



Kraków 30.11.2018

dr hab. inż. Aleksander Byrski, prof. AGH
Grupa Inteligentnych Systemów Informacyjnych
Katedra Informatyki, Wydział Informatyki, Elektroniki i Telekomunikacji
Akademia Górniczo-Hutnicza im. St. Staszica w Krakowie
Al. Mickiewicza 30, 30-059 Kraków
e-mail: olekb@agh.edu.pl
Tel: 012 328 33 00
Fax: 012 617 51 72

Recenzja rozprawy doktorskiej pt.:

Inteligentne systemy monitoringu procesów harmonogramowania w rozproszonych środowiskach dużej skali w ujęciu wieloagentowym

autorstwa:

mgr inż. Daniel Grzonka

Systemy współbieżne i rozproszone stały się aktualnie „wszędobylskie” (z ang. przewyższa mocą obliczeniową urządzenia które obciążały nasze biurka jeszcze kilka/kilkanaście lat temu, a po drugie jest maszyną wielordzeniową na której realnie można uruchamiać obliczenia wielowątkowe i wieloprocessowe (pamiętając o ograniczeniach związanych z pojemnością baterii). Zrównoleglenie obliczeń wydaje się już od wielu lat jedynym sensownym kierunkiem rozwoju infrastruktury, ze względu na zbliżanie się projektantów i producentów sprzętu komputerowego do granicy prawa Moore'a, nawet jeśli nie pod względem fizycznym to z pewnością ekonomicznym.

W skali bardziej użytecznej niezmiernie popularnymi w dzisiejszych czasach są klastry i gridy obliczeniowe a także systemy typu cloud computing. Przykładowo, dla polskich uczonych uzyskanie dostępu do takich zasobów jak superkomputery dostarczane w ramach projektu PIGrid stanowi jedynie formalność, poza tym wszyscy używamy systemów chmurowych w codziennej pracy czy rozrywce (są to serwisy streamingowe, pakiety biurowe, systemy przechowywania danych i inne).

Przedstawiona do recenzji rozprawa doskonale wpisuje się w aktualne potrzeby obszaru Informatyki związanego z obliczeniami współbieżnymi i rozproszonymi, a proponowane rozwiązania stanowią wyraźną kontrybucję do stanu wiedzy związanego z bardzo ważnym aspektem działania systemów typu cloud, choć na co dzień zupełnie (z założenia) niewidocznym dla użytkownika – czyli z harmonogramowaniem zadań które w sposób

zaplanowany mają być rozdzielone pomiędzy maszynami liczącymi. Na podkreślenie zasługuje fakt transparentności dla użytkownika tego typu usługi – jedynie objawiającej się w kategoriach Quality of Experience.

Sam problem harmonogramowania należy do klasy problemów NP-zupełnych, dlatego też niezmiernie ważnym i aktualnym zadaniem dla badaczy jest opracowywanie algorytmów metaheurystycznych, mogących podać pewne rozwiązanie suboptymalne tego problemu (jako że oczywiście uzyskanie rozwiązania w sposób deterministyczny nie jest możliwe w akceptowalnym czasie).

Doktorant w sposób nowatorski wpisuje się w badania nad problemem harmonogramowania, proponując **agentowy system (w zasadzie agentowe rozszerzenie algorytmu ewolucyjnego) mający na celu rozwiązanie tegoż problemu**, przedstawiając jednocześnie **szereg badań eksperymentalnych wraz z dyskusją wyników ilustrujących przewagę proponowanego algorytmu agentowego nad algorytmem nieagentowym, co jest Jego najważniejszym osiągnięciem** i zgodnie z moją najlepszą wiedzą stanowi istotny wkład do stanu wiedzy w dziedzinie Informatyki (w szczególności: algorytmika, sztuczna inteligencja, systemy agentowe). Jako dodatkowe, ważne osiągnięcia Doktoranta jawią się następujące:

- Doktorant rozszerzył taksonomię systemów monitoringu (K. Fatema i in.) o dodatkowy proces związany z harmonogramowaniem zadań.
- Doktorant przystosował macierz ETC stosowaną w problemach harmonogramowania do realiów systemów rozproszonych, rozszerzając jej elementy o stały i równoległy czas obliczenia, wynikający ze znanego wzoru Amdahla na maksymalne możliwe przyspieszenie zadania obliczeniowego w trakcie jego zrównoleglenia.
- Doktorant zaproponował oryginalną reprezentację problemu harmonogramowania na potrzeby zastosowanego algorytmu ewolucyjnego.
- Doktorant opracował dedykowany system symulacyjny i przeprowadził za jego pomocą szeroko zakrojone eksperymenty, które dostarczyły argumentów potwierdzających założoną przez Niego tezę.

Manuskrypt pracy liczy 175 stron. Struktura pracy mogłaby zostać opracowana w sposób trochę bardziej zbalansowany (znaleźć w niej można kilku i kilkunasto-stronicowe rozdziały, a także rozdziały o długości dwudziestu i trzydziestu stron). Po zarysowaniu celu badań we wprowadzeniu następuje 8 rozdziałów (opisujących: opis problemu, monitoring systemów rozproszonych, systemy wieloagentowe, harmonogramowanie zadań, model środowiska rozproszonego i systemu monitorującego, wyniki badań i podsumowanie), następnie zamieszczono bibliografię liczącą 129 pozycji, spis tabel oraz spis rysunków.

Rozdział 1. przedstawia motywację, cel i tezę pracy. Doktorant podkreśla ważność harmonogramowania jako jednej z podstawowych usług, która powinna być dostępna w systemach rozproszonych, zapowiada również opracowanie agentowego systemu harmonogramowania, który jak można wnioskować z tezy, przyczyni się do generacji lepszych algorytmów w krótszym czasie (w uwagach dyskusyjnych odnoszę się do pewnego braku precyzji tego stwierdzenia).

Rozdział 2. zawiera opis problemu monitorowania. Doktorant przytacza zarys konstruowanego systemu monitoringu, podaje role i akcje poszczególnych agentów, wreszcie przedstawia zawartość poszczególnych rozdziałów pracy (w moim przekonaniu, ten rozdział powinien być zintegrowany z rozdziałem pierwszym).

Rozdział 3. koncentruje się na systemach monitoringu przedstawiając przegląd popularnych rozwiązań oraz ich cech charakterystycznych. Doktorant zaczyna od popularnego środowiska AWS i prezentuje staranny i dogłębny przegląd rozwiązań znanych z literatury przedmiotu. Następnie prezentowane i dokładnie omówione zostają istniejące taksonomie systemów monitoringu. Rozdział trzeci kończy się prezentacją notacji za pomocą której można w łatwy i przejrzysty sposób opisać modele monitoringu.

Rozdział 4. zawiera podstawowe informacje z dziedziny systemów agentowych. Doktorant rozpoczyna od dyskusji definicji i cech agentów, przytaczając spójny przegląd stanu wiedzy w tym temacie. Następnie Doktorant przytacza elementy modelu formalnego opracowanego przez Wooldridge'a, pozwalające się zorientować w zarysie struktury i akcji systemu agentowego opisanego w sposób abstrakcyjny. Kolejnym modelem, którego elementy zostają przytoczone przez Doktoranta jest M-agent opracowany przez prof. Krzysztofa Cetnarowicza. Następnie Doktorant przytacza elementy kilku innych modeli opracowanych przez pracowników Katedry Informatyki AGH. Rozdział kończy się przeglądem podstawowych cech systemów wieloagentowych.

Rozdział 5. koncentruje się na problemie, którego rozwiązanie staje się głównym celem pracy Doktoranta, czyli problemie harmonogramowania. Po przedstawieniu wstępnych definicji i cech, Doktorant przytacza notację opisu problemu harmonogramowania opartą na trzelementowej krotce, następnie przedstawia szereg możliwych do przyjęcia kryteriów oceny jakości rozwiązania problemu harmonogramowania (niezbędnymi do konstrukcji metaheurystyk). W kolejnych podrozdziałach Doktorant osadza problem harmonogramowania w realiach rozproszonych środowisk obliczeniowych, m.in. przedstawiając ważny element składowy swoich algorytmów a mianowicie macierz ETC wraz z proponowanym rozszerzeniem bazującym na prawie Amdahla. W dalszym ciągu doktorant prezentuje algorytm ewolucyjny jako bazową metaheurystykę, którą zajął się w pracy.

W rozdziale 6. Doktorant przechodzi do szczegółowej dyskusji elementów proponowanego rozwiązania, czyli agentowego systemu obliczeniowego. Rozdział rozpoczyna przedstawienie reprezentacji populacji (jest to jeden z oryginalnych wyników badań Doktoranta przedstawiony w niniejszej rozprawie), następnie zaprezentowane zostają szczegóły opisu poszczególnych elementów dynamiki środowiska rozproszonego w którym ma być realizowane harmonogramowanie, oraz procesu tworzenia harmonogramu. Zaraz potem następuje szczegółowy opis komponentów stosowanej metaheurystyki, a więc algorytmu ewolucyjnego, rozpoczynając od selekcji, krzyżowania i mutacji. Na uwagę zasługuje fakt eleganckiej i zwięzłej prezentacji elementów algorytmu z wykorzystaniem funkcji charakterystycznej zbioru. W dalszej części rozdziału Doktorant przechodzi do podania szczegółów aspektów agentowych proponowanej metaheurystyki. W szczególności podaje sposób przydziału zadań do jednostek obliczeniowych oraz opisuje ewolucyjną metodę konstruowania harmonogramów dla poszczególnych pakietów zadań. W sekcji 6.4. Doktorant przedstawia model formalny systemu agentowego, a następnie przechodzi do szczegółowego opisu działania



poszczególnych, zaproponowanych w pracy agentów:

- Ag0 monitoruje proces generowania harmonogramu i ingeruje w selekcję (jest to jedyny agent tego typu w systemie),
- Ag1 informuje o aktualnym etapie wykonania zadań i gotowości jednostek obliczeniowych do przyjmowania nowych,
- Ag2 bazując na informacjach otrzymanych od Ag1 wskazuje moment inicjalizacji procesu harmonogramowania zadań (to jedyny tego typu agent w systemie),
- Ag3 wykorzystuje sztuczną sieć neuronową do monitorowania terminowości wykonywania poszczególnych zadań przypisanych do jednostki obliczeniowej i predykuje opóźnienia,
- Ag4 konsultuje oczekiwane czasy wykonywania zadań z agentami typu Ag3 i bazując na predykcji opóźnień wykonywanej przez nich tworzy macierz poprawek (to jedyny tego typu agent w systemie).

W rozdziale 7 Doktorant prezentuje szeroko zakrojone badania eksperymentalne, przy czym wyniki zostały otrzymane z użyciem dedykowanego środowiska testowego opracowanego przez Doktoranta. Wychodząc od przedstawienia sposobu generacji danych testowych oraz założeń środowiska testowego doktorant przechodzi przez wszystkie zaproponowane typy agentów i porównuje skuteczność działania nieagentowego algorytmu harmonogramowania z wersjami agentowymi. W wielu zaobserwowanych przypadkach należy stwierdzić, że wprowadzenie mechanizmów agentowych wpłynęło pozytywnie na efektywność i skuteczność działania systemu. Rozdział kończy się rozważaniami n.t. możliwości zastosowania proponowanego systemu harmonogramowania oraz krótkim przedstawieniem opracowanego środowiska testowego.

W rozdziale 8 znaleźć można podsumowanie pracy oraz najważniejsze wnioski wyciągnięte przez Doktoranta w trakcie jej realizacji.

Doktorant przygotowując pracę nie uniknął pewnych błędów i niedociągnięć, które chciałbym teraz przedstawić. Poza tym czytając pracę nasunęło mi się kilka uwag ogólnych, o odniesienie się do których uprzejmię proszę Doktoranta.

Uwagi o charakterze ogólnym i dyskusyjnym:

- Na wstępie (strona 3) pojawiają się nieprecyzyjne sformułowania, np. „bardziej realistyczny model zadań i jednostek obliczeniowych”, a szczególnie w tezie „bardziej optymalnych harmonogramów w krótszym czasie” - słowo „bardziej” sugeruje porównanie, Doktorant nie podaje z czym się chce porównywać, poza tym nie istnieje takie pojęcie jak „bardziej optymalny”, może być gorszy lub lepszy, a optymalny to po prostu „najlepszy”.
- W rozdziale 3.4 przedstawiony został model monitoringu, sprowadza się on jednak do prezentacji pewnego rodzaju krotki (?) $\alpha|\beta|\gamma$ z podaniem możliwości przypisania wartości poszczególnych symboli (np. α może oznaczać chmurę obliczeniową, przetwarzanie sieciowe, klaster obliczeniowy itd.). Nazywanie tego notacji „modelem” może być postrzegane jako pewne nadużycie. Wprowadzenie spójnych oznaczeń niewątpliwie przyczynia się do uspoźnienia wszelkich dyskusji na temat przedmiotu. W tym momencie można też po raz pierwszy dostrzec usterkę techniczną miejscowo spotykaną

w dalszych częściach pracy: elementy podanego wyrażenia (tu 3.1.) nie są poprawnie wyjaśnione: przykładowo, co oznacza symbol „pipe” - „|”? Oczywiście można się domyślić, że jest to swoisty separator, a prezentowana notacja stanowi tak naprawdę element iloczynu kartezjańskiego zbiorów możliwych wartości alfa, beta i gamma – ale czytelnik nie powinien być pozostawiony sam na sam nawet z tak prostymi domysłami. Podobne uwagi można wysunąć do notacji przytaczanej w rozdziale 5.2 (choć nie została ona zaproponowana przez Doktoranta).

- Nie jest do końca jasne w jakim celu Doktorant przytacza dwa modele agentowe, a następnie wybrane elementy jeszcze kilku modeli. Jeden z nich [24] został zastosowany w sekcji 6.4. do formalizacji konstruowanego systemu agentowego – nie wyjaśniono w jakim celu więc, Doktorant przytoczył szczegóły pozostałych modeli.
- W rozdziale 5.1. na stronie 57 podana została klasyfikacja algorytmów za pomocą których rozwiązuje się problemy harmonogramowania. Zauważyć można, że typy problemów rozwiązywanych przez pierwsze 3 algorytmy są identyczne, sposób więc ich podania jest nadmiarowy. Podanie 5 algorytmów powinno zostać opatrzone komentarzem, dlaczego wybrano akurat te 5 (nie wspominając już o cytowaniach odpowiednich prac, które również powinny się tutaj pojawić) – algorytm mrówkowy akurat jest jedną z najczęściej stosowanych technik w rozwiązywaniu problemów harmonogramowania, natomiast algorytm PSO najczęściej stosowany jest do rozwiązywania problemów ciągłych, a nie dyskretnych. Poza tym trudne problemy można próbować rozwiązywać w zasadzie za pomocą dowolnej metaheurystyki (oczywiście efekty będą zależały od jej wyboru i parametryzacji) dlatego tym bardziej brakuje w tym miejscu stosownego komentarza odnośnie intencji Doktoranta.
- Doktorant za punkt wyjścia do rozszerzenia tablicy ETC przyjmuje prawo Amdahla, które jest znane z ograniczeń (w szczególności przyjmowania stałego rozmiaru problemu niezależnie od dostępnej infrastruktury). Planista może adaptować rozmiar zadania do konkretnej mocy obliczeniowej dostępnej w sieci – wówczas przyspieszenie jest w sposób lepszy opisywane przez prawo Gustafsona, warto byłoby się pokusić o przedyskutowanie możliwości rozszerzenia macierzy ETC w oparciu o to prawo.
- Algorytm ewolucyjny przedstawiony w punkcie 5.5. jest na tyle ogólny, że łączenie go bezpośrednio z harmonogramowaniem nie stanowi specjalnej wartości dodanej w dyskusji. Lepiej było przedstawić algorytm w formie ogólnej, następnie dyskutując jego elementy niezbędne do zaimplementowania problemu harmonogramowania (funkcję przystosowania oraz operatory wariacyjne).
- Komponenty algorytmu ewolucyjnego opisywane w rozdziale 6.3. powinny być poprzedzone rozważaniami oderwanymi od samego problemu harmonogramowania. Doktorant stosuje krzyżowanie jednopunktowe, mutację polegającą na zamianie kolejności zadań oraz przystosowany do swoich potrzeb rodzaj selekcji rankingowej – wcześniej w pracy powinien się znaleźć abstrakcyjny opis algorytmu ewolucyjnego oraz często stosowanych wariantów jego komponentów.
- Doktorant w szeregu równań w rozdziale 6 stosuje w sposób bardzo elegancki w celu uproszczenia zapisu skomplikowanych formuł, funkcję charakterystyczną zbioru niestety bez dokładnego określenia tegoż zbioru. Można się domyślać, że funkcja określona jest na

zbiornicze {prawda}, gdyż powinna ona zwracać jeden w przypadku, gdy formuła w nawiasie klamrowym jest prawdziwa i zero w przeciwnym wypadku. Wniosek ten nasuwa się po przejrzaniu formuł w których argumentem funkcji charakterystycznej są nierówności – natomiast pierwsza definicja dotyczy porównania (poniżej równania 6.1). Oczywiście intencja Doktoranta jest jasna, natomiast sama funkcja powinna zostać w sposób precyzyjny zdefiniowana.

- Model agentowy przedstawiony w 6.4. w zasadzie precyzyjnie definiuje strukturę i działanie konstruowanego systemu agentowego, choć pewne jego szczegóły wymagałyby dalszego wyjaśnienia, choćby wiedza agenta k należące do K – czym tak naprawdę jest? Jak może być sformalizowana ewentualnie zaimplementowana? Niedosyt wywołuje jednak fakt, że notacja która wynika z modelu nie jest później wykorzystywana w definicji akcji poszczególnych agentów (w przeciwieństwie do [24], na której to pracy Doktorant najprawdopodobniej się wzorował).
- Doktorant proponuje system obliczeniowy z założenia mający działać na architekturze równoległej i rozproszonej, jednocześnie w pracy nie pojawiają się informacje n.t. metod implementacji i synchronizacji procesów harmonogramowania i agentów. Jeśli to ma być system agentowy, być może sam proces harmonogramowania też powinien być traktowany jako agent? Idąc dalej, niektóre agenty porozumiewają się ze sobą, w jaki sposób to się odbywa? Nie chodzi wyłącznie o szczegóły implementacyjne, ale o mechanizmy algorytmiczne które dbają o synchronizację dostępu do sekcji krytycznych – w końcu np. Ag_0 wpływa bezpośrednio na selekcję wykonywaną gdzieś w systemie (w jakimś procesie?). Tego typu wymiana informacji na pewno może prowadzić do zmiany stanu poszczególnych agentów, procesów...
- Wprowadzanie w systemie rozproszonym pojedynczych bytów typu „generator harmonogramów”, pojedyncze agenty wprowadza doń „single points of failure”- warto byłoby przedyskutować możliwości zapewnienia odporności takiego „miejscami scentralizowanego” systemu (własność „resiliency”).
- Wprowadzenie Ag_0 jako modyfikatora selekcji zostaje podsumowane przez Doktoranta podaniem algorytmu 3 (str 91), który jest typowym algorytmem sekwencyjnym, nie posiada widocznych cech agentowości. Skoro tak można przedstawić harmonogramowanie w systemie po wprowadzeniu Ag_0 to być może akcje tegoż agenta powinny zostać po prostu włączone w algorytm harmonogramowania jako opcjonalny komponent? W takim kontekście tworzenie Ag_0 (oraz co za tym idzie, agentów Ag_2 i Ag_4) może mieć cechy nadmiarowości – warto byłoby precyzyjnie uzasadnić ich wprowadzenie jako agentów (przedyskutować spełnianie przez nich przynajmniej niektórych cech agentowości przytaczanych wcześniej w pracy).
- Doktorant prezentuje swoje eksperymenty bazując na dedykowanym środowisku testowym opracowanym, jak można domyślać się z opisu, jako aplikacja jednowątkowa. Na ile implementacja wpłynęła na zachowanie cech agentowości (przede wszystkim autonomii) przez zaimplementowane agenty? Oczywiście implementacja sekwencyjna algorytmów agentowych jest jak najbardziej możliwa (choćby wykorzystując symulację zdarzeniową), ale wymagałaby ona krótkiego wyjaśnienia i doprecyzowania.
- Doktorant prezentuje wyniki przyjmując bezpieczny rozmiar próby statystycznej (każdy



eksperyment został przeprowadzony 100 razy). Naturalnie bazując na przedstawionych wykresach „box and whiskers” w miarę wyraźnie można zaobserwować które konfiguracje mogą być traktowane jako „lepsze” czy „gorsze” od pozostałych, niemniej jednak warto byłoby wykorzystać jedną z metod testowania hipotez statystycznych do upewnienia się w tym fakcie (choćby testy Kruskal-Wallisa i Dunna).

- Poszczególne agenty zaproponowane przez Doktoranta mają dobrze zdefiniowane odpowiedzialności (niektóre działają w pojedynkę, niektóre wyłącznie kooperując z innymi). Dlatego nic nie stoi na przeszkodzie, aby potencjalnie przeprowadzić testy systemu z wykorzystaniem kilku typów agentów jednocześnie (czyli pewnej kombinacji wszystkich typów agentów). Warto byłoby przynajmniej przedyskutować taką możliwość, z pewnością może ona być atrakcyjna w przyszłych badaniach planowanych przez Doktoranta.
- Pomniejsze, techniczne uwagi
 - Manuskrypt wyglądałby naturalniej i byłby bardziej praktyczny do czytania, jeśli zostałby wydrukowany dwustronnie a interlinia nie przekraczałaby 1,5.
 - Zgodnie z regułami polskiego składu nie powinno się stosować myślnika zamiast pauzy (w zasadzie wg polskiej typografii – półpauzy).
 - Przedstawiane wzory w przeważającej większości opatrzone są dokładnym wyjaśnieniem symboli w nich stosowanych, ale precyzję wywodu zwiększyłoby dokładne przedstawienie przestrzeni do których należą poszczególne zmienne.
 - Niedociągnięciem typograficznym jest rozpoczęcie podsekcji zaraz po nagłówku sekcji bez wstawienia jakiegokolwiek tekstu – np. po sekcji 6.5. natychmiast rozpoczyna się sekcja 6.5.1.
 - Na stronie 109 i wielu innych w dalszej części pracy pozostawione są za duże światła – jest to błąd składu.

Chciałbym stwierdzić, że Doktorant podjął się rozwiązania ciekawego przedsięwzięcia i konsekwentnie realizował założone cele, a za najważniejsze Jego osiągnięcie uważam opracowanie nowatorskiego algorytmu agentowego oraz przeprowadzenie szeroko zakrojonych testów jego efektywności i skuteczności. Wraz z dodatkowymi osiągnięciami, przytoczonymi przeze mnie na początku recenzji, jestem przekonany iż Doktorant dowiódł posiadania wystarczającej wiedzy teoretycznej w dyscyplinie Informatyka. Opracowane przez Doktoranta oryginalne rozwiązania postawionego problemu naukowego wskazują na posiadane przez Niego umiejętności efektywnego prowadzenia badań naukowych.

Jednocześnie stwierdzam, że przytoczone przeze mnie w tekście recenzji uwagi krytyczne i dyskusyjne nie wpływają w znacznym stopniu na poziom prezentowanej pracy i powinny być traktowane w głównej mierze, jako propozycje modyfikacji strategii prowadzenia dalszych badań przez Doktoranta.

W uzupełnieniu do wniosków merytorycznych płynących z pracy chciałbym przedstawić ocenę dorobku naukowego Doktoranta. W trakcie swoich badań naukowych opublikował on 12 artykułów naukowych, w tym 3 indeksowane na liście A MNiSW. Doktorant brał czynny udział w kilku projektach naukowo-badawczych, uczestniczył w 5 konferencjach międzynarodowych, pracował również wielokrotnie jako członek

konferencyjnych komitetów programowych. Brał udział w letnich szkołach i aktywnie prowadził działalność dydaktyczną. Jest czynnym członkiem Rady Wydziału Fizyki, Matematyki i Informatyki Politechniki Krakowskiej oraz członkiem PTI. Przytoczone parametry bibliometryczne zdecydowanie spełniają wymagania stosownej ustawy.

Zgodnie z moim rozeznaniem przedstawiona rozprawa spełnia warunki określone przez Ustawę z dnia 14 marca 2003 (wraz z późniejszymi zmianami) o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz.U. 2003 Nr 65 poz. 595), dlatego też wnoszę o dopuszczenie Pana mgr. inż. Daniela Grzonka do dalszych etapów przewodu doktorskiego.

