

ВІДГУК ОФІЦІЙНОГО ОПОНЕНТА

доктора технічних наук Гребеніка Ігоря Валерійовича на дисертаційну роботу Жолткевич Галини Григоріївни “Математичне моделювання процесів реплікації даних у розподілених сховищах”, що представлена на здобуття наукового ступеня кандидата технічних наук за спеціальністю 01.05.02 — математичне моделювання та обчислювальний методи (технічні науки)

Актуальність теми дисертації. Зважаючи на проникнення інформаційних технологій майже у всі галузі сучасної техніки, забезпечення продуктивності і коректності функціонування програмних систем є одним з головних викликів для сучасної інформатики і програмної інженерії. Перетворення обчислювальної системи на розподілену є наразі єдиним шляхом подальшого підвищення її продуктивності, враховуючи досягнення на сьогодні фізичної граници швидкодії процесорів. Проте добре відомо, що такий підхід призводить до суттєвого ускладнення процесів проєктування обчислювальних і, зокрема, інформаційних систем. Одним із джерел цього ускладнення є або цілковита неможливість, або суттєва обмеженість використання методів динамічного аналізу (тестування і симуляції) програмних артефактів, які, як правило, використовуються в програмній індустрії. В такій ситуації статичний аналіз, який базується на побудові математичних моделей програмних артефактів для подальшого їх дослідження математичними методами, стає основним методом оцінки коректності і продуктивності при розробці розподілених систем. Надзвичайно важливим компонентоможної сучасної великої системи є сховище (або база) даних, з якими операє система, причому воно також має бути розподіленим. Після робіт Е. Брюера, Н. Лінч і С. Гілберта у першому десятиріччі нашого сторіччя стало зрозуміло, що технологія централізованого зберігання даних суттєво відрізняється від технології розподіленого зберігання за причин принципового, а не технічного характеру. Внаслідок цих робіт управління реплікацією даних стало однією з фундаментальних проблем сучасної інформатики, яка і досі остаточно не вирішена. Зважаючи на викладене вище вважаю, що дисертаційна робота Жолткевич Г.Г. “Математичне моделювання процесів реплікації даних у

розділених сховищах” присвячена актуальній науково-технічній задачі забезпечення контролюваного рівня узгодженості реплік даних в розподіленому сховищі даних (РСД) при необхідних рівнях доступності даних і стійкості сховища даних до розділення на етапі проектування розподіленої системи задля управління ризиками некоректного функціонування системи, яка розв’язується авторкою шляхом побудови комплексу математичних моделей об’єкту дослідження та розробки обчислювальних методів аналізу цих моделей.

Зв’язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалася на кафедрі теоретичної та прикладної інформатики факультету математики та інформатики Харківського національного університету імені В.Н. Каразіна відповідно до завдань прикладної НДР «Моделі інформаційних процесів та методи їх обробки» ДР № 0116U003141.

Ступінь обґрунтованості і достовірності наукових положень, висновків і рекомендацій, сформульованих у дисертації. Обґрунтованість викладених у дисертаційній роботі положень базується на коректних постановках мети та завдань дослідження, що підтверджується: використаними методами теорії графів та теорії ймовірностей, розробленими програмними засобами для імітаційного моделювання розподілених сховищ даних з використанням мови UML, результатами проведених імітаційних експериментів, достатньою кількістю публікацій у наукових фахових виданнях України, міжнародних виданнях, апробацією результатів дослідження на міжнародних наукових конференціях.

Оцінка змісту дисертації та її завершеність. Матеріал дисертації викладено логічно й завершено. Дисертація включає анотації українською та англійською мовами, вступ, чотири розділи, висновки, список використаних джерел з 102 найменувань на 11 сторінках та 7 додатків на 13 сторінках, а також містить 54 рисунки. Загальний обсяг роботи складає 138 сторінок, з них 110 сторінок основного тексту.

Структура дисертації добре та логічно скомпонована, робота викладена у науковому стилі і оформлена відповідно діючих вимог.

Зміст дисертації відповідає паспорту спеціальності 01.05.02 — математичне моделювання та обчислювальні методи (технічні науки).

Матеріали дисертації достатньо повно викладені у 8 наукових працях, три з яких — у наукових фахових виданнях України, 1 — в іноземному спеціалізованому виданні (як індексовано в наукометричній базі «Scopus») і 4 тези доповідей на фахових міжнародних наукових конференціях.

Автореферат точно відповідає змістові дисертації, причому його зміст дає повне уявлення про наукову цінність і практичну значущість дослідження.

У вступі наведено загальну характеристику роботи: актуальність теми та аналіз пов'язаних з темою робіт, предмет, мета і задачі дослідження, об'єкт дослідження, наукова новизна та практичне значення отриманих результатів, а також наведений список публікацій за темою дисертаційної роботи.

У першому розділі проведений аналіз сучасних протоколів поширення реплік, забезпечення консенсусу, механізмів забезпечення CAP-гарантій даних та сучасних реалізацій реплікації у розподілених базах даних. Розглянуто розподілені бази даних, вказано їх актуальність та класифікацію, зроблено опис основних парадигм в архітектурі таких баз даних, описано модель якості баз даних та характеристики і принципи їх надійності, проаналізовано механізми внутрішньої роботи обробки запитів до БД. Також проведений порівняльний аналіз баз даних з відкритим кодом та їх архітектури. В розділі також сформульовані CAP-гарантії та CAP-теорему й подані результати порівняння рішень, що імплементовані в різних реалізаціях РСД для забезпечення тих чи інших CAP-гарантій. Це дозволило авторці показати відсутність структурованого стандартизованого рішення, що могло б послужити механізмом регулювання CAP-гарантій та механізмом прийняття рішень, що дасть можливість знайти баланс між гарантіями, залежачи від конкретної специфікації потреб системи РСД, що і стало підставою для даного дослідження.

Другий розділ присвячений побудові комплексу математичних моделей для опису мережі розподіленої системи, централізованої моделі зберігання даних та її узагальнення для випадку розподіленого сховища даних. Модель мережі крім

графа мережі містить додатково характеристики вузлів і ребер графа, що забезпечує можливість подальшого моделювання відмов вузлів, оцінювання ймовірності доставки повідомлень мережею та вартість такої доставки. Моделі централізованого і розподіленого сховищ даних включають механізми виконання транзакцій читання та запису даних. Описано також метод реплікації даних у РСД у вигляді відповідного алгоритму.

У третьому розділі введені ймовірнісні оцінки рівня виконання САР-гарантій. На мою думку, в цьому як раз і полягає основна новація роботи — замість забезпечення виконання всіх САР-гарантій, що неможливо з огляду на САР-теорему, розробити математичний і програмний інструментарій оцінювання рівня виконання цих гарантій і розробка методів їх збалансування. Таким чином, поставлена задача вирішується через забезпечення контролюваного відхилення рівнів виконання гарантій від ситуації повного задоволення вимог шляхом балансування цих рівнів за допомогою розробленого і описаного в цьому розділі методу.

Четвертий розділ містить опис програмних моделей, які реалізують комплекс експериментальних досліджень, а також аналіз результатів цих досліджень.

Нові наукові результати дисертаційної роботи. У дисертаційній роботі отримані наступні нові наукові результати:

- зазнала подальшого розвитку стохастична модель розподіленої системи за рахунок забезпечення формалізації виконання САР-гарантій, яка, на відміну від існуючих моделей, об'єднує формалізовані визначення для САР-гарантій в одну цілісну структуру і дає підставу для оцінки гарантій на іншому рівні;
- вперше запропоновано систему статистичних метрик для оцінювання виконання САР-гарантій у РСД, що базується на формалізації визначень САР-гарантій, і замість бінарної оцінки гарантій РСД пропонується оцінювати систему гарантій за допомогою ймовірнісних величин, що дає підставу для розвитку метрик для оцінки рівня САР-гарантій;
- удосконалено методи обчислення ймовірностей втрати даних на основі запропонованої моделі, що дає можливість оцінки САР-гарантій і формування

метрик для оцінки з більшою точністю, з врахуванням реальних умов таких систем, і, таким чином, це дає підставу для формування методу регулювання САР-характеристик на рівні, потрібному конкретній розподіленій системі;

– зазнали подальшого розвитку методи забезпечення САР-вимог шляхом розробки нової стратегії контролюваного балансування узгодженості у формі спеціального методу контролюваного балансування узгодженості, що може бути інтегрований з методами реплікації у РСД.

Практична цінність отриманих результатів. На основі запропонованих методів та моделей реалізовано метод регулювання САР-гарантій у формі алгоритму для покриття такого рівня САР-гарантій, які будуть необхідні конкретному РСД. Цей метод може бути застосований на етапі проектування системи РСД для запобігання помилок і неоптимальних рішень побудови розподіленої мережі, а також невірного вибору або застосування методів забезпечення САР-гарантій, тобто зменшення ризиків, які коштуватимуть дорожче на етапі застосування такої системи. Також такий метод містить елементи, які мають працювати на етапі застосування системи, а саме регулювання рівня узгодженості з метою утримування її на необхідному рівні. Розроблені в дисертаційній роботі моделі, метрики та методи доведені до прототипу бібліотеки утиліт, які можуть бути використані як компоненти систем моніторингу РСД з метою контролювання САР-гарантій і оперативного реагування на інциденти в системі. Тож, такі методи використовуються для підвищення ефективності роботи РСД, мережі в яких налічують сотні і тисячі вузлів.

Результати дисертаційної роботи були використані при підготовці відповідних спеціальних освітніх компонентів за двома програмами у Запорізькому національному і Херсонському державному університетах.

Зауваження щодо дисертаційної роботи. Звичайно, робота не вільна від певних недоліків.

1. Перший підрозділ другого розділу присвячений дуже важливій для розподілених систем проблемі — проблемі відсутності глобального часу, що

призводить до експоненційного зростання складності аналізу процесів реплікації, і хоча в підрозділі наведений дуже цікавий математичний результат — узагальнення на цей випадок відомої в теорії впорядкованих множин теореми Шпільрайна про продовження — в подальшому цей результат не використовується, окрім прояснення унікальності моделі годинників Лемпорта.

2. Запропонований авторкою метод балансування дуже подібний до ідей фон Неймана, використаних ним у теорії ігор — у випадку відсутності класичного розв'язку гри пропонується розглядати її стохастичне узагальнення шляхом введення так званих змішаних стратегій. Було б цікаво продовжити цю аналогію з метою з'ясування можливості використання інших результатів, відомих з теорії ігор.

3. Як базову модель сховища даних авторка розглядає модель словника. Звичайно, ця модель достатньо загальна і використовується у noSQL базах даних, проте цікавим було б розглянути більш структуровану модель даних наприклад, реляційну.

4. В роботі вказано на можливість імплементації запропонованих моделей і методів у системи управління РСД з відкритим кодом, на жаль авторка не наводить даних про реалізацію такої імплементації.

Крім цього слід зазначити, що робота містить стилістично невдалі вирази, вузько професійний жаргон (наприклад, слово “метрика” вживається не у загально прийнятому математичному значенні, а у значенні показника чи індикатора, як це зазвичай використовується фахівцями з програмної інженерії).

Загальний висновок. Вважаю що дисертація Жолткевич Г.Г. “Математичне моделювання процесів реплікації даних у розподілених сховищах” є завершеною науковою роботою, в якій вирішується актуальна науково-технічна задача забезпечення контролюваного рівня узгодженості реплік даних в розподіленому сховищі даних при необхідних рівнях доступності даних і стійкості сховища даних до розділення.

Дисертація відповідає паспорту спеціальності 01.05.02 — математичне моделювання та обчислювальні методи (технічні науки) та всім вимогам до кандидатських дисертацій згідно з пп. 9, 11, 12 Порядку присудження наукових ступенів, затвердженого Постановою Кабінету міністрів України 24.07.2013 року № 567, а її авторка Жолткевич Галина Григоріївна заслуговує на присудження наукового ступеня кандидата технічних наук за спеціальністю 01.05.02 — математичне моделювання та обчислювальні методи (технічні науки).

Офіційний опонент:

доктор технічних наук, професор,
завідувач кафедри системотехніки
Харківського національного
університету радіоелектроніки

I. В. Гребенник

“15” 03 2021 року

Підпис проф. І.В. Гребенника засвідчує.

Перший проректор Харківського національного
університету радіоелектроніки



I.B. Рубан