

Recenzja wniosku dr Joanny Berlińskiej o nadanie stopnia doktora habilitowanego

Niniejsza opinia sporządzona została w związku z powołaniem mnie w dniu 28/09/2021 na recenzenta w postępowaniu o nadanie stopnia doktora habilitowanego dr Joannie Berlińskiej, toczącym się na Uniwersytecie Adama Mickiewicza w Poznaniu.

OCENA

Po przeanalizowaniu materiałów przedstawionych jako osiągnięcie naukowe mające być podstawą nadania stopnia doktora habilitowanego stwierdzam, iż moim zdaniem

nie jest spełniony warunek

wymieniony w art. 219 ust. 1 pkt 2. Mianowicie, *jakkolwiek Kandydatka posiada w dorobku osiągnięcia naukowe albo artystyczne, stanowiące ~~znaczny~~ wkład w rozwój określonej dyscypliny*, jednak jestem zdania, że określenie wkładu jako *znaczny* jest nieuzasadnione. Przedstawiony cykl publikacji jest niezwykle monotematyczny – rozważany jest jeden problem algorytmiczny z drobnymi modyfikacjami, zaś przedstawione rozwiązania są w dużym stopniu powtarzalne i rutynowe. Jakkolwiek należy wyrazić uznanie dla solidności przedłożonych prac, należy również zauważyć że przedstawiony dorobek nie zawiera ani głębokich wyników, ani innowacyjnych technik w istotny sposób zmieniających stan wiedzy i technik algorytmicznych.

Zgodnie z Ustawą, niniejsza recenzja odnosi się wyłącznie do pytania o spełnienie warunku z art. 219 ust. 1 pkt 2.

SZCZEGÓLOWE OMÓWIENIE PRZEDSTAWIONYCH OSIĄGNIĘĆ BADAWCZYCH

H1. Communication scheduling in data gathering networks with limited memory.

Praca ta jest w moim odczuciu jednym z mocniejszych punktów przedstawionego osiągnięcia naukowego. W rozważanym modelu każdy z sensorów na bieżąco produkuje dane, które mają zostać przesłane do stacji bazowej. Jednocześnie każdy z sensorów ma ograniczony bufor. Przepelnienie tego bufora wskutek braku transmisji jest rozumiane jako niepowodzenie algorytmu.

Problem jest interesujący i ważny. Słabością dokonanej analizy są założenia o rundzie: wszystkie sensory startują z takim samym stopniem zapełnienia bufora, zaś kolejność transmisji jest ustalona (co zmniejsza znacząco obszar poszukiwań). Podana recepta tak więc służy wyłącznie "przeżyciu" i zachowaniu niezmiennika - zapełnienia bufora w tym samym stopniu przez każdy

sensor. Problemem dla zachowania niezmiennika może być brak wystarczającej wielkości danych do transmisji (dotyczy to początkowych sensorów) oraz przepełnienie bufora (dotyczy to ostatnich sensorów). Autorka wyprowadza kilka zależności pozwalających na wyznaczenie wielkości transmisji w rundzie dla zachowania niezmiennika.

Dalsza część pracy to empiryczne badania nad zachowaniem algorytmu w kolejnych rundach (wiadomo, że wartość początkowa dla zapelnienia buforów może zmieniać się z rundy na rundę, co po pewnym czasie może spowodować niepowodzenie).

- H2. Scheduling for data gathering networks with data compression.** Praca zajmuje się problemem wysyłania danych do stacji bazowej w sytuacji, gdy wysyłający ma do wyboru dwie opcje: dane nieskompresowane i dane skompresowane. W drugim przypadku rozmiar maleje o stały czynnik, ale kompresja wymaga czasu proporcjonalnego do długości danych. Stawianym pytaniem jest możliwość wyboru kolejności wysyłania i kompresji poszczególnych danych tak aby całkowity czas transmisji był jak najmniejszy. Formulowany jest też odpowiedni problem decyzyjny.

Głównym wynikiem pracy jest redukcja problemu 3-PARTITION do instancji stawianego pytania. Redukcja jest dosyć bezpośrednia ale i elegancka. W dalszej części Autorka bada eksperymentalnie kilka dosyć naturalnych heurystyk typu zachłannego.

- H3. Scheduling Data Gathering with Maximum Lateness Objective.** W kolejnej odmianie tego samego problemu zbierania informacji chodzi nie o optymalizację łącznego czasu dla wszystkich transmisji, ale o minimalizację łącznego czasu przekroczenia terminu dla wszystkich zadań (*dataset lateness*). Z wcześniejszych prac na ten temat ma wynikać, że problem jest NP-zupełny.

W rozdziale 4 Autorka dokonuje kilku obserwacji. Po pierwsze pokazuje przykład dlaczego optymalne rozwiązanie nie może być uzyskane przez rozwiązanie typu *non-preemptive*. W drugiej propozycji wskazuje na możliwość ograniczenia poszukiwań do szczególnych rozwiązań, gdzie transmisja o wcześniejszym terminie nie może być przerwana przez transmisję o późniejszym terminie. Na tej podstawie konstruuje algorytm heurystyczny DS z jednym parametrem γ . Mianowicie, decyzja o przerwaniu transmisji podejmowana po porównaniu wartości d_j oraz $d_i - \gamma \cdot S$, gdzie d_i oraz d_j są terminami zakończenia transmisji, zaś S czasem potrzebnym do wystartowania transmisji. W sumie eksplorowane są możliwości pomiędzy algorytmami *non-preemptive EDD* a *preemptive EDD*. Wyniki eksperymentalne pokazują, zgodnie z intuicją jaką można mieć, że jakość rezultatów będzie pomiędzy oboma wersjami EDD, zaś γ wpływa na to, do której wersji EDD jest bliżej.

- H4. Scheduling in a Data Gathering Network to Minimize Maximum Lateness.** Praca ta rozważa bardzo podobny model co praca [H3]. W pracy [H4] przyjmuje się, że $S = 0$, natomiast po każdej transmisji stacja bazowa musi wykonać przetwarzanie otrzymanych danych w czasie proporcjonalnym do ich długości. Przetwarzanie może odbywać się równolegle z kolejną transmisją, wsku-

tek czego nowy problem poza uproszczeniem ($S = 0$) narzuca też poważną komplikację.

Autorka przedstawia redukcję problemu 3-PARTITION do postawionego problemu, tym samym wykazując jego NP-trudność. Redukcja jest prawie taka sama jak w pracy [H2].

Druga część pracy to eksperymentalne porównanie heurystyk FIFO, EDD, SRT.

H5. Makespan Minimization in Data Gathering Networks with Dataset Release Times. Kolejny problem zbierania informacji to sytuacja, w której dane do przesłania nie są gotowe do transmisji od razu – dla każdego sensora określony jest moment, w którym dane te są gotowe do transmisji.

W Propozycji 1 Autorka wskazuje, że dowód NP-zupełności z pracy z 1981 roku działa również dla obecnego modelu. Autorka przytacza też kilka innych obserwacji: np. optymalne rozwiązanie w modelu *non-preemptive* może być dłuższe od optymalnego rozwiązania w modelu *preemptive*. Najciekawszą częścią jest modyfikacja algorytmu Johnsona $J(\gamma)$. Autorka pokazuje, że algorytmy te dają współczynnik aproksymacji ≤ 2 (co nie jest do końca zaskakujące, bo największy *release time* jest mniejszy od optymalnego czasu T^* zakończenia pracy algorytmu). Następnie wskazane są konkretne przykłady pozwalające na stwierdzenie, że współczynnik aproksymacji wynosi dokładnie 2. Następnie Autorka bada eksperymentalnie zachowanie algorytmu. Jest to, obok pracy H[1] jedno z ciekawszych miejsc w dorobku.

H6. Heuristics for scheduling data gathering with limited base station memory.

Praca rozważa kolejny wariant problemu zbierania danych, w którym nakłada się ograniczenie na łączny rozmiar pamięci stacji bazowej. Odebranie danych o rozmiarze α_i wymaga wolnej pamięci o rozmiarze α_i . Co więcej, pamięć ta jest zajęta przez okres $A \cdot \alpha_i$ już po otrzymaniu wiadomości ze względu na konieczność przetworzenia tej wiadomości przez stację bazową. Komunikacja i przetwarzanie danych odbywają się w trybie *non-preemptive*, przetwarzanie danych przez stację bazową odbywa się w trybie FIFO. Zadaniem jest zoptymalizowanie kolejności transmisji tak aby łączny czas na odebranie i przetworzenie wszystkich danych był jak najmniejszy.

Pierwszym wynikiem pracy [H6] jest obserwacja, że w ogólności rozważany problem jest NP-trudny, nawet gdy parametry A i C (koszt przetwarzania i koszt transmisji) są równe 1. W tym celu Autorka przedstawia prostą redukcję do problemu BIN PACKING. Dalsza część pracy to ewaluacja prostych algorytmów heurystycznych. Eksperymenty zostały przeprowadzone dla relatywnie małych rozmiarów bufora. Jest to o tyle słuszne, że przy bardzo małym buforze w istocie możemy mieć do czynienia z sytuacją, gdy i tak tylko jedno zadanie mieści się w buforze i problem optymalizacyjny redukuje się do zadania wyjściowego (z podwyższonym czasem transmisji). Z kolei przy dużym buforze ograniczenie stacja bazowa może zawsze mieć odpowiednio dużo pamięci by bez zwłoki startować z kolejną transmisją. Badane przypadki dotyczą wartości α_i dobieranych losowo (zapewne z jednostajnym

prawdopodobieństwem) dla kilku wybranych wartości bufora. Wyniki eksperymentalne rzucają nieco światła na zachowanie się heurystyk dla takich danych.

H7. Scheduling in data gathering networks with background communications.

Praca zajmuje się sytuacją, w której prędkość transferu do stacji bazowej w niektórych okresach zmniejsza o stały czynnik. W odróżnieniu od innych prac, zakłada się, że rozpoczęcie komunikacji nie wiąże się każdorazowo ze stałym opóźnieniem. Dlatego w sytuacji *preemptive scheduling* wskazany algorytm wielomianowy jest oczywisty. Z kolei dla przypadku *non-preemptive scheduling* redukcja do problemu 3-PARTITION jest oczywista i jest niemal powtórzeniem dowodu z jednej z wcześniejszych prac.

Konstrukcja algorytmu opartego na programowaniu dynamicznym jest oczywista, z kolei uproszczenia zawarte w propozycjach 3, 4, 5 tylko w niewielkim stopniu zmieniają złożoność tego algorytmu.

W rozdziale 6 Autorka rozważa kilka prostych heurystyk: *gTime*, *gRate*, *gSlow-time*. Każda z nich posiada intuicyjną motywację, jednak równie łatwe jest wskazanie przypadków, w których heurystyki te działają słabo, lub w których jedna z nich okazuje się dawać wyniki optymalne przy słabych wynikach pozostałych heurystyk. Atrakcyjnie z pozoru brzmią wyniki dotyczące współczynnika aproksymacji. Niemniej jednak chodzi o trywialny fakt z Propozycji 6 – współczynnik aproksymacji nie może być większy niż δ – i wskazanie, że żadna z heurystyk nie poprawia tego współczynnika aproksymacji.

Rozdział 7 zawiera wyniki eksperymentów, jednak trzeba pamiętać, że tak jak zazwyczaj porównanie heurystyk dotyczy tylko "losowego" wyboru parametrów zadania. Zmiana rozkładu prawdopodobieństwa może dać zupełnie inne wyniki. W praktyce należałoby oczekiwać rozkładów bardzo różnych od jednostajnego. W tym momencie samo rozpoznanie sytuacji mogłoby być bardzo ciekawe.

H8. A comparison of priority rules for minimizing the maximum lateness in tree data gathering networks.

Najistotniejsza część pracy dotyczy problemu zbierania informacji nie bezpośrednio przez stację bazową (jak w większości poprzednich prac), ale w sieci będącej drzewem o głębokości 2. Jest to najprostszy przypadek dla zbierania informacji poprzez węzły pośrednie – w praktycznych przypadkach można się spodziewać większej liczby stacji wierzchołków pośrednich. Autorka porównuje trzy standardowe strategie używane przez węzły pośrednie: FIFO, EDD, SRT oraz hybrydową strategię SD.

Badania eksperymentalne mogą być wartościowe dla wyboru algorytmu w sytuacjach specyficznych. Niestety, rozważany model jest niezwykle uproszczony, choćby założenie o takim samym stopniu kompresji dla każdego zbioru danych jest nierealistyczne. Dla zorientowania się w sytuacji często należy zbadać bardzo uproszczony model teoretyczny, jednak w przypadku badań eksperymentalnych takie uproszczenia nie mają sensu. Przyznać należy, że praca obszernie dokumentuje wyniki eksperymentów.

OGÓLNA OCENA ZAPREZENTOWANEGO OSIĄGNIĘCIA NAUKOWEGO

silne strony :

- program badawczy wszechstronnie badający różne aspekty tego samego problemu, poszukujący rozwiązań na granicy, gdzie stawiane zadania algorytmiczne stają się NP-trudne,
- wysoki poziom edycyjny, staranność w dokumentowaniu wyników, ogólna rzetelność,
- samodzielność w pracy badawczej (żadna przedstawiona praca nie ma współautora, dzięki czemu wkład autorki jest bezsprzeczny).

słabe strony :

- Większość wyników ma charakter rutynowy. Choć prace takie są pozytywne i potrzebne, stanowią słabą przesłankę dla nadania stopnia doktora habilitowanego w naukach ścisłych. Nie stanowiłoby to problemu w przypadku początkującego badacza, na przykład ubiegającego się o stopień doktora. Warto zauważyć, że żadna praca nie zawiera twierdzenia, dowodzone fakty są (uczciwie) określane jako *propositions*.
- Prace konstruowane są zazwyczaj według jednolitego schematu: (a) sformułowanie modelu lekko różniącego się od poprzednio rozważanych, (b) obserwacja, że zadanie obliczeniowe jest NP-trudne, (c) budowa kilku heurystyk, (d) eksperymentalna ocena heurystyk w prostym modelu.
- Dobór danych do testowania heurystyk jest skażony bardzo silnymi uproszczeniami. Autorka nie podejmuje pytania w jaki sposób generowane instancje związane są z pytaniami występującymi w rzeczywistości obliczeniowej.
- Przedstawiony dorobek nie charakteryzuje się stopniowym wzrostem poziomu wyników, co byłoby naturalnym oczekiwaniem w stosunku do osoby ubiegającej się o stopień dra habilitowanego.

Jakkolwiek moja rekomendacja jest negatywna, nie jest to ocena o charakterze kategoriycznym. Przedstawiony dorobek nie zawiera żadnych elementów jednoznacznie dyskwalifikujących.

Mirosław Kutylowski

Mirosław Kutylowski