikacji i systemu) znaleźć można w tabeli 7-1 gdzie odczytać można niższy średni koszt dla algorytmu RMO^{ACCR} 0,98 \$/h względem 1,47 \$/h dla referencyjnego $MSMD^{ACCR}$ oraz lepsze średnie wykorzystanie zasobów odpowiednio

75,81% względem 57,31% odpowiednio dla przywołanych algorytmów.

4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Oryginalność rozprawy polega na propozycji algorytmu rafinacyjnego RMO^{ACCR} , który służy do rozwiązywania zdefiniowanego problemu jak również wykazanie, na bazie eksperymentów symulacyjnych, wyższości tego algorytmu w porównaniu do algorytmu $MSMD^{ACCR}$, dla przykładowych, losowych aplikacji. Ponadto, autor wykazał stosowalność oraz znaczące zmniejszenie kosztu wykonania dla aplikacji adaptacyjnego systemu nawigacji samochodowej oraz aplikacji zliczania liczby wystąpień wyrazów w tekście. Zwrócić należy uwagę, iż mgr Bąk zademonstrował, że zaproponowany algorytm jest w stanie wyjść przynajmniej z niektórych minimów lokalnych stosując zaproponowany zestaw modyfikacji. Pokazano to na rys. 7-2 oraz 7.1-2. Ten pierwszy obrazuje dłuższą zbieżność do rozwiązania tej samej jakości w porównaniu do $MSMD^{ACCR}$, ale także możliwość finalnego osiągnięcia rozwiązania o mniejszym koszcie.

5. Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników /zwięzłość, jasność, poprawność redakcyjna rozprawy/?

Autor, po uprzednim omówieniu szczegółów algorytmu tj. algorytmu rafinacyjnego, w tym kroków oraz dozwolonych modyfikacji M1-M4 (w rozdziale 6), przedstawił dowód możliwości uzyskania rozwiązania o dowolnej architekturze poprzez wykonanie zaproponowanych modyfikacji. Zaprezentował także miarę jakości rozwiązania i zademonstrował możliwość wyjścia algorytmu z lokalnych minimów. Algorytm przedstawiony został w klarowny i zwięzły sposób w postaci pseudokodu na rys. 6-5-1, z wyszczególnieniem możliwych modyfikacji w ramach kroku rafinacji przedstawionych na rys. 6.5-2. Co ważne, autor zadbał o uzasadnienie zbieżności algorytmu a także określenie złożoności obliczeniowej algorytmu syntezy SCCR dla najgorszego przypadku. Elementy te dowodzą umiejętności określenia podstawowych i istotnych cech algorytmu z punktu widzenia jego zastosowania w praktyce, z uwagami wyszczególnionymi poniżej.

Z kolei wyniki liczbowe z testów algorytmu zostały przedstawione w dwóch częściach: porównania proponowanego algorytmu do referencyjnego $MSMD^{ACCR}$ dla konfiguracji o losowych parametrach a następnie przedstawienie skuteczności optymalizacji poprzez minimalizację kosztu dla dwóch aplikacji o praktycznym charakterze – systemu nawigacji samochodowej dla obszaru miasta, które dzielone jest na sektory oraz zliczania liczby wystąpień wyrazów w tekście. Tutaj podejście jest wystarczające, w moim przekonaniu, chociaż porównanie praktycznych aplikacji mogłoby uwzględniać także inne algorytmy optymalizacji, w szczególności różnych grup, wymienionych przez autora w analizie źródeł.

6. Jakie są słabe strony rozprawy i jej główne wady?

Uważam, że omówienie literatury związanej z przedmiotem rozprawy powinno wyróżniać

podejścia ze względu na model aplikacji, której lub których wykonanie podlega optymalizacji (zestaw zadań, częstotliwość pojawiania się, definicja aplikacji w formie scenariusza zadań ang. workflow application etc.) oraz konkretne cele optymalizacji uwzględniające różne metryki takie jak czas wykonania, zużyta energia, optymalizacja wielokryterialna etc. W aktualnej formie różne kwestie np. energetyczne opisywane są w różnych fragmentach bardzo długiego opisu w sekcji 2.1.

Czy autor wykonywał eksperymenty odnośnie różnych rozmiarów drzewa pokazanego na rys. 7.2.2-2, pod kątem różnej wydajności przetwarzania? Mogłoby to być bardzo ciekawe z punktu widzenia optymalizacji wykonania.

Dlaczego algorytm MSMD^{ACC}R jest uznany za referencyjny? Na s. 33 oraz 89 autor stwierdza, że jest najlepszym kandydatem. Czy autor mógłby poprzeć stwierdzenie bardziej konkretnym uzasadnieniem?

Czy autor przeanalizował średnią złożoność algorytmu?

Wartościowa byłaby dyskusja odnośnie możliwości wyrażenia różnych klas aplikacji w postaci rozważanego modelu.

Autor przyznaje we wstępie do sekcji 7.2, że losowe przykłady (takie rozpatrywane są na początku rozdziału 7) nie muszą być reprezentatywne. Na ile więc praktyczne są wyniki uzyskane w ramach tego fragmentu pracy? Autor mógłby pokusić się o próbę określenia parametrów klas rzeczywistych obciążeń w chmurze. Niektóre projekty podejmowały optymalizację związaną z wykorzystaniem obciążeń rozpatrywanych w chmurze np. w projekcie KRICO <http://krico.gda.pl/>. Być może szczegóły dotyczące samych obciążeń mogłyby zostać wykorzystane przez autora.

Czy autor zidentyfikował słabe strony proponowanego algorytmu, w których istniejące podejścia mogłyby być lepsze? Takim przypadkiem wydaje się być konfiguracja w pierwszej linijce tabeli 7-1.

Uwagi szczegółowe:

- s. 29 – w stosunku do czego określono redukcję kosztu rzędu 50%?
- s. 46 - autor zakłada, że jeśli węzeł ma większą moc obliczeniową to koszt wynajęcia jest wyższy – sformułowanie jest mało precyzyjne, przydałby się opis bardziej szczegółowy oraz referencje do typowych zależności i modeli wydajnościowo-kosztowych używanych w praktyce w systemach wysokiej wydajności oraz chmurowych
- definicja 5.3.5 – w ogólności czas transmisji mógłby także uwzględniać opóźnienie komunikacyjne tzw. startup time
- s. 86 i 87 – diagramy powinny mieć oznaczone jednostki na osiach OY

Uwagi językowe:

Praca napisana jest dobrym językiem, nie znalazłem wielu uchybień w tym względzie. Tym niemniej, następujące fragmenty mogłyby być poprawione:

- s. 33 „Posiada jednak pewne różnice ...”

- s. 101 powinno być „mapowaniem i redukcją”
- s. 101 lepiej napisać „liczba zadań” zamiast „ilość zadań”

7. Przydatność rozprawy dla nauk technicznych?

Bez wątpienia tematyka podejmowana przez rozprawę, tj. optymalizacja wykonania aplikacji i wykorzystania zasobów w chmurze, stanowi aktualnie ważny temat badawczy, istotny zarówno z punktu widzenia dostawców komercyjnych, centrów superkomputerowych i oczywiście klientów, zarówno instytucjonalnych jak i prywatnych. Wynika to z rosnącej popularności wykorzystania technologii chmurowych, w szczególności dla obciążeń służących bardzo dużej liczbie użytkowników i powiązanych z nimi usług przy ograniczeniach na czas przetwarzania żądań. Można oczekiwać zwiększenia popularności tego typu obciążeń wraz z rozwojem usług dla użytkowników systemów mobilnych oraz wszechobecnych. Praca wnosi propozycję algorytmu do optymalizacji wykonania tego typu aplikacji w systemach chmurowych porównując wyniki do innego znanego algorytmu, z analizą złożoności obliczeniowej rozwiązania. Powyższe można także potraktować jako przyczynek dla dalszych badań w tym obszarze dyscypliny.

8. Do której z następujących kategorii Recenzent zalicza rozprawę:

- a/ nie spełniająca wymagań stawianych rozprawom doktorskim przez obowiązujące przepisy**
- b/ wymagająca wprowadzenia poprawek i ponownego recenzowania**
- c/ spełniająca wymagania**
- d/ spełniająca wymagania z wyraźnym nadmiarem**
- e/ wybitnie dobra, zasługująca na wyróżnienie**

Podsumowując, w moim przekonaniu, rozprawa spełnia wymagania postawione rozprawom doktorskim przez odpowiednią ustawę, w dziedzinie nauk technicznych, w dyscyplinie informatyka i wnioskuję o dopuszczenie rozprawy do publicznej obrony.



