



Д. Є. Фролов

Харківський національний університет радіоелектроніки, м. Харків, Україна

МЕТОД МІЖРІВНЕВОГО ДИНАМІЧНОГО ПЕРЕРОЗПОДІЛУ НАВАНТАЖЕННЯ У БАГАТОРІВНЕВИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

Виявлено, що в багаторівневих інформаційних системах (БІС) нерівномірний розподіл навантаження між рівнями призводить до зростання тривалості завершення процесів навіть за наявності незайнятих обчислювальних ресурсів на інших рівнях. Встановлено, що за таких умов доцільним є тимчасове переміщення процесу на інший рівень з подальшим виконанням на ньому і поверненням на рівень, на якому процес виконувався до тимчасового перенесення (попередній рівень), якщо короткочасне припинення не порушує виконання основної функції системи, а сумарні часові витрати на збереження стану, переміщення, виконання та повернення є меншими за тривалість завершення процесу на попередньому рівні. З'ясовано, що наявні підходи до динамічного розміщення мікрослужб, перенесення контейнерів, міграції віртуальних машин і прямого надання додаткових обчислювальних ресурсів не формалізують задачу тимчасового переміщення саме процесу між рівнями з обов'язковим поверненням на попередній рівень. Охарактеризовано формалізований опис рівнів, процесів, показників їх зайнятості, умов допустимості тимчасового перенесення та правил вибору рівня-донора за критерієм мінімізації сумарної тривалості завершення процесу. Удосконалено метод динамічного перерозподілу навантаження між рівнями, який передбачає виявлення рівнів з надмірною зайнятістю, відбір процесів, придатних до тимчасового перенесення, визначення скорочення тривалості завершення процесу, вибір рівня-донора, виконання процесу на ньому, повернення на попередній рівень і повторне оцінювання стану системи. Встановлено межі застосування удосконаленого методу для процесів, що допускають короткочасне припинення, мають хоча б один допустимий рівень-донор і забезпечують скорочення тривалості завершення процесу порівняно з продовженням виконання на попередньому рівні. Оцінено вплив запропонованого підходу на часові показники виконання процесів за результатами імітаційного моделювання. Встановлено, що удосконалений метод забезпечує найменшу середню тривалість завершення процесів, найбільшу їх частку, завершених у допустимий проміжок часу, та вищу стійкість часових показників за нерівномірного розподілу навантаження між рівнями.

Ключові слова: відмовостійкість інформаційної системи; обчислювальний ресурс; переміщення процесу; імітаційне моделювання; динамічне розміщення мікрослужб.

Вступ / Introduction

Сучасні багаторівневі інформаційні системи (БІС) функціонують в умовах нерівномірного розподілу навантаження між рівнями та часових обмежень на виконання процесів [3, 14]. За таких умов на одних рівнях у певні проміжки часу можуть залишатися незадіяні обчислювальні ресурси, тоді як на інших виконання процесів стає довшим через їх недостатність. Наслідком цього є зростання тривалості завершення процесів і ускладнення своєчасного виконання основної функції системи.

У відомих підходах до організації функціонування БІС поширеними є підходи з фіксованою багаторівневою організацією процесів і ресурсів [17] та підходи з перебудовою структури системи і перерозподілом функцій [6]. У роботі [7] розглянуто підхід, за якого окремий процес тимчасово переміщується на інший рівень із незадіяними обчислювальними ресурсами, виконується на ньому та після цього повертається на попередній рівень. Такий підхід відрізняється від прямого надання додаткових ресурсів процесу тим, що передба-

чає зміну місця його тимчасового виконання. За його застосування потрібно враховувати тривалість збереження стану, переміщення, поновлення виконання та повернення процесу, а також не допускати тривалого зайняття ресурсів рівня-донора. Водночас, наявні рішення переважно орієнтовані на задачі загального планування для вирівнювання навантаження у паралельних і розподілених системах на основі міграції віртуальних машин між більш і менш завантаженими вузлами. [9].

Обмеження методу, запропонованого роботі [7], полягає в тому, що він не може застосовуватися до всіх процесів без винятку. У БІС існує клас процесів, для яких навіть короткочасне припинення є недопустимим, оскільки це може порушити виконання основної функції системи. Тому удосконалення відомого методу доцільно пов'язати з визначенням умов його застосування до процесів, для яких короткочасне припинення, збереження стану, переміщення, поновлення виконання та повернення на попередній рівень є допустимими, а очікуване скорочення тривалості завершення процесу перевищує сумарні витрати часу на ці дії. У такій постановці завдання виникає потреба удосконалення методу

© Copyright 2026 by the author(s)

Фролов Дмитро Євгенович, аспірант, асистент, кафедра електронних обчислювальних машин.

Email: dmytro.frolov@nure.ua; <https://orcid.org/0009-0009-3291-3561>

Цитування за ДСТУ: Фролов Д. Є. Метод міжрівневого динамічного перерозподілу навантаження у багаторівневих інформаційних системах. Науковий вісник НЛТУ України. 2026, т. 36, № 2. С. 213–221.

Citation APA: Frolov, D. Ye. (2026). Inter-level dynamic load redistribution method in multi-level information systems. *Scientific Bulletin of UNFU*, 36(2), 213–221. <https://doi.org/10.36930/40360223>

[7] міжрівневого динамічного перерозподілу навантаження в БІС.

Об'єкт дослідження – функціонування БІС за нерівномірного розподілу навантаження між рівнями та змінної зайнятості обчислювальних ресурсів.

Предмет дослідження – методи і засоби організації міжрівневого перерозподілу навантаження у БІС, які дадуть змогу зменшити тривалість завершення процесів і підвищити частку процесів, завершених у допустимий проміжок часу.

Мета роботи – удосконалити метод динамічного перерозподілу навантаження між рівнями БІС шляхом введення правил тимчасового переміщення процесу на інший рівень з подальшим поверненням на попередній рівень, на якому він виконувався до перенесення, з урахуванням часових витрат на таке переміщення та умов допустимості його застосування для визначеного класу процесів, що дасть змогу зменшити тривалість завершення процесів і підвищити частку процесів, завершених у допустимий проміжок часу.

Для досягнення зазначеної мети визначено такі основні завдання дослідження:

- проаналізувати особливості розподілу навантаження між рівнями БІС та формалізувати умови, за яких тимчасове переміщення процесу є доцільним, що дасть змогу виявити ті процеси, для яких таке переміщення забезпечує скорочення тривалості їх завершення без порушення виконання основної функції системи;
- розробити формалізований опис рівнів, процесів, їх зайнятості та правил вибору рівня для тимчасового виконання процесу, що дасть змогу визначати допустимі рівні-донори та обирати рівень з мінімальною сумарною тривалістю завершення процесу;
- удосконалити метод динамічного перерозподілу навантаження між рівнями БІС, передбачивши виявлення процесів, придатних до тимчасового переміщення, вибір рівня-донора, виконання процесу на ньому та його повернення на попередній рівень, що забезпечить зменшення середньої тривалості завершення процесів за нерівномірного розподілу навантаження між рівнями;
- визначити обмеження застосовності удосконаленого методу для процесів, тимчасове переміщення яких не порушує виконання основної функції БІС, що дасть змогу відокремити клас процесів, для яких таке переміщення є допустимим і доцільним;
- виконати порівняльне оцінювання удосконаленого методу за часовими показниками виконання процесів, що покаже його переваги порівняно з іншими методами даної предметної області.

Аналіз останніх досліджень та публікацій. У роботі [7] розглянуто задачу динамічного розміщення мікрослужб у багаторівневого туманного середовищі. Запропонований підхід ґрунтується на трирівневій організації виконання, у межах якої здійснюються впорядкування мікрослужб за пріоритетом, оцінювання доступних ресурсів вузлів, урахування залежностей між мікрослужбами, перенаправлення запитів та масштабування екземплярів мікрослужб. Наведені результати свідчать про підвищення ефективності використання ресурсів і скорочення тривалості виконання застосунків. Водночас, ця робота орієнтована на розміщення й перерозподіл мікрослужб у туманному середовищі, тоді як умови допустимості короткотривалого припинення окремого процесу, його тимчасового міжрівневого переміщення та обов'язкового повернення на попередній рівень у ній не формалізовано.

У роботі [18] подано узагальнений огляд алгоритмів розміщення сервісів в інтегрованих середовищах хмарних, туманних і периферійних обчислень. Особливість цієї роботи – у систематизації наявних рішень та побудові класифікації, у межах якої підходи поділено на графові, евристичні та такі, що спираються на навчання за даними. Ця робота має оглядовий характер. Її важливість полягає у фіксації класів суміжних рішень у предметній області, що вказує на актуальність саме цього дослідження.

У роботі [1] наведено результати дослідження динамічного планування для вирівнювання навантаження у паралельних і розподілених системах на основі міграції віртуальних машин між більш і менш завантаженими вузлами. Запропонований підхід охоплює виявлення перевантаження, прийняття рішення щодо перенесення, прогнозування зміни навантаження та подальше перепланування задач. Його перевагою є комплексний розгляд часових показників і поєднання кількох взаємопов'язаних процедур керування. Проте, предметом цього дослідження є перенесення віртуальних машин у хмарному середовищі, а не тимчасове переміщення окремого процесу між рівнями БІС. Окрім цього, у цій роботі не розглядають формалізації, за якої процес після тимчасового виконання на іншому рівні має бути повернений на свій попередній рівень як обов'язкова умова застосування методу.

У дослідженні [4] запропоновано прикладне рішення у вигляді фреймворку для динамічної міграції контейнерів у мобільних периферійних хмарах. Підхід ґрунтується на багаторівневій архітектурі обчислень, моніторингу завантаження вузлів за показниками процесорного часу та оперативної пам'яті, а також на виборі вузла призначення для перенесення контейнера за поточного перевантаження. Особливістю результатів цього досліджень є врахування динаміки навантаження та використання механізмів динамічної міграції для підтримання працездатності сервісів у розподіленому середовищі. Проте, в цій роботі міграцію розглянуто як односпрямоване перенесення контейнера, тоді як умови тимчасового міжрівневого переміщення окремого процесу з обов'язковим поверненням на попередній рівень у ній не формалізовано. Окрім цього, цей підхід супроводжується істотними витратами на передавання стану контейнера.

У роботі [2] вирішено прикладну задачу оптимального перенесення контейнерів у гібридному середовищі кордонних обчислень. Запропонований підхід передбачає міжрівневу оркестрацію контейнерів, у межах якої рішення про їх перенесення приймають з урахуванням доступних обчислювальних ресурсів та затримки доступу до обчислювальних ресурсів. Дотичним до нашого дослідження в цій роботі є формалізований розгляд міжрівневого перерозподілу навантаження в гібридному середовищі та врахування ресурсних обмежень під час вибору нового вузла виконання. Однак запропонованому рішенню характерна висока обчислювальна складність, що обмежує їх придатність для оперативного тимчасового переміщення процесу на інший рівень з подальшим поверненням на попередній рівень.

У роботі [15] наведено результати досліджень у галузі динамічного керування ресурсами у високопродуктивних обчислювальних системах на основі асинхронної багатозадачної моделі та механізму перехоплення

завдань між вузлами. Підхід орієнтований на зменшення неповного завантаження обчислювальних вузлів і підвищення загальної ефективності використання ресурсів шляхом динамічного перерозподілу обчислювальних завдань під час виконання. Однак процес переміщення завдань має характер остаточного перерозподілу в межах єдиного обчислювального середовища, тоді як завдання тимчасового міжрівневого переміщення окремого процесу між рівнями БІС з обов'язковим поверненням на попередній рівень не розглядають.

Отже, аналіз публікацій [1, 2, 4, 7, 15, 18] показав, що сучасні дослідження охоплюють динамічне розміщення мікрослужб, класифікацію алгоритмів розміщення сервісів, перенесення контейнерів у гібридних і периферійних середовищах, а також динамічний перерозподіл обчислювальних задач у високопродуктивних системах. Проте, поза увагою розглянутих підходів залишається завдання тимчасового міжрівневого переміщення саме процесу в БІС з його виконанням на іншому рівні та обов'язковим поверненням на попередній рівень. Недостатньо опрацьованими залишаються також питання визначення класу процесів, для яких таке переміщення є допустимим, вибору рівня-донора з урахуванням сумарної тривалості збереження стану, переміщення, виконання та повернення, а також обмеження тривалості зайняття його обчислювальних ресурсів. Саме це зумовлює доцільність удосконалення методу [7] для умов, визначених у цьому дослідженні.

Матеріали та методи дослідження. Матеріалами дослідження є БІС, у яких процеси розподілені між рівнями, а під час функціонування виникають стани нерівномірного розподілу навантаження, за яких на одних рівнях спостерігається перевантаження, тоді як на інших зберігаються незадіяні обчислювальні ресурси. У роботі використано методи системного аналізу багаторівневої організації інформаційних систем, формалізації міжрівневого перерозподілу навантаження, математичного моделювання умов допустимості тимчасового переміщення процесів, побудови правил вибору рівня-донора для їх тимчасового виконання та імітаційного моделювання для оцінювання ефективності удосконаленого методу.

Результати дослідження та їх обговорення / Research results and their discussion

Аналіз особливостей розподілу навантаження між рівнями виконання процесів та умов тимчасового їх переміщення на інші рівні. У БІС навантаження між рівнями змінюється в часі. У різні моменти функціонування на одних рівнях зберігаються незадіяні обчислювальні ресурси, тоді як на інших виникає надмірна зайнятість, що призводить до зростання тривалості завершення процесів.

У цьому дослідженні аналізують стани БІС, за яких подальше виконання процесу на попередньому рівні є менш доцільним, ніж його тимчасове переміщення на інший рівень. За такої постановки навантаження визначають як характеристику зайнятості рівня виконання закріплених за ним процесів, а нерівномірний розподіл навантаження – як невідповідність між потребою процесів в обчислювальних ресурсах і фактично доступним обсягом цих ресурсів на різних рівнях.

Зростання тривалості завершення процесів зумовлюється збільшенням потоку обчислювальних запитів або

зменшенням доступних ресурсів на окремому рівні. Якщо на цьому рівні подальше виконання процесу втрачає доцільність за часовими показниками, процес має розглядатися як кандидат для тимчасового переміщення на рівень-донор, де наявні незадіяні обчислювальні ресурси, достатні для його виконання.

Специфікою постановки задачі є тимчасове переміщення процесу на рівень-донор, де наявні ресурси, достатні для його виконання. Такий підхід дає змогу уникнути тривалого зайняття ресурсів рівня-донора одним процесом і зменшити ймовірність обмеження доступу інших процесів до ресурсів попереднього рівня. Визначальною рисою запропонованого підходу є тимчасовий характер переміщення процесу з обов'язковим поверненням на попередній рівень.

Тимчасове переміщення процесу є доцільним тільки за одночасного виконання таких умов:

- процес не належить до класу процесів, для яких короткочасне припинення виконання є недопустимим;
- очікувана тривалість завершення процесу на попередньому рівні перевищує сумарні витрати часу на його збереження стану, переміщення, виконання на рівні-донорі та повернення;
- на рівні-донорі наявні незадіяні обчислювальні ресурси, достатні для виконання процесу;
- повернення процесу на попередній рівень після завершення тимчасового виконання не порушує виконання інших процесів.

Отже, доцільність тимчасового переміщення процесу визначають поєднанням чотирьох чинників: зайнятості ресурсів попереднього рівня, наявності незадіяних ресурсів на рівні-донорі, допустимості короткочасного припинення конкретного процесу та очікуваного скорочення тривалості його завершення. Саме така постановка конкретизує умови застосування удосконаленого методу [7] у напрямі тимчасового переміщення тільки тих процесів, для яких допустимими є їх короткочасне припинення, виконання на рівні-донорі та обов'язкове повернення на попередній рівень, що відмежовує його від підходів статичного закріплення процесів за рівнями [13] і прямого надання додаткових ресурсів [5].

Формалізація рівнів, процесів і правил вибору рівня для тимчасового виконання процесу. Нехай множину рівнів БІС задають у вигляді

$$L = \{l_{ij} \mid j = \overline{1, n_i}; i = \overline{1, m}\}, \quad (1)$$

де: L – множина всіх рівнів БІС; l_{ij} – j -ий рівень i -ої підсистеми; m – кількість підсистем; n_i – кількість рівнів у i -й підсистемі.

Відповідно до виразу (1), для кожного рівня l_{ij} у момент часу t задають множину процесів, що виконуються на ньому, а саме:

$$P_{ij}(t) = \{p_{ijr} \mid r = \overline{1, s_{ij}(t)}\}, j = \overline{1, n_i}; i = \overline{1, m}, \quad (2)$$

де: p_{ijr} – r -й процес на рівні l_{ij} ; $s_{ij}(t)$ – кількість процесів, що виконуються на рівні l_{ij} у момент часу t .

Для процесу p_{ijr} та рівня l_{ij} вводять тривимірні вектори ресурсних потреб процесу $\vec{a}_{ijr}(t)$ і тривимірний вектор доступних обчислювальних ресурсів рівня $\vec{b}_{ij}(t)$ (усі компоненти векторів $\vec{a}_{ijr}(t)$ та $\vec{b}_{ij}(t)$ визначаються на фіксованому інтервалі виконання Δt), а саме:

$$\vec{a}_{ijr}(t) = (a_{ijr}^e(t), a_{ijr}^m(t), a_{ijr}^d(t)), \quad \vec{b}_{ij}(t) = (b_{ij}^e(t), b_{ij}^m(t), b_{ij}^d(t)), \quad (3)$$

де: $a_{ijr}^e(t)$ – середня потреба процесу p_{ijr} у процесорному часі на інтервалі виконання $[t, t + \Delta t]$, виражена відно-

шенням потрібного процесорного часу до тривалості інтервалу Δt ; $a_{ijr}^m(t)$ – обсяг оперативної пам'яті, потрібний процесу p_{ijr} під час виконання на інтервалі $[t, t+\Delta t]$, Б; $a_{ijr}^d(t)$ – середня швидкість введення-виведення, потрібна процесу p_{ijr} на інтервалі виконання $[t, t+\Delta t]$, Б/с; $b_{ij}^c(t)$ – доступний на рівні l_{ij} процесорний час на інтервалі виконання $[t, t+\Delta t]$, виражений відношенням доступного процесорного часу до тривалості інтервалу Δt ; $b_{ij}^m(t)$ – доступний на рівні l_{ij} обсяг оперативної пам'яті на інтервалі виконання $[t, t+\Delta t]$, Б; $b_{ij}^d(t)$ – доступна на рівні l_{ij} середня швидкість введення-виведення на інтервалі виконання $[t, t+\Delta t]$, Б/с.

На основі виразів (3) показник зайнятості рівня l_{ij} визначають як найбільше з відношень сумарної потреби процесів до фактично доступних обчислювальних ресурсів рівня:

$$\rho_{ij}(t) = \max \left\{ \frac{\sum_{r=1}^{s_{ij}(t)} a_{ijr}^c(t)}{b_{ij}^c(t)}, \frac{\sum_{r=1}^{s_{ij}(t)} a_{ijr}^m(t)}{b_{ij}^m(t)}, \frac{\sum_{r=1}^{s_{ij}(t)} a_{ijr}^d(t)}{b_{ij}^d(t)} \right\}, j = \overline{1, n_i}; i = \overline{1, m}, \quad (4)$$

де: $\rho_{ij}(t)$ – показник зайнятості рівня l_{ij} ; $\sum a_{ijr}^c(t)$ – сумарна потреба всіх процесів рівня l_{ij} у процесорному часі на інтервалі виконання; $\sum a_{ijr}^m(t)$ – сумарна потреба всіх процесів рівня l_{ij} в оперативній пам'яті; $\sum a_{ijr}^d(t)$ – сумарна потреба всіх процесів рівня l_{ij} у швидкості введення-виведення.

Формулу (4) використовують для виявлення рівнів, на яких зайнятість досягає або перевищує допустимий рівень, а також рівнів, на яких у певний момент часу зберігаються незадіяні обчислювальні ресурси. Для цього вводять вектор незадіяних обчислювальних ресурсів рівня:

$$\bar{v}_{ij}(t) = \bar{b}_{ij}(t) - \sum_{r=1}^{s_{ij}(t)} \bar{a}_{ijr}(t), j = \overline{1, n_i}; i = \overline{1, m}, \quad (5)$$

де $\sum \bar{a}_{ijr}(t)$ – сумарна ресурсна потреба всіх процесів, що виконуються на рівні l_{ij} .

Віднімання у (5) виконують покомпонентно. Оскільки не всі процеси допускають навіть короткочасне припинення, для кожного з них визначають ознаку допустимості його тимчасового перенесення. Для цього вводять двійкову величину:

$$\mu_{ijr} = \begin{cases} 1, & \delta_{ijr} = 1, \tau_{ijr}^c \leq \tau_{ijr}^{\max}, \\ 0, & \delta_{ijr} = 0 \vee \tau_{ijr}^c > \tau_{ijr}^{\max}, \end{cases} \quad (6)$$

де: μ_{ijr} – ознака допустимості тимчасового перенесення процесу p_{ijr} ; δ_{ijr} – ознака наявності у процесу p_{ijr} фіксованої точки призупинення, у якій стан процесу може бути збережений без втрати коректності подальшого виконання; τ_{ijr}^c – тривалість припинення процесу p_{ijr} , потрібна для збереження його стану, переміщення та поновлення виконання на іншому рівні; $\tau_{ijr}^{\max}$ – гранично допустима тривалість призупинення процесу p_{ijr} , за якої не порушується виконання основної функції БІС.

Відповідно до (6) рівень-донор для тимчасового виконання процесу вибирають тільки з множини тих рівнів, які одночасно мають достатній незайнятий обсяг обчислювальних ресурсів, є сумісними з процесом за архітектурними та програмними умовами та забезпечують його повернення на попередній рівень у допустимий проміжок часу. Таку множину задають у вигляді

$$\Omega_{ijr}(t) = \left\{ l_{\alpha\beta} \in L \setminus \{l_{ij}\} \mid \bar{a}_{ijr}(t) \leq \bar{v}_{\alpha\beta}(t), \kappa_{ijr, \alpha\beta} = 1, \tau_{ijr, \alpha\beta}^p(t) \leq \tau_{ijr}^{\max} \right\}, \quad (7)$$

де: $l_{\alpha\beta}$ – можливий рівень-донор; $\bar{a}_{ijr}(t) \leq \bar{v}_{\alpha\beta}(t)$ – покомпонентна умова, за якої ресурсна потреба процесу p_{ijr} не перевищує незайнятий обсяг обчислювальних ресурсів рівня $l_{\alpha\beta}$; $\kappa_{ijr, \alpha\beta}$ – ознака сумісності процесу p_{ijr} з рівнем $l_{\alpha\beta}$, що враховує допустимість його виконання на цьому рівні за архітектурними та програмними умовами; $\tau_{ijr, \alpha\beta}^p(t)$ – сумарна тривалість переміщення процесу p_{ijr} на рівень $l_{\alpha\beta}$ для його тимчасового виконання та повернення на попередній рівень.

Для кожного рівня $l_{\alpha\beta} \in \Omega_{ijr}(t)$ сумарну тривалість завершення процесу після його тимчасового переміщення визначають так:

$$T_{ijr, \alpha\beta}(t) = \tau_{ijr, \alpha\beta}^{\downarrow}(t) + \tau_{ijr, \alpha\beta}^e(t) + \tau_{ijr, \alpha\beta}^{\uparrow}(t), \quad (8)$$

де: $T_{ijr, \alpha\beta}(t)$ – сумарна тривалість завершення процесу p_{ijr} у разі його тимчасового виконання на рівні $l_{\alpha\beta}$; $\tau_{ijr, \alpha\beta}^{\downarrow}(t)$ – тривалість збереження стану процесу p_{ijr} та його перенесення з рівня l_{ij} на рівень $l_{\alpha\beta}$; $\tau_{ijr, \alpha\beta}^e(t)$ – тривалість виконання процесу p_{ijr} на рівні $l_{\alpha\beta}$; $\tau_{ijr, \alpha\beta}^{\uparrow}(t)$ – тривалість повернення процесу p_{ijr} на попередній рівень l_{ij} .

На основі (8) рівень-донор для тимчасового виконання процесу обирають за правилом мінімуму сумарної тривалості завершення:

$$l_{ijr}^* = \arg \min_{l_{\alpha\beta} \in \Omega_{ijr}(t)} T_{ijr, \alpha\beta}(t). \quad (9)$$

Тимчасове переміщення процесу вважають доцільним тільки за умови:

$$T_{ijr, l^*}(t) < T_{ijr}^0(t), \quad (10)$$

де: $T_{ijr, l^*}(t)$ – мінімальна сумарна тривалість завершення процесу після тимчасового переміщення; $T_{ijr}^0(t)$ – тривалість завершення процесу без переміщення, тобто за умови його подальшого виконання на попередньому рівні l_{ij} .

Остаточне правило прийняття рішення про тимчасове перенесення процесу задають у вигляді

$$u_{ijr}(t) = \begin{cases} 1, & \mu_{ijr} = 1, \Omega_{ijr}(t) \neq \emptyset, T_{ijr, l^*}(t) < T_{ijr}^0(t), \\ 0, & \text{в інших випадках.} \end{cases} \quad (11)$$

У виразі (11) значення $u_{ijr}(t) = 1$ означає, що процес доцільно тимчасово перенести на рівень-донор; значення $u_{ijr}(t) = 0$ означає, що процес має залишитися на попередньому рівні.

Отже, наведений вище математичний апарат задає послідовну схему формалізації рівнів, процесів, їх зайнятості та правил вибору рівня-донора для тимчасового виконання процесу. Спочатку за (1) і (2) задають множини рівнів і процесів, далі за (3)-(5) визначають потреби процесів, доступні та незайняті обчислювальні ресурси рівнів, після чого за (6) встановлюють допустимість тимчасового перенесення конкретного процесу. На завершальному етапі за (7)-(11) задають множину допустимих рівнів-донорів, обчислюють сумарну тривалість завершення процесу, вибирають рівень-донор та приймають рішення щодо доцільності його тимчасового виконання на іншому рівні.

Удосконалення методу динамічного перерозподілу навантаження між рівнями. Метод [7] орієнтований на процеси, для яких короткочасне припинення, збереження стану, переміщення та поновлення виконання не

порушують основної функції системи. У межах цього дослідження його удосконалення пов'язане з формалізованим відбором процесів, придатних до тимчасового перенесення, вибором рівня-донора, виконанням процесу на ньому, поверненням на попередній рівень і повторним оцінюванням зайнятості рівнів.

На першому етапі методу виконують неперервне спостереження за показниками зайнятості рівнів, визначеними формулою (4). Для цього вводять порогове значення показника зайнятості рівня θ , досягнення або перевищення якого свідчить про недоцільність подальшого виконання процесу на попередньому рівні за часовими показниками. Тоді множину процесів, що підлягають розгляду в момент часу t , задають як

$$\Pi(t) = \{ p_{ijr}(t) \mid \rho_{ij}(t) \geq \theta, \mu_{ijr} = 1, \Omega_{ijr}(t) \neq \emptyset \}, \quad (12)$$

де θ – допустиме порогове значення показника зайнятості рівня.

Формула (12) означає, що метод не застосовують до кожного процесу безумовно. До розгляду залучають тільки ті процеси, які одночасно належать рівням з надмірною зайнятістю, допускають короткочасне припинення та мають хоча б один допустимий рівень-донор для тимчасового виконання. Отже, критичні процеси, що не допускають тимчасового перенесення, виключаються з подальшого розгляду на цьому етапі.

На другому етапі для кожного процесу з множини $\Pi(t)$ визначають скорочення тривалості завершення процесу внаслідок тимчасового перенесення. Цю величину обчислюють як різницю між тривалістю завершення процесу на попередньому рівні та мінімальної сумарної тривалості його завершення після тимчасового виконання на обраному рівні-донорі $l_{ijr}^*(t)$ (9), а саме:

$$\Delta_{ijr}(t) = T_{ijr}^0(t) - T_{ijr,l^*}(t). \quad (13)$$

Формулу (13) використовують для змістовної оцінки доцільності тимчасового перенесення процесу:

- якщо $\Delta_{ijr}(t) > 0$, то тимчасове перенесення скорочує тривалість завершення процесу;
- якщо $\Delta_{ijr}(t) \leq 0$, тимчасове перенесення не дає часової переваги.

На третьому етапі формують впорядковану послідовність процесів, які доцільно тимчасово переміщувати першочергово. Якщо кілька процесів одного або різних рівнів претендують на ті самі незайняті обчислювальні ресурси рівня-донора, перевагу надають процесу з більшим значенням $\Delta_{ijr}(t)$. Якщо значення скорочення тривалості завершення процесу збігаються, перевагу надають процесу з меншою сумарною тривалістю переміщення та повернення. Такий порядок дає змогу швидше зменшити зайнятість рівнів, що працюють з надмірним навантаженням, та обмежити тривалість зайняття незайнятих ресурсів рівня-донора одним процесом.

На четвертому етапі виконують тимчасове перенесення процесу. На попередньому рівні визначають найближчу допустиму точку припинення, в якій стан процесу може бути збережений без втрати коректності подальшого виконання. Після цього процес призупиняють, його стан зберігають і переміщують на обраний рівень-донор $l_{ijr}^*(t)$, де процес поновлюють і виконують до настання умови завершення тимчасового етапу. Таку умову задають як завершення всього процесу або завершення тієї частини оброблення, після якої попередній рівень виходить зі стану надмірної зайнятості. Отже, у

межах запропонованого методу тимчасове перенесення не є остаточною зміною місця виконання процесу, а має обмежений у часі характер.

На п'ятому етапі виконують повернення процесу на попередній рівень. Після завершення тимчасового виконання на рівні-донорі стан процесу повторно зберігають, переміщують на попередній рівень і поновлюють його виконання. Повернення є обов'язковою складовою методу, оскільки забезпечує збереження початкової прив'язки процесу до його робочого середовища та не допускає перетворення тимчасового перенесення на довготривале зайняття ресурсів рівня-донора. Отже, повернення процесу є передостаннім етапом методу.

На шостому етапі виконують повторне обчислення показників зайнятості всіх задіяних рівнів за формулами (4) і (5), після чого метод або завершує роботу, або переходить до нового циклу відбору процесів. Умову зупинення методу динамічного перерозподілу навантаження в момент часу t задають співвідношенням

$$(\forall l_{ij} \in L) \rho_{ij}(t) < \theta \vee \Pi(t) = \emptyset, \quad (14)$$

де: $(\forall l_{ij} \in L) \rho_{ij}(t) < \theta$ означає, що в БІС відсутні рівні з надмірною зайнятістю; $\Pi(t) = \emptyset$ означає, що в поточний момент часу немає процесів, які одночасно придатні до тимчасового перенесення та мають хоча б один допустимий рівень-донор.

Отже, логічний вираз (14) задає дві умови зупинення методу: або в БІС вже відсутні рівні з надмірною зайнятістю, або наявні процеси не задовольняють обмеженням допустимості тимчасового перенесення. Це унеможливило помилкове застосування методу до процесів, які мають залишатися на попередньому рівні навіть за високого значення показника зайнятості.

З огляду на зазначене вище, удосконалений метод динамічного перерозподілу навантаження між рівнями реалізує таку послідовність дій: виявлення рівнів з надмірною зайнятістю; відбір процесів, придатних до тимчасового перенесення, за формулою (12); визначення скорочення тривалості завершення процесу за формулою (13); вибір рівня-донора за введеними правилами; тимчасове виконання процесу на іншому рівні; повернення процесу на попередній рівень; повторне обчислення показників зайнятості та перевірка умови зупинення за формулою (14). На відміну від методу [7], удосконалений метод передбачає формалізований відбір процесів, для яких тимчасове перенесення є допустимим і доцільним за часовими показниками, з обов'язковим поверненням процесу на попередній рівень.

Визначення обмежень застосування методу. Запропонований метод динамічного перерозподілу навантаження між рівнями застосовують тільки до тих процесів БІС, для яких короткочасне припинення, збереження стану, тимчасове переміщення на інший рівень, поновлення виконання та повернення на попередній рівень не порушують виконання основної функції системи. Тому першим обмеженням застосування методу є виконання умови допустимості тимчасового перенесення процесу, заданої формулою (6). Якщо ця умова не виконується, процес має залишатися на попередньому рівні незалежно від значення показника зайнятості.

Другим обмеженням застосування методу є наявність хоча б одного допустимого рівня-донора для тимчасового виконання процесу. Це означає, що для процесу має існувати непорожня множина допустимих рів-

нів-донорів, задана формулою (7). Якщо жоден інший рівень не має достатнього незайнятого обсягу обчислювальних ресурсів або не забезпечує сумісності з процесом, тимчасове перенесення не виконують.

Третім обмеженням застосування методу є скорочення тривалості завершення процесу внаслідок тимчасового перенесення. Навіть якщо процес допускає перенесення, а інший рівень формально може бути використаний як рівень-донор, метод застосовують тільки тоді, коли сумарна тривалість збереження стану, переміщення, виконання та повернення є меншою за тривалість завершення процесу на попередньому рівні. Цю вимогу задають формулами (9)-(11). Отже, метод доцільно застосовувати до процесів, для яких витрати часу на тимчасове їх перенесення компенсуються скороченням загальної тривалості його завершення.

За відсутності хоча б однієї з наведених умов процес не підлягає тимчасовому перенесенню, а метод до нього не застосовують.

Порівняльне оцінювання методу за часовими показниками виконання процесів. Порівняльне оцінювання запропонованого методу виконано за двома часовими показниками: середньою тривалістю завершення процесів і часткою процесів, завершених у допустимий проміжок часу. Для зіставлення розглянуто три підходи: статичне закріплення процесів за рівнями [13], пряме надання додаткових обчислювальних ресурсів [5] та удосконалений метод. Оцінювання виконано за результатами імітаційного моделювання, проведеного на серверному обладнанні Харківського національного університету радіоелектроніки (в межах реалізації проекту Світового банку UIHERP "Удосконалення вищої освіти заради результатів" [10]). На рис. 1 наведено динаміку середньої тривалості завершення процесів, а на рис. 2 – динаміку частки процесів, завершених у допустимий проміжок часу. Зафарбовані ділянки на обох рисунках відповідають інтервалам, у межах яких у моделі штучно задавали зменшення доступного обсягу обчислювальних ресурсів окремих рівнів. Саме ці інтервали є показовими для порівняння досліджуваних підходів.

Аналіз рис. 1 засвідчує, що за статичного закріплення процесів за рівнями [13] середній час їх завершення різко зростає після входження БІС у режим нерівномірного навантаження. Уже в межах першого інтервалу зменшення доступних ресурсів значення цього показника зростає від рівня, близького до 1 ум. од., до понад 6 ум. од., а надалі досягає максимуму, близького до 11 ум. од. Така динаміка свідчить про накопичення процесів на перевантажених рівнях і підтверджує, що за відсутності міжрівневого перерозподілу навантаження система втрачає здатність своєчасно завершувати оброблення.

Метод прямого надання додаткових ресурсів (або метод прямого рахунку) [5] – це підхід до фінансового чи матеріального планування, що передбачає безпосереднє розрахування потреби в ресурсах для кожного окремого елемента, виду діяльності або об'єкта. Він використовується як у державному регулюванні, так і в управлінні підприємствами. Метод прямого надання додаткових ресурсів дає кращі результати, однак повністю не усуває наслідків перевантаження. У межах першого інтервалу зменшення доступних ресурсів середня тривалість завершення процесів зростає приблизно до 3,5 ум. од., а в межах другого – до 2,3 ум. од. Це свід-

чить про те, що пряме надання додаткових ресурсів частково зменшує на них навантаження. Проте, процеси залишаються закріпленими за попередніми рівнями і далі конкурують за отримання локальних ресурсів.

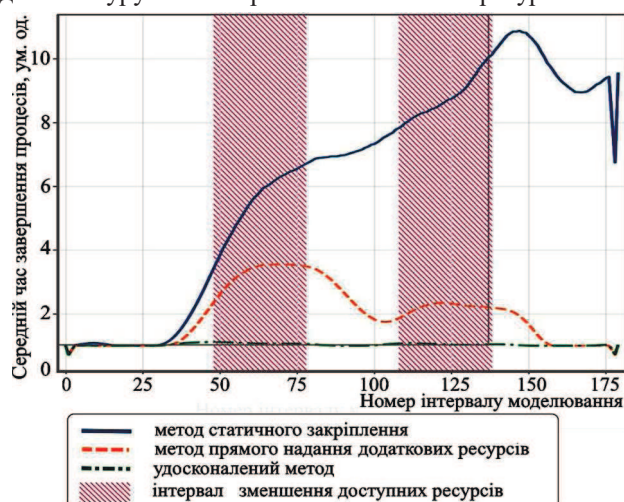


Рис. 1. Динаміка середньої тривалості завершення процесів за різних методів / Dynamics of the average process completion time for different methods

Для удосконаленого методу крива на рис. 1 упродовж основної частини інтервалу моделювання утримується поблизу 1,0-1,1 ум. од. Це означає, що тимчасове переміщення процесів на рівні-донори з незайнятими обчислювальними ресурсами дає змогу не допустити істотного накопичення затримок навіть у періоди зменшення доступних ресурсів окремих рівнів.

Отже, за показником середньої тривалості завершення процесів удосконалений метод забезпечує режим функціонування, за якого локальне перевантаження не переходить у тривале зростання тривалості завершення процесів.

Аналогічний висновок випливає з даних рис. 2. За статичного закріплення процесів за рівнями [13] частка процесів, завершених у допустимий проміжок часу, після входження в перший інтервал зменшення доступних ресурсів різко зменшується майже до нуля і надалі в межах другого інтервалу вже не відновлюється до прийнятного рівня. Це вказує на те, що накопичене відставання переводить БІС у режим, за якого своєчасне завершення оброблення стає поодиноким випадком.

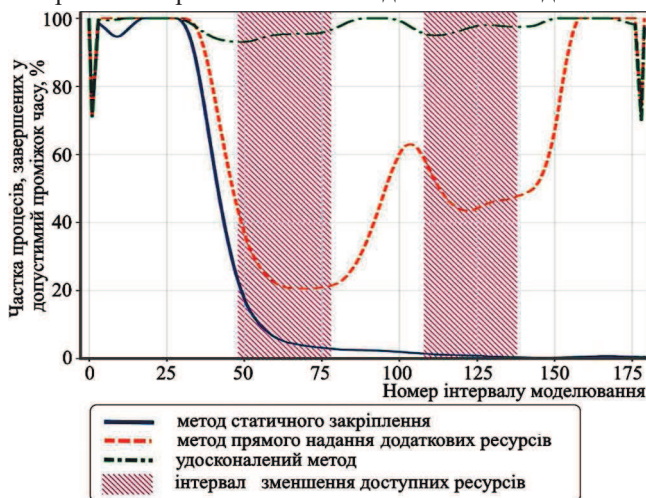


Рис. 2. Динаміка частки процесів, завершених у допустимий проміжок часу, за різних підходів / Dynamics of the share of processes completed within the allowable time interval for different methods

Метод прямого надання додаткових ресурсів [5] забезпечує кращі результати, однак не гарантує стійкого дотримання часових вимог. У межах першого інтервалу зменшення доступних ресурсів частка своєчасно завершених процесів зменшується приблизно до 20-25 %, а в межах другого переважно змінюється в діапазоні 45-65 % з подальшим відновленням після завершення впливу. Це означає, що пряме надання додаткових ресурсів послаблює наслідки їх перевантаження, але не усуває їх повністю.

Для удосконаленого методу частка процесів, завершених у допустимий проміжок часу, протягом основної частини інтервалу моделювання зберігається переважно в межах 93-100 %. Це означає, що тимчасове переміщення процесів на рівні-донори з незайнятими обчислювальними ресурсами дає змогу підтримувати високий рівень своєчасного завершення процесів навіть у періоди зменшення доступних ресурсів окремих рівнів. Отже, за цим показником удосконалений метод забезпечує найбільш стійкий режим функціонування серед розглянутих підходів.

Незначні хвилеподібні відхилення кривої в межах інтервалів зменшення доступних ресурсів не свідчать про втрату дієвості удосконаленого методу. Вони відображають вплив часу, потрібного для збереження стану процесу, його переміщення, поновлення виконання на рівні-донорі та повернення на попередній рівень. Проте, й за наявності цих часових витрат показники залишаються близькими до номінального рівня.

Окремого пояснення потребують локальні провали кривих на початку та наприкінці інтервалу моделювання, зафіксовані на рис. 1 і 2. Зазначені відхилення мають одночасний характер для всіх трьох підходів і спостерігаються на крайових ділянках вибірки, тобто в точках, де основний режим моделювання ще не сформовано або вже завершено. Такі відхилення доцільно трактувати як крайові ефекти числового опрацювання часових рядів і початкового налаштування стану БІС. Тому під час інтерпретації результатів основну увагу потрібно зосереджувати на центральній частині інтервалу моделювання, де поведінка всіх трьох підходів має стійкий характер.

Отже, порівняльне оцінювання за двома часовими показниками показало, що удосконалений метод забезпечує найменшу середню тривалість завершення процесів і найбільшу частку процесів, завершених у допустимий проміжок часу. Порівняно зі статичним закріпленням процесів за рівнями [13] він не допускає накопичення затримок в інтервалах зменшення доступних ресурсів, а порівняно з методом прямого надання додаткових ресурсів [5] забезпечує вищу стійкість часових показників. Це підтверджує доцільність застосування тимчасового міжрівневого переміщення процесів у БІС для того класу процесів, для якого таке переміщення є допустимим.

Обговорення результатів дослідження. У дослідженні [8] запропоновано двокритеріальний підхід до перенесення контейнерів у промислових периферійних середовищах, у межах якого одночасно враховують зменшення затримки та підвищення надійності. Побудовано оптимізаційну модель, що описує затримки передавання, зв'язку й оброблення, а також імовірність збоїв під час перенесення контейнерів, і запропоновано алгоритм пошуку раціонального рішення. Зазначений

підхід є близьким до цього дослідження за логікою врахування часових витрат під час переміщення обчислювального об'єкта. Водночас, у цій роботі розглядають перенесення контейнерів унаслідок зміни положення промислових терміналів, тоді як у цьому дослідженні формалізовано тимчасове міжрівневе переміщення саме процесу в БІС за нерівномірного розподілу навантаження, з вибором рівня-донора та обов'язковим поверненням процесу на попередній рівень.

У роботі [19] досліджено методику надійного планування робочих процесів у хмарному середовищі за використанням нестійких обчислювальних екземплярів, які можуть бути дотерміново відкликани системою. Для підвищення імовірності своєчасного завершення задач використано імітаційне моделювання Монте-Карло та повторний запуск перерваних задач на стійкіших обчислювальних засобах. Ця робота є релевантною в частині врахування часових обмежень і нестабільності ресурсного середовища. Проте, запропоноване у цій роботі рішення орієнтоване на перепризначення або повторне виконання задач у хмарному середовищі, а не на тимчасове переміщення вже активного процесу між рівнями БІС із збереженням стану, поновленням виконання на рівні-донорі та процесами міграції процесу.

У дослідженні [16] запропоновано підхід до виявлення незайнятих обчислювальних ресурсів у розподілених екзаскалярних системах для підтримки динамічних і взаємопов'язаних подій. Запропоноване рішення орієнтоване на випереджальне призначення ресурсів на основі прогнозування зміни робочого навантаження, що є важливим для систем зі змінною інтенсивністю оброблення даних. Ця робота підтверджує доцільність врахування поточного стану ресурсного середовища під час прийняття рішень щодо перерозподілу навантаження. Проте, в цій роботі вирішують завдання пошуку і призначення незайнятих ресурсів у розподіленій архітектурі, але не визначають, які саме процеси доцільно тимчасово переміщувати між рівнями БІС, за яких умов таке переміщення є допустимим і як забезпечити повернення процесу після завершення його виконання на рівні-донорі.

У роботі [11] запропоновано адаптивний метод міграції сервісів у мобільних кордонних обчисленнях, орієнтований на зменшення затримки та енергетичних витрат за дефіциту обчислювальних ресурсів. Підхід ґрунтується на моніторингу стану ресурсного середовища, виявленні вузлів із незайнятими обчислювальними ресурсами та прийнятті рішення про міграцію сервісу на інший вузол для зменшення черг і часу очікування. Значущість цієї роботи полягає у підтвердженні доцільності врахування поточного стану ресурсів під час міграції обчислювальних об'єктів. Водночас, у цій роботі не зазначено, як саме можливе повернення процесу на нативний рівень. Окрім цього, запропоноване авторами рішення орієнтоване на централізований збір інформації про стан середовища, що не охоплює задачу формалізованого визначення допустимості короткочасного припинення виконання процесу.

У дослідженні [12] розглянуто метод міграції процесів, що побудований на механізмах аналізу залежностей між сервісами та специфікою прийняття рішень на основі глибокого навчання з підкріпленням. Запропонований підхід орієнтований на зменшення затримки виконання процесів і вирівнювання завантаження між серве-

рами за рахунок урахування критичних шляхів у структурі міжпроцесної взаємодії та поточного стану вузлів. Однак проблемою цього підходу є локалізація рішення у горизонтальній площині рівня, що, власне, і надає йому високу швидкодію, тоді як обчислювальні ресурси на інших рівнях, навіть за умови їх незайнятості, не можуть бути задіяні цими процесами.

Унаслідок обговорення результатів дослідження встановлено, що сучасні наукові дослідження у даній предметній області охоплюють перенесення контейнерів у периферійних середовищах, надійне планування задач у хмарних обчисленнях, виявлення незайнятих обчислювальних ресурсів у розподілених системах, а також адаптивну міграцію сервісів у середовищах мобільних периферійних обчислень і міграцію сервісів з урахуванням залежностей між їх компонентами. Водночас, поза увагою розглянутих підходів залишається завдання тимчасового міжрівневого переміщення саме процесу в БІС з урахуванням допустимості його короткочасного припинення, вибору рівня-донора за критерієм мінімізації сумарної тривалості збереження стану, переміщення, виконання та повернення, а також формалізованого визначення меж застосування методу. Це підтверджує доцільність виконаного дослідження та коректність обраного напрямку удосконалення методу динамічного перерозподілу навантаження між системними рівнями.

Отже, внаслідок виконаної роботи можна сформулювати такі наукову новизну та практичну значущість результатів дослідження.

Наукова новизна отриманих результатів дослідження – удосконалено метод динамічного перерозподілу навантаження між рівнями БІС, який, на відміну від підходів статичного закріплення процесів за рівнями та прямого надання додаткових обчислювальних ресурсів, передбачає тимчасове переміщення процесу на рівень-донор із незайнятими обчислювальними ресурсами, вибір такого рівня за критерієм мінімізації сумарної тривалості збереження стану, переміщення, виконання та повернення з нього, а також урахування допустимості короткочасного припинення процесу без порушення виконання основної функції БІС, що забезпечує вищу стійкість часових показників виконання процесів за нерівномірного розподілу навантаження між рівнями.

Практична значущість результатів дослідження – можна використати під час побудови та вдосконалення БІС, у яких виникає нерівномірне навантаження між рівнями, зокрема у віртуалізованих, хмарних і туманних середовищах, для скорочення середньої тривалості завершення процесів, підвищення частки процесів, завершених у допустимий проміжок часу, та вибору доцільного рівня-донора для тимчасового виконання процесу за нестачі локальних обчислювальних ресурсів.

Висновки / Conclusions

Удосконалено метод динамічного перерозподілу навантаження між рівнями БІС шляхом введення правил тимчасового переміщення процесу на інший рівень з подальшим поверненням на попередній рівень, на якому він виконувався до перенесення, з урахуванням часових витрат на таке переміщення та умов допустимості його застосування для визначеного класу процесів, що дало змогу зменшити тривалість завершення процесів і підвищити частку процесів, завершених у допустимий проміжок часу. За результатами проведеного дослідження можна зробити такі основні висновки.

1. Встановлено, що нерівномірний розподіл навантаження між рівнями БІС призводить до зростання тривалості завершення процесів навіть за наявності незайнятих обчислювальних ресурсів на інших рівнях. Це підтвердило доцільність тимчасового переміщення процесу на інший рівень за умови непорушення виконання основної функції системи.

2. Сформовано формалізований опис рівнів, процесів, показників зайнятості та правил вибору рівня-донора. Це дало змогу задати умови допустимості тимчасового перенесення процесу, визначити множину допустимих рівнів-донорів і встановити правило вибору рівня за мінімальною сумарною тривалістю збереження стану, переміщення, виконання та повернення.
3. Запропоновано послідовність реалізації удосконаленого методу, яка включає виявлення рівнів з надмірною зайнятістю, відбір процесів, придатних до тимчасового перенесення, вибір рівня-донора, виконання процесу на ньому, повернення на попередній рівень і повторне оцінювання стану системи.
4. Визначено меж застосування удосконаленого методу. Показано, що його застосування є обґрунтованим тільки для процесів, які допускають короткочасне припинення, мають хоча б один допустимий рівень-донор і забезпечують скорочення тривалості завершення процесу порівняно з продовженням виконання на попередньому рівні.
5. За результатами порівняльного оцінювання встановлено, що удосконалений метод забезпечує найменшу середню тривалість завершення процесів і найбільшу частку процесів, завершених у допустимий проміжок часу, серед розглянутих підходів, а також підвищує стійкість часових показників виконання процесів у БІС за нерівномірного розподілу навантаження між рівнями.

References

1. Albalawi, N. S. (2025). Dynamic scheduling strategies for cloud-based load balancing in parallel and distributed systems. *Journal of Cloud Computing*, 14(33). <https://doi.org/10.1186/s13677-025-00757-6>
2. Aleyadeh, S., Moubayed, A., Heidari, P., & Shami, A. (2022). Optimal Container Migration/Re-Instantiation in Hybrid Computing Environments. *IEEE Open Journal of the Communications Society*, 3, 15–30. <https://doi.org/10.1109/ojcoms.2022.3140272>
3. Chepurina, I., & Frolov, D. (2025). A method for increasing the productivity of a distributed firewall based on Proxmox in corporate computer networks. *Innovative technologies and scientific solutions for industries*, 3(33), 180–188. <https://doi.org/10.30837/2522-9818.2025.3.180>
4. Choudhury, S., Maheshwari, S., Sesar, I., & Raychaudhuri, D. (2022). ShareOn: Shared Resource Dynamic Container Migration Framework for Real-Time Support in Mobile Edge Clouds. *IEEE Access*, 10, 66045–66060. <https://doi.org/10.1109/ACCESS.2022.3183122>
5. Choularas, S., & Sotiriadis, S. (2022). Auto-scaling containerized cloud applications: A workload-driven approach. *Simulation Modelling Practice and Theory*, 121, article number 102654. <https://doi.org/10.1016/j.simpat.2022.102654>
6. Dodonov, O. G., Putyatin, V. G., Dodonov, V. O., Kutsenko, S. A., Germaniuk, A. P., Izvarin, I. V., & Kravchuk, K. O. (2024). Technology for ensuring the survivability of territorially distributed information computer systems in a unified information space. *Data Recording, Storage & Processing*, 26(1), 121–143. <https://doi.org/10.35681/1560-9189.2024.26.1.308659>
7. Hossain, M. R., Whaiduzzaman, M., Barros, A., & Fidge, C. (2024). Dynamic microservice placement in multi-tier Fog networks. *Internet of Things*, article number 101224. <https://doi.org/10.1016/j.iot.2024.101224>

8. Jin, X., He, S., & Chen, Y. (2024). Container migration for edge computing in industrial Internet with joint latency reduction and reliability enhancement. *Scientific Reports*, 14(1), article number 25394. <https://doi.org/10.1038/s41598-024-77086-2>
9. Kuchta, O. P., & Pirko, I. B. (2024). Analysis of dynamic pathfinding algorithms under variable environmental conditions. *Scientific Bulletin of UNFU*, 34(7), 46–51. <https://doi.org/10.36930/40340706>
10. Kyzym, M., & Yurchenko, O. (2025). Features of digital transformation in the management of higher education institutions in Ukraine. *Economy and Society*, 82, article number 61. <https://doi.org/10.32782/2524-0072/2025-82-61>
11. Li, L., Lv, J., Wang, S., Hu, C., Sun, G., & Li, N. (2026). Latency and energy-aware adaptive service migration in mobile edge computing. *Scientific Reports*, 16, article number 6178. <https://doi.org/10.1038/s41598-026-36711-y>
12. Liu, Y., Liu, Z., & Yao, Y. (2026). Research on Dependency-Aware Service Migration Strategy in the Internet of Vehicles Integrating a Graph Attention Network and Deep Reinforcement Learning. *Applied Sciences*, 16(2), article number 943. <https://doi.org/10.3390/app16020943>
13. Petrovska, I., & Kuchuk, H. (2022). Static allocation method in a cloud environment with a service model IAAS. *Advanced Information Systems*, 6(3), 99–106. <https://doi.org/10.20998/2522-9052.2022.3.13>
14. Ponochovnyi, Y. L., & Kharchenko, V. S. (2020). Dependability Assurance Methodology of Information and Control Systems Using Multipurpose Service Strategies. *Radioelectronic and Computer Systems*, 3, 43–58. <https://doi.org/10.32620/reks.2020.3.05>
15. Posner, J., Ellersick, T., Bietendorf, N., Huber, D., Schreiber, M., & Schulz, M. (2025). Toward Dynamic Resource Management: An Asynchronous Many-Task (AMT) Runtime System leveraging Dynamic Processes with PSets (DPP). *SN Computer Science*, 6(8), article number 941. <https://doi.org/10.1007/s42979-025-04405-3>
16. Rezaie, S., Khaneghah, E. M., Pavone, M. F., Hajirajabi, M., Safari, R., Suvizi, A., & Showkatabad, A. R. (2025). OutExaRD: Enhancing Resource Discovery for Support Dynamic and Interactive Events on Out-of-Distributed Exascale Systems. *International Journal of Networked and Distributed Computing*, 13, article number 18. <https://doi.org/10.1007/s44227-025-00059-y>
17. Ruban, I. V., & Tkachov, V. M. (2025). Multilevel model of an information system on a mobile platform and formalization of its survivability criteria. *Visnyk of Kherson National Technical University*, 2 (3(94)), 299–409. <https://doi.org/10.35546/kntu2078-4481.2025.3.2.51>
18. Taleb, I., Guillaume, J.-L., & Duthil, B. (2025). A Survey on Services Placement Algorithms in Integrated Cloud-Fog / Edge Computing. *ACM Computing Surveys*, 57(11), article number 271. <https://doi.org/10.1145/3729214>
19. Wu, Q., Fang, J., Zeng, J., Wen, J., & Luo, F. (2024). Monte Carlo Simulation-Based Robust Workflow Scheduling for Spot Instances in Cloud Environments. *Tsinghua Science and Technology*, 29(1), 112–126. <https://doi.org/10.26599/tst.2022.9010065>

D. Ye. Frolov

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

INTER-LEVEL DYNAMIC LOAD REDISTRIBUTION METHOD IN MULTI-LEVEL INFORMATION SYSTEMS

It has been found that, in multi-level information systems, an uneven distribution of workload across levels leads to an increase in process completion time even when idle computing resources are available at other levels. It has been established that, under such conditions, temporary relocation of a process to another level, followed by its execution there and return to the level at which the process was executed before the temporary relocation, referred to as the source level, is expedient, provided that short-term suspension does not disrupt the execution of the systems main function and that the total time required for state preservation, relocation, execution, and return is shorter than the time needed to complete the process at the source level. It has been found that existing approaches focus on microservice placement, container and virtual machine migration, or the allocation of idle resources; however, approaches based on dynamic microservice placement, container migration, virtual machine migration, and direct allocation of additional computing resources do not formalize the problem of temporary relocation of the process itself between levels with mandatory return to the source level. A formalized description of levels, processes, their utilization indicators, admissibility conditions for temporary relocation, and rules for selecting a donor level according to the criterion of minimizing the total process completion time has been developed. The method of dynamic load redistribution across levels has been improved; it includes detecting overutilized levels, selecting processes eligible for temporary relocation, determining the reduction in process completion time, choosing a donor level, executing the process at that level, returning it to the source level, and re-evaluating the system state. The applicability limits of the improved method have been established for processes that allow short-term suspension, have at least one admissible donor level, and ensure a reduction in completion time compared with continued execution at the source level. The influence of the proposed approach on the time-related performance indicators of process execution has been assessed by means of simulation modeling. It has been established that the improved method provides the lowest average process completion time, the highest proportion of processes completed within the allowable time interval, and greater stability of time-related performance indicators under uneven workload distribution across levels.

Keywords: fault tolerance of the information system; computing resource; process relocation; simulation modeling; dynamic deployment of microservices.