

РЕЦЕНЗІЯ

**доктора філософії, доцента,
завідувача кафедри програмного забезпечення комп'ютерних систем
Чернівецького національного університету імені Юрія Федьковича
ГАЗДЮК КАТЕРИНИ ПЕТРІВНИ
на дисертаційне дослідження
Кириченка Олександра Олексійовича
на тему «Оптимізація безсерверних обчислень у хмарних середовищах»,
представлену на здобуття наукового ступеня доктора філософії
за спеціальністю 121 – Інженерія програмного забезпечення
галузі знань 12 – Інформаційні технології**

Актуальність теми дисертації.

Стрімкий розвиток та еволюція хмарних технологій зумовили появу інноваційних обчислювальних парадигм, серед яких особливе місце посідають безсерверні обчислення (Serverless computing). Сучасні безсерверні платформи, такі як AWS Lambda, Google Cloud Functions та Azure Functions, реалізують подійно-орієнтовану модель виконання, забезпечують автоматичне масштабування ресурсів та пропонують гнучку модель оплати за фактично використані потужності. Це суттєво спрощує і прискорює процеси розробки, розгортання та адміністрування складних хмарних застосунків, позбавляючи розробників необхідності керувати фізичною чи віртуальною інфраструктурою. Попри значні переваги, практичне розгортання безсерверних архітектур у розподілених системах з високим або нестабільним навантаженням стикається з низкою критичних обмежень. Проблема «холодного старту» (Cold Start) є однією із найвагоміших з них, тобто виникнення суттєвих додаткових затримок під час першого виклику функції або після періоду її тривалого простою. Ще одним обмеженням є реактивний характер автоматичного масштабування, яка полягає в

тому, що традиційні вбудовані механізми хмарних провайдерів виділяють додаткові ресурси лише за фактом перевищення наперед визначених порогових метрик навантаження, що призводить до неефективного використання ресурсів, утворення черг та непередбачуваних затримок. Застосування моделей машинного навчання та штучного інтелекту (наприклад, нейромереж) для прогнозування навантаження вимагає накопичення колосальних обсягів історичних даних, що також є критичним обмеженням, оскільки потребує постійного перенавчання моделей при зміні динаміки трафіку і не надає чітких аналітичних гарантій якості обслуговування (SLA).

Для розв'язання цих проблем безсерверні платформи доцільно розглядати через призму теорії систем масового обслуговування (СМО). Проте класичні моделі СМО нездатні повною мірою відобразити реальну стохастичну структуру безсерверного середовища. На практиці вхідний потік завдань є гетерогенним — він формується як суміш запитів із багатьох незалежних джерел подій із різними статистичними характеристиками, а інтенсивність їх обробки має змінний характер. Таким чином, актуальність дослідження зумовлена об'єктивною потребою у подоланні технологічних обмежень безсерверних обчислень шляхом розробки нових математичних моделей неоднорідної структури та проактивних інформаційних технологій. Створення аналітичних методів, які враховують нерівномірність навантаження та стохастичну природу вхідних потоків без необхідності збору надлишкових історичних даних, має важливе науково-прикладне значення. Це дозволяє підвищити продуктивність розподілених систем, мінімізувати кількість «холодних стартів», оптимізувати фінансові витрати на хмарну інфраструктуру та забезпечити нормативні гарантії якості обслуговування користувачів.

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

Наукові дослідження були виконані здобувачем на кафедрі програмного забезпечення комп'ютерних систем Чернівецького національного університету

імені Юрія Федьковича в межах держбюджетних науково-дослідних робіт «Дослідження, моделювання та розробка програмного забезпечення складних динамічних систем» (№ 0121U109232) та «Інформаційні технології в аспекті сучасних задач прийняття рішень» (номер держреєстрації 0121U109159).

Ступінь обґрунтованості наукових положень, висновків, рекомендацій, сформульованих у дисертації.

Наукові положення, які представлені в дисертаційній роботі, є добре обґрунтованими, а також належно висвітлені у відповідних розділах дисертації. Основні результати, отримані здобувачем та винесені на захист, цілком відповідають меті та завданням роботи, обговорювалися на наукових семінарах кафедри програмного забезпечення комп'ютерних систем ННІФТКН, наукових конференціях та опубліковані у фахових виданнях та у наукових журналах, які індексуються в базах даних Scopus. Достовірність отриманих результатів ґрунтується на використанні загальноприйнятих експериментальних і теоретичних підходів й методів дослідження та не викликає сумнівів. У роботі проведено ґрунтовний огляд літературних джерел з тематики дисертаційного дослідження.

Наукова новизна

Наукова новизна результатів дисертаційного дослідження полягає в наступному:

1. Вперше доведено граничні еволюції для процесу зміни довжини черги у безсерверних обчислювальних середовищах у схемі усереднення та схемі дифузійної апроксимації з використанням апарату напівмарковських випадкових еволюцій. На відміну від відомих рішень, схема усереднення визначає детерміновану траєкторію середнього навантаження λ -функції, а схема дифузійної апроксимації описує випадкові відхилення від цієї траєкторії через стохастичне диференціальне рівняння Іто. Математичним результатом даного аналізу стало встановлення необхідної та достатньої умови обмеженості черги для неоднорідної СМО. Обґрунтованість підходу підтверджується строгою математичною базою, а

достовірність – результатами імітаційного моделювання, що фіксують точність відтворення стохастичних коливань навантаження.

2. Дістала подальшого розвитку математична модель безсерверної обчислювальної системи (на прикладі платформи AWS Lambda) як неоднорідної системи масового обслуговування $H_k(t)/M/\infty$ зі змішаними режимами роботи. Наукова новизна полягає в тому, що, на відміну від класичного підходу Markov-Modulated Poisson Process (MMPP), де інтенсивність змінюється синхронно під впливом єдиного марковського ланцюга, запропонована модель розглядає вхідний процес як суміш незалежних неоднорідних пуассонівських процесів з різних джерел подій (API Gateway, S3, SQS). Кожне джерело оперує власними статистичними характеристиками, що забезпечує більш адекватний опис дисперсії сумарного потоку завдань та підвищує точність прогнозування динаміки системи.

3. Удосконалено підхід до оцінки кількості відхилених завдань (throttling rate) у безсерверних системах з обмеженою чергою на основі нормального наближення. Запропоноване рішення є математичним узагальненням існуючих результатів для СМО типу $M/M/\infty$ з марковською модуляцією на випадок неоднорідного вхідного процесу зі змішаними режимами роботи та неперервним часом. Алгоритмічні та розрахункові принципи підходу, дозволяють здійснювати проактивний аналітичний контроль параметрів системи, що має пряме практичне застосування для визначення та гарантування відповідності вимогам угод про рівень послуг (Service Level Agreement, SLA).

4. Удосконалено алгоритм параметричної оцінки та оптимізації конфігурації безсерверної системи. На відміну від поширених існуючих підходів, побудованих на методах машинного навчання («чорних скриньках»), розроблений алгоритм базується на аналітичних оцінках теорії масового обслуговування. Це дозволило отримати явні розрахункові формули для пошуку оптимальних параметрів конфігурації (обсягу виділеної пам'яті, тайм-аутів), повністю усуваючи

обчислювально містку необхідність збору великих історичних логів та постійного перенавчання нейронних мереж.

5. Вперше спроектовано та реалізовано наскрізну інформаційну технологію (архітектуру фреймворку) для розподіленої обробки даних з використанням безсерверних технологій. На відміну від класичних монолітних чи віртуалізованих рішень, обчислювальні процеси, збереження станів та маршрутизація запитів органічно розподілені між хмарними сервісами з підтримкою горизонтального масштабування. Оригінальність рішення полягає в інтеграції керованої черги повідомлень та реалізації концепції слабкої зв'язності між елементами, що забезпечує незалежне масштабування компонентів, швидке відновлення після збоїв та оптимізацію використання ресурсів. Експериментальні вимірювання підтверджують ефективність створеної технології, фіксуючи суттєве зниження частоти «холодних стартів» та підвищення продуктивності обчислень порівняно з базовими реактивними механізмами масштабування хмарних провайдерів.

Достовірність отриманих наукових результатів підкріплена коректним застосуванням математичного апарату теорії ймовірностей, математичної статистики, теорії систем масового обслуговування, апарату напівмарковських випадкових еволюцій. Результати експериментальної верифікації та імітаційного моделювання, проведених на реальних профілях навантаження безсерверних функцій, повністю узгоджуються з теоретичними висновками автора, що підтверджує наукову та практичну цінність роботи.

Практичне значення одержаних результатів.

Запропонований підхід до проактивного масштабування ресурсів на основі аналітичних моделей теорії масового обслуговування, математична модель безсерверної системи $H_k(t)/M/\infty$ та розроблена архітектура розподіленого фреймворку для хмарних обчислень використовуються у практичній діяльності компанії Finker Finance B.V. та ФОП Вербицькою Світланою Ігорівною для

оптимізації витрат на хмарну інфраструктуру та зниження затримок обробки даних. А результати теоретичних та практичних досліджень впроваджено у навчальний процес кафедр математичних проблем управління і кібернетики та програмного забезпечення комп'ютерних систем Чернівецького національного університету імені Юрія Федьковича.

Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності.

За своїм змістом дисертаційна робота здобувача Кириченка О.О. повністю відповідає Стандарту вищої освіти зі спеціальності 121 – Інженерія програмного забезпечення та напрямам досліджень відповідно до освітньої програми «Інженерія програмного забезпечення».

Представлена дисертація є цілісним і завершеним науковим дослідженням, яке демонструє вагомий особистий внесок автора у розвиток інженерії програмного забезпечення, зокрема в частині оптимізації безсерверних обчислень та моделювання неоднорідних процесів у хмарних середовищах.

Аналіз звіту з перевірки рукопису на наявність текстових збігів та ознак академічного плагіату дозволяє стверджувати, що дисертаційна робота Кириченка Олександра Олексійовича виконана здобувачем повністю самостійно. У праці відсутні прояви фабрикації, фальсифікації, неправомірних запозичень чи компіляції. Усі наукові ідеї, концепції, фактичні дані та текстові фрагменти, що належать іншим дослідникам, супроводжуються коректними й належно оформленими посиланнями на першоджерела.

Мова та стиль викладення результатів.

Дисертаційна робота написана українською мовою.

Робота характеризується належним рівнем наукового викладу та логічною структурою подання матеріалу: спочатку наведено теоретичні засади дослідження безсерверних обчислювальних систем та систем масового обслуговування із застосуванням апарату напівмарковських випадкових еволюцій, далі – авторські

математичні та алгоритмічні розробки для проактивного моделювання неоднорідних потоків і оцінки відхилених завдань, потім – архітектурні рішення щодо розподіленого хмарного фреймворку та, нарешті, експериментальна перевірка й імітаційне моделювання ефективності запропонованих підходів. Мова викладення є грамотною, технічно точною, з доречним використанням термінів, характерних для інженерії програмного забезпечення, теорії масового обслуговування та хмарних обчислень (Serverless). Теоретична частина супроводжується обґрунтованими схемами, формулами і прикладами, що суттєво підвищує доступність викладеного матеріалу. Робота легко сприймається, незважаючи на складність тематики.

Дисертація складається зі вступу, 4 розділів, висновків, списку використаних джерел та додатків. Загальний обсяг дисертації становить 245 сторінок, з яких 142 сторінки основного тексту, 26 сторінок – список використаних джерел та 47 сторінок – додатки.

У вступі автор обґрунтовує актуальність обраної теми в контексті сучасного стану розвитку безсерверних обчислень (Serverless computing) та хмарних інформаційних систем. Чітко сформульовано об'єкт, предмет, мету і завдання дослідження, обрані методи та інструменти, а також викладено елементи наукової новизни й практичного значення.

Перший розділ присвячено концептуалізації предметного поля дослідження. У ньому здійснюється ретельний критичний аналіз науково-практичних джерел у сегменті Serverless-рішень. Досліджено специфіку функціонування сервісів класів FaaS та BaaS, принципи побудови систем, що реагують на події, та існуючі підходи до динамічного виділення хмарних потужностей. Автор фокусує увагу на ключових деструктивних чинниках технології – проблемі «холодного старту», ризиках технологічної прив'язки до одного провайдера та дефіциті ефективних засобів наскрізного моніторингу. Підсумком розділу є класифікація метрик

продуктивності та виокремлення коло колізій, розв'язання яких потребує залучення нових методів оптимізації.

Другий розділ фокусується на системно-архітектурному проектуванні. Обґрунтовано високу ефективність подієво-орієнтованого керування (EDA), що забезпечує ізоляцію функцій, слабку зв'язаність та фінансову оптимізацію системи. На основі компаративного аналізу комунікаційних патернів («Запит-Відповідь», «Публікація/Підписка», шлюзи) доведено доцільність застосування черг повідомлень для згладжування пікових навантажень. Також порівняно технології обміну даними в реальному часі AWS AppSync та WebSockets через API Gateway. Окремо типізовано вади стандартних реактивних стратегій виділення ресурсів, які через пороговий характер роботи провокують масові «холодні старты». Це дозволило автору висунути наукову гіпотезу про необхідність переходу до проактивного прогнозування трафіку. Підсумком розділу стала розробка оригінальної багаторівневої структури хмарного фреймворку, в якій логічно розмежовано процеси імпорту, обробки, телеметрії та прогнозного масштабування ресурсів, а також сформульовано систему функціональних і якісних вимог до нього.

У третьому розділі розроблено математичну модель безсерверного середовища як гетерогенної СМО зі змішаними режимами роботи. Завдяки апарату напівмарковських випадкових еволюцій доведено умови стабільності черги та граничні закони її зміни у схемах усереднення й дифузійної апроксимації через рівняння Іто. Для ідентифікації параметрів вхідного трафіку адаптовано EM-алгоритм під метрики AWS CloudWatch і SQS. На основі теорії масового обслуговування побудовано оригінальний аналітичний алгоритм оптимізації конфігурації платформи, який генерує оптимальні параметри у явному вигляді й усуває потребу в навчанні нейромереж.

Четвертий розділ демонструє емпіричну верифікацію створеного програмного інструментарію та аналітичних моделей на базі хмарної

інфраструктури AWS через три серії експериментів. У першому тесті визначено можливість нейромережевого підходу DeepAR, який на масиві з понад 1 млн подій знизив кількість «холодних стартів» на 27%, проте виявив обмеження через високу чутливість до обсягів історичних даних. Другий експеримент підтвердив переваги авторського проактивного методу на основі напівмарковських процесів, що забезпечив прискорення обробки на 25,8% та падіння рівня «холодних стартів» до 3% при оптимізації вартості конфігурації функцій. Третій етап, реалізований як стохастичне моделювання методом Монте-Карло, довів валідність моделі суміші потоків і переваги оптимального протоколу розподілу завдань (мінімізація Error Rate до 0,01881). Зіставлення результатів наочно продемонструвало, що розроблена аналітична модель працює стабільніше за методи машинного навчання та надійно гарантує нормативні критерії SLA.

У висновках підсумовано результати дослідження; додатки містять акти впровадження, список опублікованих праць автора, лістинг коду основних модулів програмного забезпечення.

Оприлюднення результатів дисертаційної роботи.

Наукові результати дисертації висвітлені у 11 наукових публікаціях здобувача, серед яких: 2 статті у виданнях, у наукометричній базі Scopus, 2 статті у наукових виданнях, включених до переліку наукових фахових видань України, а також 7 публікацій у матеріалах міжнародних та всеукраїнських науково-практичних конференцій.

Результати дисертації були апробовані на 7 наукових конференціях, серед яких міжнародні: «Проблеми інформатики та комп'ютерної техніки (ПІКТ)» (Чернівці, 2023-2024), «Problems of Science and Technology: the Search for Innovative Solutions» (Мюнхен, Німеччина, 2024); 13th International Conference on Electronics, Communications and Computing's (IC ECCO) (Кишинів, Молдова, 2024); «Trends and Prospects for the Development of Science and Education» (Оксфорд, Велика Британія, 2025); 6th International Workshop on Intelligent

Information Technologies & Systems of Information Security (IntelITSIS 2025) (Хмельницький, 2025); «Scientific Progress: Theories, Applications and Global Impact» (Брага, Португалія, 2026).

Наукові результати, описані в дисертаційній роботі, повністю висвітлені у наукових публікаціях здобувача. В усіх публікаціях дотримано принципів академічної доброчесності. Особистий внесок автора у спільних роботах чітко простежується та повністю відповідає результатам, зарахованим за темою дисертаційного дослідження.

Недоліки та зауваження до дисертаційної роботи.

1. У розділі 1, присвяченому аналізу теоретичних основ безсерверних обчислень та огляду літератури, наведено ґрунтовний опис моделей FaaS та BaaS. Однак у тексті відсутня чітка та наочна систематизація наявних методів прогнозування навантаження. Доцільно було б узагальнити переваги, недоліки та сфери застосування розглянутих підходів у вигляді порівняльної таблиці або концептуальної класифікаційної схеми, що дозволило б чіткіше та переконливіше обґрунтувати вибір математичного апарату теорії черг для оптимізації безсерверних систем.

2. Під час порівняння двох підходів до реалізації комунікації в реальному часі (AWS AppSync та WebSockets через API Gateway) у розділі 2 подано детальні якісні критерії їх функціонування. Проте автором не наведено детального аналізу чутливості та фінансової затратності обох рішень при різко мінливих сценаріях інтенсивності запитів (наприклад, при стрибку від нуля до десятків тисяч з'єднань), що могло б посилити інженерну обґрунтованість вибору архітектури.

3. Експериментальні дослідження у розділі 4 виконано на базі реальних наборів даних хмарної платформи AWS, що містять інформацію про транзакції та події. Проте для підтвердження універсальності запропонованої математичної моделі суміші потоків та проактивного алгоритму масштабування на основі

напівмарковських процесів доцільно було б розширити експериментальну базу за рахунок трафіку з інших доменних областей, наприклад, медіа-стрімінгу або телеметрії індустріальних сенсорів IoT.

4. У роботі недостатньо уваги приділено аналізу обчислювальної складності параметричної оцінки на основі EM-алгоритму під час обробки метрик у реальному часі. Оцінка асимптотичної складності розробленого восьмикрокового алгоритму за умов обробки великих масивів метрик з AWS CloudWatch дозволила б точніше визначити затримки, які створює сам оптимізаційний фреймворк.

5. У дисертаційній роботі спостерігається певна нерівномірність між теоретичною та експериментальною частинами: при фундаментальному та глибокому математичному обґрунтуванні моделей на основі напівмарковських випадкових еволюцій і стохастичних диференціальних рівнянь Іто експериментальна валідація у розділі 4 могла б бути більш розширеною та різноплановою. Зокрема, проведення серій експериментів на базі інших популярних хмарних провайдерів підвищило б рівень універсальності та практичної цінності зроблених висновків.

Вважаю, що висловлені зауваження не є визначальними і не зменшують загальну наукову новизну та практичну значимість результатів та не впливають на позитивну оцінку дисертаційної роботи.

Висновок про дисертаційну роботу.

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії Кириченка Олександра Олексійовича на тему «Оптимізація безсерверних обчислень у хмарних середовищах» виконана на належному науковому рівні, не порушує принципів академічної доброчесності та є закінченим науковим дослідженням, сукупність теоретичних та практичних результатів якого розв'язує наукове завдання, що має істотне значення для інформаційних технологій. Дисертаційна робота за актуальністю, практичною цінністю та науковою

новизною повністю відповідає вимогам чинного законодавства України, що передбачені в п. 6, 7, 8, 9 «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого Постановою Кабінету Міністрів України від 12 січня 2022 р. № 44.

Здобувач Кириченко Олександр Олексійович заслуговує на присудження ступеня доктора філософії в галузі знань 12 – Інформаційні технології за спеціальністю 121 – Інженерія програмного забезпечення.

Рецензент:

завідувач кафедри програмного забезпечення
комп'ютерних систем
Чернівецького національного університету
імені Юрія Федьковича,
доктор філософії, доцент

Катерина ГАЗДЮК

Підпис *Газдюк К.* засвідчую
Ученій секретарі Чернівецького національного
університету імені Юрія Федьковича
Липовська Н.
" 29 " травня 2026

