

РЕЦЕНЗІЯ

на дисертаційну роботу **Мітікова Миколи Юрійовича**
на тему: **«Ідентифікація аномалій в роботі програмного забезпечення на
основі аналізу знімків пам'яті»**,
представлену на здобуття ступеня доктора філософії
за спеціальністю 113 Прикладна математика

Актуальність теми дослідження.

У дисертаційній роботі М.Ю. Мітікова досліджується задача формалізації, алгоритмізації та автоматичної діагностики сценаріїв надмірного використання оперативної пам'яті на основі аналізу знімків пам'яті. Розроблені математичні моделі, алгоритмічні процедури та продукційна система орієнтовані переважно на програмні застосунки керованого середовища виконання .NET/CLR і враховують особливості структури керованої купи, системи типів та механізмів автоматичного керування пам'яттю.

Актуальність теми дослідження не викликає сумнівів, оскільки зі зростанням складності програмних систем та обсягів даних підвищуються вимоги до ефективного використання оперативної пам'яті. Існуючі засоби моніторингу та профілювання переважно орієнтовані на виявлення витоків пам'яті та контроль загальних показників споживання ресурсів, однак не завжди дозволяють виявити сценарії структурно надмірного використання пам'яті, пов'язані з дублюванням даних, неефективним поданням об'єктів, надлишковими службовими витратами колекцій та іншими формами структурно надлишкового подання інформації в оперативній пам'яті. У зв'язку з цим розроблення математичних моделей, алгоритмічних процедур та засобів експертного аналізу знімків пам'яті для ідентифікації таких сценаріїв є актуальним науковим і практичним завданням.

Актуальність роботи зумовлена потребою у підвищенні продуктивності та масштабованості сучасних програмних систем, для яких неефективне використання оперативної пам'яті може призводити до зростання апаратних витрат, погіршення швидкодії та зниження надійності функціонування програмного забезпечення. Тематика роботи відповідає напряму прикладної математики, пов'язаному з моделюванням та алгоритмічним аналізом складних інформаційних систем.

Зв'язок роботи з науковими програмами, планами, темами.

Дисертаційне дослідження виконано відповідно до планів наукової роботи кафедри комп'ютерних технологій Дніпровського національного університету імені Олеся Гончара в межах тем «Детерміновані та стохастичні алгоритми комп'ютерного моделювання об'єктів та процесів різної природи» (номер державної реєстрації 0122U001467) та «Розробка високоефективних комп'ютерних алгоритмів для аналізу та ідентифікації параметрів математичних моделей» (номер державної реєстрації 0125U002277).

Формулювання наукової задачі, нове вирішення якої одержане в

дисертації.

Метою дисертаційної роботи є розроблення формальної моделі знімка пам'яті, критеріїв та алгоритмічних процедур ідентифікації сценаріїв надмірного використання оперативної пам'яті у програмних застосунках керованого середовища виконання, а також побудова експертної системи для виявлення фактів надмірного використання пам'яті, надання рекомендацій щодо їх усунення та оцінювання ефекту від упровадження запропонованих змін.

Основні завдання дослідження:

- проаналізувати існуючі підходи до подання, інтерпретації, аналізу та зменшення використання оперативної пам'яті у програмних застосунках керованого середовища виконання;
- побудувати формальну модель знімка пам'яті як скінченної типізованої структури, що відображає об'єкти пам'яті, їхні типи, зв'язки, значення, розміри та службові елементи середовища виконання;
- сформулювати задачу ідентифікації сценаріїв надмірного використання пам'яті на основі формалізованої структури знімка пам'яті;
- визначити критерії, предикати та кількісні характеристики для виявлення окремих класів надмірного подання інформації в оперативній пам'яті;
- розробити алгоритмічні процедури обчислення характеристик надмірності та оцінювання можливого зменшення обсягу пам'яті після допустимих перетворень подання даних;
- побудувати продукційну модель застосування правил до формалізованого подання знімка пам'яті для визначення сценаріїв надмірності, рекомендацій та кількісних оцінок;
- перевірити запропоновані моделі, критерії та алгоритмічні процедури на знімках пам'яті програмних застосунків.

Об'єктом дослідження є знімок пам'яті як зафіксований стан оперативної пам'яті програмного процесу у визначений момент часу.

Предметом дослідження є математичні моделі, критерії, алгоритмічні процедури та продукційні правила ідентифікації сценаріїв надмірного використання оперативної пам'яті за формалізованим поданням її знімка, а також кількісні оцінки можливого зменшення обсягу пам'яті після допустимих перетворень подання даних.

Новизна наукових положень, висновків і рекомендацій.

Наукова новизна одержаних результатів полягає в такому:

1. Вперше запропоновано формальну модель знімка пам'яті програмного процесу як скінченної типізованої структури об'єктів, у якій байтове представлення пам'яті інтерпретується через множини об'єктів, типів, адрес, розмірів, значень полів і зв'язків між об'єктами, що задає математично визначене подання стану оперативної пам'яті.

2. Вперше сформульовано задачу ідентифікації надмірного використання оперативної пам'яті як задачу виявлення класів надлишкового подання інформації у типізованій структурі об'єктів з урахуванням класів еквівалентності за інформаційним вмістом, безнадлишкового представлення

та кількісного коефіцієнта надмірності.

3. Дістав подальшого розвитку алгоритмічний підхід до виявлення дублікатів незмінюваних об'єктів у знімках пам'яті шляхом попереднього групування об'єктів за типом і швидко обчислюваними ознаками інформаційного вмісту з подальшою перевіркою еквівалентності всередині отриманих підмножин, що зменшує кількість порівнянь порівняно з повним попарним перебором.

4. Дістав подальшого розвитку метод кількісного оцінювання можливого зменшення використання пам'яті для основних класів надмірного подання даних, зокрема дублювання незмінюваних об'єктів, надлишкового діапазону типів полів, втрат через вирівнювання об'єктів, неефективного використання хеш-колекцій і стискання масивів байтів зі збереженням інформаційного вмісту.

5. Дістала подальшого розвитку продукційна модель представлення знань для аналізу знімків пам'яті, у якій правила задаються предикатами над формалізованою структурою знімка пам'яті, а результат логічного виведення визначається як множина сценаріїв надмірного використання пам'яті, відповідних рекомендацій та кількісних оцінок.

6. Дістала подальшого розвитку система метрик валідації рекомендацій щодо зменшення використання пам'яті, яка пов'язує виявлені сценарії надмірності, прогнозоване зменшення обсягу пам'яті та фактичну зміну споживання пам'яті після реалізації рекомендацій за допомогою кількісних показників точності, повноти та похибки прогнозування.

Практичне значення одержаних результатів.

Практичне значення одержаних результатів полягає у застосуванні запропонованих моделей, критеріїв, продукційних правил та алгоритмічних процедур для автоматизованого аналізу знімків пам'яті, виявлення сценаріїв надмірного використання оперативної пам'яті, формування рекомендацій щодо оптимізації структур даних та оцінювання очікуваного ефекту від впровадження запропонованих змін. Програмна реалізація використовується як інструментальний засіб аналізу знімків пам'яті, перевірки працездатності запропонованих моделей і алгоритмів та оцінювання ефективності рекомендацій щодо зменшення використання оперативної пам'яті. Практичне значення результатів дослідження підтверджується актом впровадження від компанії Relewise ApS, згідно з яким розроблені методи аналізу знімків пам'яті та виявлення аномалій використання оперативної пам'яті були використані у процесах розробки та діагностики програмної платформи Relewise.

Обґрунтованість і достовірність наукових положень, висновків і рекомендацій, які захищаються.

Дисертаційна робота виконана на належному науковому рівні. Обґрунтованість і достовірність одержаних результатів забезпечуються коректністю математичних постановок задач, узгодженістю моделі, критеріїв і алгоритмічних процедур, використанням методів теорії множин, дискретної математики, структурного та алгоритмічного аналізу даних, продукційного представлення знань і математичної статистики.

Результати перевірено на представницькій вибірці промислових знімків пам'яті високонавантажених .NET-застосунків, де підтверджено можливість автоматизованого виявлення сценаріїв структурної надмірності використання оперативної пам'яті та отримання кількісних оцінок потенційного ефекту від впровадження рекомендацій. Для сформульованих рекомендацій проведено порівняння прогнозованого та фактичного зменшення обсягу використовуваної пам'яті після реалізації запропонованих змін у програмному коді, що підтвердило практичну придатність розроблених моделей і алгоритмічних процедур. Основні результати дослідження відображено у 18 наукових працях.

Особистий внесок здобувача.

Усі результати дисертаційної роботи, що виносяться на захист, отримані автором особисто. У працях, опублікованих у співавторстві, особистий внесок здобувача визначено в тексті дисертації.

Дисертантом проведено аналіз літературних джерел, сформульовано задачу дослідження, побудовано формальну модель знімка пам'яті, визначено критерії та класи надмірного подання інформації, розроблено алгоритмічні процедури ідентифікації сценаріїв надмірного використання оперативної пам'яті, продукційну модель та експертну систему підтримки прийняття рішень, виконано експериментальну перевірку запропонованих підходів, узагальнено та проаналізовано отримані результати.

Структура та обсяг дисертації.

Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, переліку використаних джерел зі 104 найменувань та додатків. Загальний обсяг роботи становить 170 сторінок, з них 147 сторінок основного тексту. Робота містить 23 рисунки та 7 таблиць.

У вступі обґрунтовано актуальність теми, сформульовано мету і завдання, визначено об'єкт, предмет та методи дослідження, наведено наукову новизну і практичне значення результатів.

У першому розділі наведено аналітичний огляд методів та підходів виявлення проблем продуктивності програмного забезпечення та аналізу знімків пам'яті. Розмежовано витoki пам'яті та стабільне надмірне використання ресурсу. Приділено увагу онтологічним, логічним, rule-based та експертним підходам до інтерпретації діагностичних ознак.

У другому розділі сформовано математичну основу дослідження: побудовано формальну модель знімка пам'яті, введено критерії та класи надмірного використання, подано процедури аналізу дублювання, хеш-колекцій, діапазонів типів, вирівнювання й стискання масивів байтів.

У третьому розділі наведено результати експериментальної перевірки. Виконано кількісний аналіз промислових знімків, досліджено прискорений пошук дублікатів та перевірено вплив запропонованих змін на використання пам'яті.

У четвертому розділі подано продукційну експертну систему, визначено базу фактів, базу правил, механізм логічного виведення та структуру результатів застосування правил.

Висновки по роботі є логічними, відповідають поставленим завданням і впливають із теоретичних та експериментальних результатів, отриманих автором.

Зауваження до дисертації

Позитивно оцінюючи дисертаційну роботу в цілому, вважаю за необхідне висловити такі зауваження:

1. Потребує додаткового обґрунтування ряд положень, заявлених у розділі наукової новизни як отриманих вперше. Зокрема, доцільно було б детальніше показати відмінності запропонованої формальної моделі знімка пам'яті, постановки задачі ідентифікації надмірного використання пам'яті та продукційної системи правил від існуючих підходів до аналізу надлишкового використання пам'яті, представлених у сучасній науковій літературі.

2. У підрозділах 2.5.2, 2.5.4 та 2.5.5 оцінювання надмірності використання пам'яті здійснюється переважно за статичним зрізом – одним знімком пам'яті $M(u)$. Водночас звуження типів полів, оцінювання доцільності заміни структур даних та виявлення дублікатів базуються на характеристиках, зафіксованих у конкретний момент часу, що створює ризик хибнопозитивних рекомендацій у разі зміни профілю навантаження або розширення фактичного діапазону значень під час подальшої експлуатації системи. Доцільно було б поширити аналіз серії знімків $M(u_1), \dots, M(u_q)$, який у підрозділі 2.5.1 використовується для виявлення витоків пам'яті, і на перевірку стійкості виявленої структурної надмірності та стабільності спостережуваних характеристик об'єктів.

3. У правилах експертної системи, наведених у підрозділі 4.3, використовуються численні порогові параметри (τ_{range} , τ_{pad} , τ_{cap} , τ_D , τ_{cmp} та інші), від значень яких безпосередньо залежить спрацювання відповідних продукційних правил. Водночас у роботі недостатньо детально описано методику визначення цих параметрів, межі їх застосовності та вплив на результати аналізу. Доцільно було б подати процедуру калібрування порогових значень, а також результати аналізу чутливості системи до їх зміни, що дозволило б більш повно оцінити стійкість запропонованої експертної системи.

4. У розділі 3 зазначено, що дослідження виконувалося на вибірці понад 80 промислових знімків пам'яті, проте детальний аналіз наведено лише для окремих характерних прикладів. У роботі недостатньо висвітлено статистичні характеристики застосування запропонованих критеріїв та правил на повній вибірці, зокрема частоту їх спрацювання, кількість підтверджених і непідтверджених рекомендацій та частку випадків, для яких відповідні сценарії надмірного використання пам'яті були відсутні. Наведення таких узагальнених характеристик дозволило б більш повно оцінити репрезентативність одержаних результатів.

5. У підрозділі 2.5.6 при оцінюванні доцільності стискання масивів типу `System.Byte[]` використовується предикат $Adm_{cmp}(o)$, який визначає допустимість застосування стискання до конкретного об'єкта. Водночас у

роботі недостатньо формалізовано процедуру автоматичного обчислення цього предиката виключно за даними знімка пам'яті. Потребує додаткового уточнення механізм автоматичного відокремлення масивів, придатних до стискання, від службових буферів та інших об'єктів, для яких таке перетворення є недоцільним.

Висловлені зауваження не знижують позитивної оцінки дисертації в цілому, а також розроблених автором основних наукових положень, висновків та отриманих результатів проведених досліджень.

Зміст дисертації відповідає вимогам, що висуваються до наукових робіт такого рівня, вирізняється логічністю, системністю та послідовністю викладення. Структура роботи узгоджується з її назвою, метою і завданнями дослідження. Використані джерела супроводжуються бібліографічними посиланнями; підстав вважати, що в роботі порушено принципи академічної доброчесності, не встановлено. Зміст та результати дисертації відповідають спеціальності 113 «Прикладна математика».

Загальний висновок.

Дисертаційна робота **Мітікова Миколи Юрійовича** на тему *«Ідентифікація аномалій в роботі програмного забезпечення на основі аналізу знімків пам'яті»* є завершеним, самостійно виконаним науковим дослідженням, у якому отримано нові науково обґрунтовані результати, що мають теоретичне та практичне значення для задач аналізу та оптимізації використання оперативної пам'яті програмних систем.

Враховуючи обсяг виконаних досліджень, наукову новизну, обґрунтованість і практичну цінність отриманих результатів, дисертація повністю відповідає вимогам Наказу Міністерства освіти і науки України № 40 від 12 січня 2017 р. та Постанови Кабінету Міністрів України № 44 від 12 січня 2022 р. зі змінами, а її автор Мітіков Микола Юрійович заслуговує присудження ступеня доктора філософії за спеціальністю 113 Прикладна математика.

Рецензент

доцент кафедри
інженерії програмного забезпечення
та інформаційних технологій
Дніпровського національного університету
імені Олеся Гончара
кандидат фіз.-мат. наук, доцент



Лілія БОЖУХА

Підпис канд. фіз.-мат. наук, доцента Божухи Л.М. засвідчую:

Проректор з наукової роботи
Дніпровського національного університету
імені Олеся Гончара
кандидат біологічних наук, професор



Олег МАРЕНКОВ