

## **Recenzja rozprawy doktorskiej**

*mgr. inż. Mirosława Ławrynowicza*

### **zatytułowanej:**

*Algorytmy szeregowania zadań z nieustalonymi terminami gotowości  
na maszynach dowolnych*

### **1. Problem badawczy i jego znaczenie**

W rozprawie rozważa się problem szeregowania zadań na maszynach równoległych, gdzie czas wykonania zadania oraz termin jego gotowości zależy od zadania oraz od maszyny, na której jest wykonywane. Rozważa się różne sposoby zdefiniowania terminu gotowości zadania: deterministyczny określony a priori, niedeterministyczny oraz wynikający z lokalizacji maszyny.

Problemy szeregowania zadań stanowią od lat przedmiot badań naukowych ściśle związany z praktycznymi zastosowaniami, szczególnie w zarządzaniu zasobami w systemach komputerowych, produkcyjnych i usługowych. W teorii szeregowania zadań problem szeregowania definiuje się jako poszukiwanie przydziału zasobów do zadań w czasie, którego celem jest minimalizacja (lub maksymalizacja) przyjętego kryterium przy określonych ograniczeniach. Zarówno potrzeby wynikające z praktycznych zastosowań, jak również rosnące możliwości obliczeniowe współczesnych komputerów powodują rozwój coraz bardziej złożonych modeli i metod optymalizacji w ramach teorii szeregowania. Szczególnie istotne staje się uwzględnienie założeń, dzięki którym model matematyczny zagadnienia stanowi coraz dokładniejszy opis rzeczywistej sytuacji decyzyjnej, co prowadzi do poprawy jakości podejmowanych decyzji.

Niniejsza rozprawa dotyczy analizy modeli szeregowania zadań z uwzględnieniem nowych ograniczeń polegających na uzależnieniu terminu dostępności zadania od maszyny, na której jest ono wykonywane. Analizę ograniczono do problemów szeregowania zadań na maszynach równoległych, gdzie zadanie może być wykonane na dowolnej maszynie, przy czym czas wykonania zadania zależy od maszyny (tzw. maszyny dowolne lub niezależne) z kryterium minimalizacji długości uszeregowania. W klasycznej teorii szeregowania termin gotowości definiuje się jako zależny jedynie od zadania. W niniejszej rozprawie rozważa się trzy modele terminów gotowości zależnych od zadań i maszyn. W pierwszym modelu przyjęto, że termin gotowości jest znany a priori, w związku z tym parametry instancji są reprezentowane w postaci macierzy liczb. W drugim modelu przyjęto, że termin gotowości zadania może przyjąć dowolną wartość z określonego przedziału, przy czym rozkład prawdopodobieństwa nie jest znany. Ze względu na niedeterministyczny charakter danych, jako kryterium w tym modelu przyjęto minimalizację maksymalnego odchylenia od decyzji optymalnej dla danego scenariusza (tzw. żalu) tj. kryterium Savage'a. Rozważa się dwa przypadki: terminy gotowości zależne tylko od zadania oraz zależne zarówno od zadania, jak i od maszyny. Trzeci model wynika z sytuacji gdzie mamy do czynienia równocześnie z problemem lokalizacji maszyn oraz szeregowania zadań. Termin gotowości zadania zależy wtedy od odległości zadania od maszyny, na której będzie

WPLYNĘŁO

1

11-10-2023

RDN-IT / 188/2023



wykonywane, przy czym lokalizacje zadań są zdefiniowane a priori. Dla powyższych problemów przedstawiono analizę złożoności obliczeniowej oraz zaproponowano algorytmy dokładne i heurystyczne wraz z oceną ich efektywności na podstawie eksperymentów obliczeniowych. Podejście takie jest poprawne metodologicznie a uzyskane wyniki stanowią wkład w rozwój teorii szeregowania zadań. Przytoczone przykłady ilustrujące zastosowania analizowanych modeli potwierdzają potencjalne znaczenie praktyczne prezentowanych wyników.

## 2. Wkład autora

W rozdziale 1.5. sformułowano podstawowy cel rozprawy jako: „opracowanie i przebadanie algorytmów szeregowania dla problemów  $Rm|r_{ij}|Cmax$ ,  $Rm|r_j^- \leq r_j \leq r_j^+|Reg(Cmax)$ ,  $Rm|r_{ij}^- \leq r_{ij} \leq r_{ij}^+|Reg(Cmax)$  oraz dla problemu ScheLoc z kryterium długości uszeregowania i dyskretnym podproblemem rozmieszczenia maszyn dowolnych.” Cel ten został osiągnięty poprzez realizację kilku celów cząstkowych zrealizowanych w ramach analizy poszczególnych problemów.

W szczególności, analiza deterministycznego problemu  $Rm|r_{ij}|Cmax$  obejmuje wykazanie własności problemu oraz wskazanie szczególnych przypadków rozwiązywalnych w czasie wielomianowym lub pseudowielomianowym, implementację metody wyznaczania wartości dolnego ograniczenia długości uszeregowania oraz opracowanie trzech algorytmów heurystycznych, dedykowanych oraz metaheurystyki symulowanego wyżarzania (Simulated Annealing). Ocenę efektywności zaproponowanych algorytmów przeprowadzono na podstawie eksperymentów obliczeniowych.

W ramach analizy problemów  $Rm|r_j^- \leq r_j \leq r_j^+|Reg(Cmax)$  oraz  $Rm|r_{ij}^- \leq r_{ij} \leq r_{ij}^+|Reg(Cmax)$  wykazano szereg własności wykorzystanych do projektowania algorytmów heurystycznych dla rozważanych problemów. Również w tym przypadku zastosowano zarówno algorytmy dedykowane, jak i metaheurystykę przeszukiwania tabu (Tabu Search). Przeprowadzono eksperymenty obliczeniowe w celu oceny efektywności proponowanych algorytmów.

Dla problemu równoczesnego rozmieszczenia maszyn i szeregowania zadań (ScheLoc) opracowano algorytm genetyczny oraz algorytm hybrydowy składający się z heurystyki zachłannej i metaheurystyki symulowanego wyżarzania. Zaproponowano również podejście polegające na dekompozycji problemu ScheLoc na dwa etapy. W pierwszym etapie wybiera się podzbiór lokalizacji maszyn a w drugim etapie rozwiązuje się problem szeregowania zadań. Analizowano dwa sposoby realizacji etapu pierwszego: jeden, polegający na minimalizacji największej odległości spośród wybranych lokalizacji od zadania, drugi polegający na minimalizacji sumy odległości wybranych lokalizacji do wszystkich zadań. Ponadto przedstawiono koncepcję, według której w pierwszym kroku rozwiązuje się problem szeregowania zadań na maszynach dowolnych z niedeterministycznymi terminami gotowości zadań. Następnie metodą pełnego przeglądu wybiera się lokalizację maszyn, która minimalizuje długość uszeregowania z zachowaniem przydziału zadań do maszyn oraz kolejności wykonywania zadań na każdej maszynie ustalonych w pierwszym kroku. Rozważono dwa warianty realizacji tej koncepcji różniące się co do zasady realizacją pierwszego kroku. W pierwszym wariantcie najpierw rozwiązuje się problem  $Rm|r_j^- \leq r_j \leq r_j^+|Reg(Cmax)$ , gdzie zakres zmienności terminów gotowości zadań wyznacza się jako przedział, którego lewy koniec przyjmuje wartość minimum odległości zadania od możliwych lokalizacji a prawy koniec przyjmuje wartość maksimum tych odległości. W drugim wariantcie zakres zmienności terminów gotowości zależy zarówno od zadań, jak i od maszyn i jest wyznaczony w sposób rekurencyjny. W konsekwencji w kroku pierwszym drugiego wariantu rozwiązuje się problem  $Rm|r_{ij}^- \leq r_{ij} \leq r_{ij}^+|Reg(Cmax)$ . Efektywność



algorytmów reprezentujących opisane podejścia porównano na podstawie przeprowadzonych eksperymentów obliczeniowych.

Sformułowano również złożoną tezę pracy, która jednak brzmi nieco sztucznie. Moim zdaniem ograniczenie się do wniosków z przeprowadzonych badań byłoby w tym przypadku wystarczające.

Niektóre wyniki cząstkowe prezentowane w rozprawie były wcześniej publikowane, głównie w materiałach konferencyjnych. Spójne ujęcie wielu aspektów szeregowania zadań z różnymi sposobami definiowania terminów gotowości zadań w ramach rozprawy napisanej w języku polskim uważam za dodatkowe wartościowe osiągnięcie Autora.

### 3. Poprawność

Przyjęta w rozprawie metodyka badań, polegająca na ustaleniu złożoności obliczeniowej badanego problemu a następnie wykazaniu jego szczególnych własności i zaprojektowaniu algorytmów przybliżonych, których efektywność potwierdza się na podstawie eksperymentów obliczeniowych jest właściwa dla teorii szeregowania zadań. Przedstawione dowody są poprawne a plan eksperymentów i analiza wyników nie budzą zastrzeżeń.

Mam jednak kilka uwag szczegółowych, związanych przede wszystkim z prezentacją materiału badawczego.

Pojęcie harmonogramu (uszeregowania) jest w literaturze jednoznacznie zdefiniowane i obejmuje zarówno przydział zadań do maszyn, jak i określenie terminów zakończenia (równoważnie rozpoczęcia) wykonywania każdego zadania. W definicji podanej na stronie 12 stwierdzono, że uszeregowanie jest to rozwiązanie problemu szeregowania oraz, że „ograniczono się wyłącznie do zagadnień optymalnego podejmowania decyzji (w skrócie do zagadnień optymalizacji)”. Nie jest oczywiste, czy oznacza to, iż uszeregowaniem (rozwiązaniem) jest tylko rozwiązanie optymalne (raczej nie, bo w dalszej części pracy nazywa się rozwiązaniem i uszeregowaniem również rozwiązania przybliżone), czy chodzi o sformułowanie problemu (wersja decyzyjna vs. wersja optymalizacyjna) dla celów analizy złożoności obliczeniowej. Ponadto w proponowanej definicji nie mówi się w sposób jawny, że uszeregowanie to przydział zasobów do zadań w czasie. Na s. 44 znajduje się poprawne stwierdzenie, że harmonogram jest reprezentowany przez macierz binarną  $x$ , należy jednak pamiętać, że harmonogram otrzymuje się dopiero poprzez zastosowanie do macierzy  $x$  równania rekurencyjnego (2.1) wraz z założeniem, że dla  $x_{ijk} = 0$  termin zakończenia zadania  $i$  równa się zero. Brak konsekwencji w stosowaniu definicji uszeregowania powoduje niejasności w opisie podejścia  $Rm|r_j^- \leq r_j \leq r_j^+ |Reg(C_{max})$  (s.108). Mianowicie, w drugim etapie: „przy ustalonym harmonogramie wybieramy rozmieszczenie maszyn, aby uzyskać jak najmniejszą długość uszeregowania w deterministycznym problemie ScheLoc”. Skoro harmonogram (uszeregowanie) jest ustalone, to znaczy, że ustalone są też terminy zakończenia zadań, w tym  $C_{max}$ . Prawdopodobnie chodzi o to, że na pierwszym etapie ustala się jedynie przydział zadań do maszyn i kolejność wykonania zadań na każdej maszynie (macierz  $x$ ), co nie jest kompletnym harmonogramem (uszeregowaniem).

Część uwag dotyczy terminologii i ma charakter dyskusyjny. Uważam, że dbałość o zachowanie poprawności i rozwój języka polskiego należą do obowiązków przedstawicieli świata nauki i dlatego pozwalam sobie te uwagi przedstawić.



Bardzo mało intuicyjnym, i nie występującym dotąd w literaturze, pojęciem jest „niepewność przedziałowa”. Modele, w których występuje co najmniej jeden parametr taki, że w momencie podejmowania decyzji znany jest jedynie zbiór wartości, które parametr ten może przyjąć zostały nazwane „modelami strategicznymi” (p. M. Siudak, *Badania operacyjne*, Oficyna Wydawnicza Politechniki Warszawskiej 1994). Niepewność przedziałowa kojarzyć się może raczej z reprezentacją danych niepewnych za pomocą zbiorów rozmytych lub przybliżonych. Nie jest poprawne traktowanie równoważnie modelu, w którym „wystąpienie dowolnej wartości parametru z ustalonego przedziału jest równo prawdopodobne” (parametr ma rozkład równomierny) z modelem, w którym „niewielka ilość danych uniemożliwia estymację parametrów rozkładu prawdopodobieństwa” (s. 20). W pierwszym przypadku można rozwiązywać problem za pomocą metod probabilistycznych, w drugim natomiast najczęściej stosuje się metody teorii gier. Kwestia ta wymaga co najmniej bardziej szczegółowego omówienia.

Na stronie 14 pojawia się następujące wyjaśnienie: „Pod pojęciem niepewności rozumiemy nieprecyzyjnie określoną wartość parametru...” Jest to w zasadzie opisanie terminu (niepewność) przez inny niezdefiniowany termin (nieprecyzyjność). W badaniach operacyjnych przyjęto klasyfikację ze względu na charakter parametrów na modele deterministyczne i nondeterministyczne. Zachowanie przyjętej terminologii znacznie ułatwiłoby lekturę. Ponadto określenia „nieustalony”, „niepewny”, „nondeterministyczny”, „nieprecyzyjny”, są stosowane zamiennie, jednak nie są to pojęcia równoznaczne. Czytelnikowi trudno się zorientować, czy jest to zabieg stylistyczny (unikanie powtórzeń), czy chodzi o różne modele.

Neologizmy, budzące moje wątpliwości to „determinizacja” (s.41) (co prawda występuje w teorii automatów pojęcie „determinizacja automatu skończonego, ale ma zupełnie inne znaczenie) oraz nieustaloność (s. 39).

Drobne uwagi dotyczą:

- określenia użytego na s. 20: „jest wektorem precyzyjnie określonych parametrów”. W tym przypadku nie chodzi o precyzję, ale o fakt, że parametry mają charakter deterministyczny,
- braku konsekwencji w określeniu zmiennych: decyzja, zmienna decyzyjna, zmienna optymalizacyjna (tego ostatniego określenia dotąd nie spotkałam),
- określenie problemów ScheLoc i  $Rm|_{r_{ij}}|C_{max}$  jako „tożsame” – w teorii złożoności obliczeniowej mówi się o redukcji wielomianowej między problemami, ewentualnie można mówić o problemach równoważnych,
- nie objaśniono oznaczenia „t” we wzorze (2.6),
- odwołania do wzorów nie zawsze zgadzają się z numeracją tychże, na przykład w Tabeli 4.1 na s. 110 zamiast (4.12) i (4.13) powinno być odpowiednio (4.13) i (4.14),
- nie rozumiem w jaki sposób zarówno dolną, jak i górną granicę długości uszeregowania można wyznaczyć jednym wzorem (2.19) – wzór (2.19) wyznacza dolne ograniczenie,
- doceniam, że praca jest napisana w języku polskim, użyto jednak wiele terminów w języku angielskim mających powszechnie używane tłumaczenia na język polski, np. partition problem – problem podziału, 3-partition problem – problem trójpodziału, subset sum problem – problem sumy podzbioru (s. 46), Simulated Annealing – symulowane wyżarzanie Tabu Search – przeszukiwanie tabu (s.17),
- na s. 29 wspomniano o problemach szeregowania zadań „z nierównymi terminami gotowości”, poprawnie mówi się o zadaniach z *dowolnymi* terminami gotowości, gdyż jest to



- pojęcie bardziej ogólne, uwzględniające zarówno terminy zależne od zadań (potencjalnie różne) i niezależne od zadań,
- nie przedstawiam listy literówek, gdyż są one nieliczne i nie wpływają na pozytywną ocenę rozprawy.

#### 4. Wiedza kandydata

Wiedzę kandydata oceniam na podstawie Rozdziału 1 (Wstęp) oraz cytowanej bibliografii.

Wprowadzenie do tematyki pracy wyjaśnia genezę i motywację podjętych zadań szczegółowych, które zasadniczo zmierzają do zaproponowania efektywnych metod rozwiązywania problemu ScheLock polegającego na równoczesnym wyborze lokalizacji maszyn i szeregowaniu zadań. W kolejnym podrozdziale przedstawiono bardzo skrótowo przegląd technik optymalizacji wykorzystanych w pracy. Autor cytuje klasyczne pozycje omawiające rozważane podejścia, co jest dobrym rozwiązaniem biorąc pod uwagę bogactwo literatury w tym zakresie. Moje zastrzeżenia budzi jedynie charakterystyka podejścia zachłannego, określenie „krótkowzroczne decyzje” nie jest precyzyjne. Uważam, że zwięzła definicja: ”Algorytm zachłanny, to algorytm, który w celu wyznaczenia rozwiązania w każdym kroku dokonuje zachłannego, tj. najlepiej rokującego w danym momencie wyboru rozwiązania częściowego” (T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Wprowadzenie do algorytmów*, WNT 2007) lepiej oddaje istotę algorytmu zachłannego. W podsumowaniu zestawienia technik optymalizacji wykorzystanych w pracy (s. 19) stwierdza się, że „wydajność przeglądu zupełnego można poprawić stosując algorytm z nawrotami”. Czy rozważano możliwość zastosowania algorytmu podziału i ograniczeń? Następny podrozdział poświęcono modelowaniu niepewności w problemach szeregowania zadań. Mimo cytowania dosyć licznych i podstawowych pozycji literatury, ten podrozdział napisany jest chaotycznie, terminologia nie jest konsekwentnie stosowana. Przejrzystą klasyfikację modeli badań operacyjnych ze względu na charakter dostępnych informacji przedstawiono m.in. w skrypcie M. Siudak, *Badania operacyjne*, Oficyna Wydawnicza Politechniki Warszawskiej 1994. Wykorzystanie tej klasyfikacji znacznie uporządkowałoby prezentację tej problematyki. Pojawienie się w tym miejscu kolejnego podrozdziału zatytułowanego „Przegląd literatury” jest nieco zaskakujące, gdyż poprzednie podrozdziały zawierały odpowiednie powołania na literaturę. Zdecydowanie brakuje określenia już tytule podrozdziału, jakiego zagadnienia dotyczy prezentowany tutaj przegląd literatury.

Bibliografia, zawierająca 161 pozycji obejmuje zagadnienia dotyczące podstaw teorii szeregowania zadań, podejmowania decyzji w warunkach niepewności, technik optymalizacji wykorzystanych w pracy oraz pewnych szczególnych problemów szeregowania zadań związanych z modelami omawianymi w rozprawie. Lista jest bogata, jednak uważam, że są bardziej reprezentatywne monografie dotyczące teorii szeregowania zadań niż cytowana praca Pinedo, m. in. J. Błażewicz, K. H. Ecker, E. Pesch, G. Schmidt, M. Sterna, J. Węglarz, *Handbook on Scheduling, From Theory to Practice*, Springer Cham 2019 czy *Handbook of Scheduling, Algorithms, Models, and Performance Analysis*, Joseph Y-T. Leung (ed.) Chapman and Hall/CRC 2004. W szczególności warto zapoznać się z pracami, w których wprowadzono terminologię dotyczącą teorii szeregowania zadań w języku polskim, jak *Teoria szeregowania zadań*, E.D. Coffmann (red), WNT 1980 oraz J. Błażewicz, *Złożoność obliczeniowa problemów kombinatorycznych*, WNT 1988, gdyż ułatwia to poprawną pod względem językowym redakcją publikacji. Prezentowany przegląd jest jednak wyczerpujący i świadczy o dobrej znajomości aktualnego piśmiennictwa w tematyce szeroko powiązanej z problemem lokalizacji maszyn i szeregowania zadań. W bibliografii znalazłam jeden błąd w nazwisku Rinnooy Kan: zamiast A.R. Kan powinno być A.H.G. (Alexander Hendrik George) Rinnooy Kan. Jest to jedna z najważniejszych postaci w teorii szeregowania zadań, choć już od wielu lat prowadząca głównie działalność polityczną a nie naukową.

## 5. Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach stwierdzam, że rozprawa zawiera oryginalne rozwiązanie problemu naukowego, pozwala stwierdzić, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka techniczna i telekomunikacja oraz umiejętność samodzielnego prowadzenia pracy naukowej i w związku z tym spełnia wymagania zdefiniowane przez z art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (z późniejszymi zmianami), a zatem wnoszę o jej dopuszczenie do publicznej obrony.

  
\_\_\_\_\_  
Podpis