dr hab. Beata Bylina
Instytut Informatyki
Uniwersytet Marii Curie-Skłodowskiej
Pl. M. Curie-Skłodowskiej 5
20-031 Lublin
email: beata.bylina@umcs.pl

Lublin, 24 października 2023

Recenzja rozprawy doktorskiej

Tytuł rozprawy:	**Modelowanie dynamiki transmisji internetowych za pomocą aproksymacji płynnej**
Autor rozprawy:	**mgr inż. Monika Nycz**
Promotor rozprawy:	**prof. dr hab. inż. Tadeusz Czachórski**
Dziedzina:	**nauki inżynieryjno-techniczne**
Dyscyplina:	**informatyka techniczna i telekomunikacja**

1 Temat i cel rozprawy

W dzisiejszych czasach możliwość sprawnej realizacji transmisji internetowych jest niezbędna, zarówno w nauce, technice, jak i przemyśle. Transmisja internetowa związana jest z przesyłaniem informacji, gdzie jako medium używana jest sieć rozległa, a mianowicie Internet. Informacje przesyłane są z jednego miejsca w inne. Odległość może być zarówno bardzo duża, jak i bardzo mała. Przesyłane informacje podzielone są na fragmenty. Ilość informacji, a zatem również ilość fragmentów, może być mała (np. jeśli przesyłamy krótki tekst składający się tylko ze znaków) lub bardzo duża (np. podczas przekazywania materiału audiowizualnego). Zatem transmisja internetowa charakteryzuje się zmienną intensywnością. Analiza szybkości, z jaką zmienia się dynamika transmisji internetowej, jest bardzo ważna, ponieważ związana jest z zapewnieniem niezawodności dostępu do zasobów i jakości usług oferowanych przez Internet. Badanie dynamiki transmisji internetowej dla rzeczywistej sieci jest bardzo skomplikowane, dlatego w tej sytuacji bardzo istotne jest tworzenie modeli matematycznych. Powinny one szybko i dokładnie generować wyniki modelowania. Takimi modelami są modele oparte na aproksymacji płynnej, które w przeciwieństwie do innych modeli są słabo zbadane, dlatego uzasadnienie wyboru tematu nie przedstawia żadnych wątpliwości.

Autorka postawiła sobie za cel opracowanie zarówno metod, jak i narzędzi programowych do modelowania złożonych i rozległych sieci komputerowych oraz zastosowanie ich do badania wybranych scenariuszy dynamiki transmisji internetowych. Z tym celem głównym wiążą się niemniej ważne cele dodatkowe, mianowicie wykorzystanie przetwarzania równoległego do numerycznego rozwiązywania bardzo dużych układów równań różniczkowych pierwszego rzędu opisujących modele aproksymacji płynnej oraz zastosowanie technologii bazodanowych do gromadzenia i przetwarzania danych dotyczących parametrów sieci. Pozwala to na dokonywanie analiz sieci złożonych i rozległych — a taką jest Internet — wraz z występującymi w nich stanami nieustalonymi oraz porównanie opracowanych metod z już istniejącymi.

2 Zawartość i charakter rozprawy

Rozprawa składa się z 7 rozdziałów. Pierwszy z nich rozpoczyna się krótkim wstępem, w którym zawarte jest wprowadzenie do pracy, przegląd matematycznych modeli stosowanych do badania sieci komputerowych, a także tezy i zakres rozprawy oraz studium literatury. Rozdział 2 poświęcony jest porównaniu metod przybliżonych, a mianowicie aproksymacji dyfuzyjnej i płynnej na przykładzie trzech scenariuszy badawczych dla pojedynczego węzła sieci szkieletowej różniącymi się algorytmami zarządzania kolejką. Rozdział 3 to przegląd informacji dotyczących topologii sieciowych oraz ich podziału, a także generatorów stosowanym dla sieci Internet. Szczegółowo opisany jest generator sieci rzeczywistych, a mianowicie projekt OPTE. Pozwala on na wizualizacje sieci globalnej. Kolejne trzy rozdziały dotyczą modelowania dynamiki transmisji internetowych za pomocą aproksymacji płynnej. Rozdział 4 stanowi szczegółowy opis numerycznej metody wyznaczania modelu aproksymacji płynnej wraz z jego wersją równoległą. W szczegól-

ności, dokładnie została przedstawiona motywacja do utworzenia efektywnego algorytmu równoległego na komputery o architekturach wielordzeniowe z pamięcią wspólną. Dodatkowo zbadano wpływ przypisania wątków logicznych do fizycznych rdzeni na wydajność obliczeniową. Kolejny to bardzo ciekawy rozdział opisujący alternatywne metod wyznaczania modelu aproksymacji płynnej z wykorzystaniem systemów bazodanowych. Zwłaszcza interesujące jest przedstawienie przekształcenia klasycznej implementacji w językach programowania wysokiego poziomu na język zrozumiały dla przetwarzania bazodanowego, strumieniowego i grafowego oraz procesu ETL. Rozdział 6 dotyczy rozważań nad nowymi modelami aproksymacji płynnej. Ostatni rozdział jest podsumowaniem pracy.

Rozprawa jest więc samodzielnym dziełem, w którym Doktorantka opisuje metody matematyczne i narzędzia programowe powołując się także na swoje publikacje w międzynarodowych czasopismach naukowych.

Najważniejsze osiągnięcia Doktorantki przedstawione w rozprawie to:

- porównanie pod względem dokładności dwóch metod przybliżonych, a mianowicie aproksymacji dyfuzyjnej i płynnej na przykładzie pojedynczego węzła z różnymi typami kolejki;
- opracowanie metody wyznaczania tras połączeń i parametrów routerów dla rozległych sieci komputerowych;
- stworzenie efektywnego równoległego algorytmu dla metody aproksymacji płynnej na komputery o architekturach wielordzeniowych z pamięcią wspólną;
- wykorzystanie technologii bazodanowych do gromadzenia; przechowywania i obliczania parametrów sieci rozległych przy użyciu aproksymacji płynnej;
- rozszerzenie klasycznego modelu aproksymacji płynnej.

Wszystkie zagadnienia są przez Autorkę bardzo szczegółowo omówione. Dodatkowo dobrze jest opisana metodologia badawcza oraz uzasadnienie jej wyboru.

3 Teza rozprawy

W rozdziale 1 Doktorantka wskazuje, w sposób wyraźny, aż trzy tezy i brzmią one następująco:

1. *Metoda aproksymacji płynnej, po opracowaniu metod i narzędzi programowych, jest efektywnym sposobem badania sieci rozległych z uwzględnieniem występujących w nich stanów nieustalonych.*
2. *Przetwarzanie współbieżne można zastosować do numerycznego rozwiązywania równań stanowiących model aproksymacji płynnej.*
3. *Możliwe jest zastosowanie technologii właściwych dla baz danych i hurtowni danych do modelowania za pomocą aproksymacji płynnej, przechowywania danych uzyskanych podczas modelowania oraz ekstrakcji wiedzy dotyczącej stanu sieci komputerowej.*

Tezy są sformułowane jasno i poprawnie. Odzwierciedlają one dokładnie to, nad czym Autorka w rozprawie rzeczywiście pracowała. Tezy są także oryginalne i wykraczają poza obecny stan wiedzy zarówno w tematyce modelowania z użyciem aproksymacji płynnej, jak i przetwarzania równoległego wykorzystanego do przyspieszenia numerycznego rozwiązywania równań różniczkowych opisujących model oraz wykorzystanie technologii bazodanowych w bardzo nietypowy sposób.

Rozprawa wykazuje prawdziwość tez. Przedstawiony przez Kandydatkę opis równoległego algorytmu aproksymacji płynnej na komputery wielordzeniowe z pamięcią wspólną, a także przeprowadzone testy numeryczne oceniające czas wykonania algortymu dla różnej liczby wątków potwierdzają prawdziwość tezy drugiej. Opis przekształcenia klasycznego algorytmu wyznaczania modelu aproksymacji płynnej na język zrozumiały dla języka charakterystycznego do przetwarzania związanego z szeroko rozumianymi bazami danych, a także wyniki przeprowadzonych numerycznych doświadczeń potwierdzają trzecią tezę pracy. Wszystkie opracowane metody i narzędzia pozwalają na przeprowadzanie analiz sieci rozległych z uwzględnieniem występujących w nich stanów nieustalonych, co potwierdza tezę pierwszą.

Wszystkie te fakty pozwalają na stwierdzenie, że tak ważny dla oceny pracy doktorskiej aspekt, jakim jest postawienie i udowodnienie tez badawczych, został spełniony.

4 Analiza źródeł i zastany stan wiedzy

Bibliografia recenzowanej rozprawy doktorskiej jest obszerna i obejmuje aż 204 pozycje. Są to zarówno pozycje w języku polskim, jak i angielskim. Bibliografia daje wgląd do pozycji, na podstawie których została napisana rozprawa. Źródła są zarówno dość ogólne, jaki i szczegółowe (głębiej dotykających istoty tematu). Są to książki, artykuły, dokumentacje, jaki i dokumenty zamieszczone w Internecie. Prace cytowane są w odpowiednim kontekście.

Źródła te dobrze przedstawiają bieżący stan wiedzy na tematy poruszane w pracy. Daje to możliwość poszerzenia wiedzy czytelnika, a jednocześnie prezentuje szeroką wiedzę Doktorantki w dziedzinie modelowania dynamiki transmisji internetowych za pomocą narzędzi matematycznych z wykorzystaniem dostępnych technologii komputerowych. W większości są to źródła nowe (ostatnie pochodzą z 2023 roku), co całkowicie naturalne (i pożądane) w tego rodzaju tematyce. Autorka cytuje także starsze, bardziej klasyczne publikacje, co pozwala osadzić pracę w całości dorobku naukowego dyscypliny oraz dyscyplin pokrewnych. Pozwala to pokazać, że zainteresowanie modelowaniem działania pracy sieci komputerowych przy użyciu różnych narzędzi matematycznych nie jest nowe, ale i nie słabnie, a także wymaga dalszego rozwoju — w szczególności ze względu na rosnący wciąż rozmiar sieci Internet.

5 Znaczenie wyników

Kandydatka w swej rozprawie zajmuje się modelowaniem dynamiki transmisji internetowych za pomocą aproksymacji płynnej. Sam pomysł nie jest oczywiście niczym nowym.

Ciekawym wynikiem są wnioski płynące z porównania dwóch metod przybliżonych, a mianowicie aproksymacji dyfuzyjnej i płynnej z symulacją przy użyciu symulatora zdarzeń dyskretnych OMNeT++ i wynikające z nich wskazanie, że aproksymacja płynna może być stosowana dla sieci o setkach tysięcy węzłów, zaś aproksymacja dyfuzyjna dla sieci znacznie mniejszych.

Innym ciekawym wynikiem jest powstanie równoległego efektywnego algorytmu aproksymacji płynnej. Dodatkowo przedstawione wyniki eksperymentów numerycznych pokazują, że wybór odpowiedniej liczby wątków oraz odpowiedniego przypisania tychże wątków do rdzeni fizycznych ma wymierny wpływ na wydajność zaproponowanego algorytmu równoległego.

Innym aspektem, który stanowi o oryginalności rozwiązania problemu związanego z wyznaczaniem parametrów model aproksymacji płynnej jest użycie technologii bazodanowej w nietypowym kontekście. Bazy danych i hurtowanie danych są technologiami przeznaczonymi do gromadzenia, przechowywania i przetwarzania informacji. Zwykle te informacje nie są związane z danymi numerycznymi, ale z życiem codziennym. Luka co do zakresu wykorzystania technologii bazodanowych do obliczeń numerycznych, została szeroko opisana w niniejszej rozprawie. Doktorantka bardzo trafnie uwzględniła możliwości zastosowania technologii baz danych i hurtowni danych do wyznaczenia modelu aproksymacji płynnej. Wydaje się, że gromadzenie i przechowywanie informacji o parametrach modeli opisujących działanie sieci Internet jest istotnym, ale niedocenianym aspektem modelowania. Pozwala na łatwe i w miarę szybkie pozyskiwanie informacji o stanie sieci.

6 Redakcja rozprawy i prezentacja wyników

Struktura pracy i jej układ są poprawne. Tytuły rozdziałów i podrozdziałów brzmią komunikatywnie, umożliwiaj szybką orientację. Kolejność poszczególnych rozdziałów i podrozdziałów układa się w logiczną sekwencję. Co do zasady objętość poszczególnych rozdziałów i podrozdziałów jest poprawna. Zastrzeżenie budzi jedynie niepotrzebny według recenzentki podział rozdziału 4.1 na kolejny i to tylko jeden podrozdział. Takie rozdrobnienie wydaje się za szczegółowe w tym konkretnym przypadku.

W pracy zawarto liczne rysunki, które mają za zadanie zobrazowanie zagadnień poruszanych w rozprawie. To zadanie jest doskonale wykonane, ponieważ rysunki przekazują czytelnikowi najważniejszych treści. Zdecydowanie też usprawniają proces zrozumienia przez czytelnika skomplikowanych mechanizmów i algorytmów opisanych w pracy. Każdy rysunek zawarty w pracy ma konkretne znacznie i jest poprawnie podpisany oraz wyjaśniony w tekście. Także interpretacja wykresów zawarta w pracy jest poprawna.

Język pracy jest klarowny i zrozumiały. Pracę dobrze się czyta — napisana jest starannie zarówno pod względem językowym, jak i typograficznym. Można, co prawda, znaleźć bardzo drobne błędy (o których niżej), ale nie jest to coś, co umniejszałoby znaczenie

pracy. Duża liczba ilustracji, wykresów oraz tabel ułatwia zrozumienie trudnych zagadnień.

Podsumowując, Doktorantka zachowuje normy przygotowania i redagowania tekstu naukowego.

7 Słabe strony i uwagi krytyczne

Zdecydowana większość tez, twierdzeń i poglądów wyrażonych w pracy w ocenie recenzentki jest trafna. Praca ma bardzo niewiele elementów, do których można mieć wyraźne uwagi krytyczne.

- Wykresy na rysunkach 2.1 i 2.7 niestety nie są w pełni formalnie poprawne, bo możemy z nich odczytać, że prawdopodobieństwo odrzucenia ostatniego pakietu jest w przedziale $[0; 1]$, a z opisu zawartego w tekście wynika, że prawdopodobieństwo jest dokładnie równe 1.

- Pewien niedosyt pozostawia rozdział 3 (*Topologie sieciowe*). Wskazany byłby dokładniejszy przegląd literatury dotyczący teorii możliwych podziałów topologii sieci, a także rozszerzenie informacji na temat narzędzi służących do generowania struktury sieci.

- Wątpliwość budzi także liczba testowanych wątków w rozdziale 4.2.3. Użycie aż 600 czy 800 wątków programowych na komputerach zawierających znacznie mniej fizycznych rdzeni wydaje się nie do końca uzasadnione. W wielu scenariuszach testowanie liczby wątków większej niż 100 nie wpływa na czas wykonania.

- Niektóre kwestie poruszone w pracy wydają się zbyt obszerne. Dotyczy to kwestii przypisania wątków programowych do fizycznych rdzeni (podrozdział 4.2.3.2 — *Scenariusz dystrybucyjny*). Za nie do końca trafione należy uznać szereg wykresów dotyczących rozmieszczenia wątków na poszczególnych procesorach (wykresy 4.21–4.36). Umieszczenie w pracy szesnastu rysunków, a w każdym z nich aż trzech wykresów utrudnia ich analizę. Warto byłoby dodać czytelne podsumowanie z jasnymi, komunikatywnymi wnioskami.

- W pracy pewne zastrzeżenie budzi redakcja podpisów pod rysunkami. Czasami podpisy pod rysunkami kończą się kropkami, a czasmi nie. Ogólną dobrą jakość i czytelność wykresów i diagramów z rzadka zaburza dobór rozmiaru czcionek (zbyt małego). Dodatkowo na rysunkach pojawiają się napisy nie uwzględnione w tekście, np: na rysunku 4.12 (strona 60) można znaleźć tekst „Wyznaczenie wartości L_0 dla routerów". W całym tekście nie ma żadnej zmiennej o tej nazwie. Jest to mylące.

- Do pełnego zrozumienia i docenienia pracy Doktorantki brakuje w tekście rozprawy wskazania repozytorium z pełnym kodem źródłowym, co jest już pewnym standardem w nauce (ze względu na konieczność powtarzalności doświadczeń). Dodatkowo pozwoliłoby to na zbadanie zaproponowanego algorytmu równoległego aproksymacji płynnej na najnowszych architekturach wielordzeniowych.

- Bardzo drobnymi problemami są nieliczne błędy interpunkcyjne/typograficzne (jak brakujące przecinki, użycie łącznika w miejsce pauzy itp.).

Powyższe uwagi krytyczne nie wpływają na merytoryczną wartość pracy w żaden sposób, a są jedynie uchybieniami zrozumiałymi przy tego rodzaju i wielkości pracy, a także subiektywnym zdaniem recenzentki.

8 Podsumowanie i wniosek końcowy

Praca jest wartościowa, ciekawa, na wysokim poziomie merytorycznym. Po analizie rozprawy Doktorantki mogę stwierdzić, że jest ona przygotowana rzetelnie i wnosi znaczący wkład w dyscyplinę *informatyka techniczna i telekomunikacja*. Potwierdza ona też zdolność Kandydatki do prowadzenia dalszej pracy naukowej samodzielnie.

Stwierdzam, że recenzowana rozprawa pt. „Modelowanie dynamiki transmisji internetowych za pomocą aproksymacji płynnej" spełnia warunki określone w Ustawie *Prawo o szkolnictwie wyższym i nauce*. W związku z tym wnioskuję o dopuszczenie rozprawy doktorskiej Pani mgr. inż. Moniki Nycz do publicznej obrony i dalszych etapów postępowania doktorskiego.

Beata Bylina