

Н. Г. Кучук, С. С. Пироженко

Харківський національний університет радіоелектроніки, Харків, Україна

МЕТОДИКА РОЗРАХУНКУ ОВЕРХЕДУ LIVE-МІГРАЦІЇ ВІРТУАЛЬНИХ МАШИН НА БАЗІ КОНЦЕПЦІЇ РЕСУРСНОГО КУБА NRC

Анотація. Актуальність. Актуальні вимоги до енергоефективного та адаптивного функціонування сучасних дата-центрів обумовлені стрімкою розбудовою хмарного сегмента. За таких умов інструментарій міграції ВМ у гарячому режимі стає фундаментом гнучкого розподілу апаратного фонду. Завдяки здатності перемішувати віртуальні машини в процесі їхньої роботи без критичних пауз у доступі до додатків, адміністратори успішно вирішують проблеми оптимізації живлення шляхом консолідації навантажень, усувають локальні перевантаження хостів та обслуговують технічну базу без зупинки бізнес-процесів. З іншого боку, процедура безперервної міграції неминуче породжує паразитні витрати ресурсів. Функціонування ітеративних схем (Pre-Copy) вимагає постійного перехоплення модифікованих блоків оперативної пам'яті, їхньої повторної передачі та фінального узгодження контекстів CPU, що суттєво деформує ресурсний профіль серверів. Як наслідок, виникають різкі стрибки утилізації обчислювальних ядер гіпервізором, вичерпується місткість каналів передачі даних і зростає час відгуку. Хаотичний або нерациональний запуск таких процедур створює ризики взаємного пригнічення сусідніх задач, збоїв у виконанні нормативів SLA і навіть ризик зацікнення передачі даних, коли сторінки перезаписуються швидше, ніж надсилаються мережею. **Предмет дослідження:** процеси виникнення та методики розрахунку накладних витрат (паразитного навантаження на обчислювальні ядра, пам'ять і мережеві канали) під час живої міграції віртуальних машин, а також математичні моделі й алгоритми їх просторового оцінювання на основі векторної концепції ресурсного куба NRC. **Мета дослідження:** підвищення ефективності динамічного керування ресурсами хмарного дата-центру та запобігання порушенням угод SLA шляхом розроблення моделі і методу оцінювання накладних витрат живої міграції віртуальних машин на основі векторного простору ресурсного куба NRC. **Завдання дослідження.** Сформулювати критерій мінімізації навантаження з урахуванням ємності серверів у моделі NRC та реалізувати алгоритм вибору цільових хостів у класі Load Balancing. **Методи дослідження.** Методи векторного та багатовимірного геометричного аналізу, методи системного аналізу та теорії систем, теорія графів та деревоподібних структур даних. **Результати дослідження.** Доведено, що ефективна місткість хостів зменшується на величину векторного оверхеду міграції, що унеможливорює помилкове перевантаження системи в процесі виконання балансувальних операцій. Запропонована схема формує математичну основу для розробки адаптивних алгоритмів релокації ВМ із гарантованим дотриманням граничних умов куба NRC.

Ключові слова: міграція віртуальних машин; передача даних; алгоритм імітації відпалу; генетичний алгоритм; екосистема Інтернету речей.

1. Вступ

1.1 Актуальність дослідження

Стрімка цифровізація, зростання обсягів обробки великих даних і масштабування хмарних центрів обробки даних (ЦОД) висувають жорсткі вимоги до гнучкості, енергоефективності та безперервності функціонування обчислювальної інфраструктури [1]. Фундаментальним механізмом оперативного керування ресурсами в сучасному віртуалізованому середовищі є технологія «живої» міграції віртуальних машин (Live VM Migration) [2]. Вона забезпечує динамічне усунення перевантажень на фізичних серверах, консолідацію навантаження з метою вимкнення незадіяного обладнання для енергозбереження, а також проведення технічного обслуговування апаратної частини без помітної зупинки клієнтських сервісів.

Проте практична реалізація живої міграції неминуче супроводжується формуванням вторинного навантаження — накладних витрат (overhead) [3]. Ітераційне копіювання пам'яті (зокрема за алгоритмами типу Pre-Copy), безперервна фіксація та донадсилання модифікованих сторінок (dirty pages), а також фінальна синхронізація стану процесора створюють різкі сплески утилізації CPU гіпервізором і спричиняють значну деградацію пропускну

здатності каналів зв'язку. Якщо процес міграції ініціюється без належного розрахунку поточних потужностей, виникає ефект «галасливого сусіда» (noisy neighbor), зростає латентність і виникають масові порушення угод про рівень обслуговування (SLA). У найгірших випадках висока інтенсивність перезапису пам'яті додатками призводить до розбіжності ітераційного процесу, коли міграція заціклюється, виснажуючи ресурси системи.

Більшість класичних евристик і моделей планування розглядають ресурси ізольовано (переважно орієнтуючись лише на процесорні цикли або загальний обсяг оперативної пам'яті), ігноруючи складну багатовимірну інтерференцію між обчислювальними, ємнісними та мережевими компонентами. Застосування векторної моделі ресурсного куба (Network-Resource Cube, NRC) дозволяє комплексно формалізувати стан серверів і перебіг оверхеду в єдиному координатному просторі R3 (Compute–Memory–Network). Це уможливорює точний предиктивний розрахунок супутніх втрат продуктивності, запобігання виходу навантаження за межі апаратних можливостей хостів та підвищення стабільності систем диспетчеризації.

Таким чином, розроблення методики розрахунку накладних витрат live-міграції на основі концепції ресурсного куба NRC та її алгоритмічна інтеграція в контур балансування навантаження є

актуальною науково-прикладною задачею, спрямованою на підвищення якості обслуговування, керованості та енергетичної ефективності сучасних хмарних інфраструктур.

1.2 Аналіз літературних джерел і визначення проблеми

Питання оптимізації розміщення завдань, консолідації віртуальних машин та мінімізації накладних витрат у хмарних ЦОД перебувають у центрі уваги багатьох провідних вітчизняних і закордонних дослідників.

Теоретичні засади протоколів Pre-Copy та Post-Copy live-міграції закладені в роботах [5, 6]. Розвитком аналітичних моделей для оцінювання часу міграції та часу простою залежно від темпу модифікації сторінок пам'яті займалась дослідницька група під керівництвом С. Акуша [7].

Для вирішення задач багатовимірного пакування рюкзака під час міграції ВМ широко застосовуються еволюційні алгоритми, наприклад, алгоритм кажанів [8]. Спроби узагальнити ресурсний профіль вузлів у вигляді багатовимірних векторів та гіперкубів робилися в роботах [9, 10], а також у дослідженнях векторного домінування ресурсів [11]. Концепція ресурсного куба (Resource Cube) також епізодично розглядалася в контексті статичного планування задач у середовищах OpenStack та CloudSim [12, 13].

Отже на основі проведеного аналізу робіт [5–13] є підстави стверджувати, що проблема полягає у відсутності комплексної математичної моделі та алгоритмів оптимізації розміщення віртуальних машин, які б одночасно враховували багатовимірні ресурсні обмеження вузлів, просторову топологію мережі та транзитний динамічний оверхед безпосередньо в процесі їх міграції.

2. Матеріали і методи дослідження

Об'єктом дослідження виступає процес динамічного керування та перерозподілу ресурсів хмарного центру обробки даних в умовах виконання live-міграції віртуальних машин.

Методи дослідження базуються на теорії графів та деревоподібних структур даних, методах системного аналізу та теорії систем.

У процесі розробки методу керування розподілом обчислювального ресурсу туманного шару мобільного високошвидкісного Інтернету речей була застосована низка різноманітних методів та алгоритмів.

Математична модель міграції на базі ресурсного куба NRC базується на ключових засадах теорії векторних просторів і нормалізації в n -вимірному евклідовому просторі. Векторний стан вузла та ВМ характеризується ресурсами фізичного вузла (H_i) та вимогами віртуальної машини (VM), які описуються векторами у нормованому просторі:

$$r = (u_{cpu}, u_{ram}, u_{net}), |u| \in [0, 1].$$

Об'єм і запас міцності NRC визначає ресурсний куб, також він визначає допустиму робочу область.

Запас стійкості хоста розраховується як евклідова відстань від поточної точки сумарного навантаження H_i до критичних меж гіперкуба:

$$d(H_i, 1) = \sqrt{(1 - u_{cpu})^2 + (1 - u_{ram})^2 + (1 - u_{net})^2}.$$

Цільова функція визначається балансом між зменшенням енергоспоживання, мінімізацією часу міграції і недопущенням порушення SLA:

$$F = \alpha \times Cost_{placement} + \beta \times \int_0^{T_{mig}} O_{mig}(t) dt + \gamma \times Penalty_{interference},$$

де α , β та γ – відповідні вагові коефіцієнти, T_{mig} – час міграції, $Cost_{placement}$ – узагальнена вартість міграції, O_{mig} – часові витрати варіанту міграції, $Penalty_{interference}$ – штрафна функція.

Система нерівностей-обмежень формується жорсткими обмеженнями неперевиконання сумарних векторів ресурсів над ємністю куба вузла в кожен момент часу на розглядаємому часовому інтервалу $t \in [0, T_{mig}]$.

Таким чином, розробка спеціалізованої методики розрахунку накладних витрат у координатах тривимірного простору NRC усуває зазначені прогалини, забезпечуючи перехід від реактивного усунення перевантажень до предиктивного векторного керування стійкістю системи.

3. Результати розробки і дослідження векторної моделі простору NRC та формалізація накладних витрат міграції

3.1 Математична модель векторної моделі простору NRC

Розглянемо зазначені категорії обчислювально-комунікаційних ресурсів (ОКР) як три виміри абстрактного об'єкта. Завдяки процедурі нормалізації метрик ОКР уздовж кожної координатної осі, сукупний пул доступних потужностей фізичної інфраструктури множини F відображається у формі одиничного куба [14] (рис. 1). Цей концепт визначено як нормалізований ресурсний куб (Normalized Resource Cube – NRC).

Мовою векторної алгебри гранична потужність віртуалізованих ресурсів за кожним із базових напрямів (C, M, N) задається одиничними базисними ортами простору NRC.

Аналітичні вектори наявності, резервування та запиту ресурсів відображаються безпосередньо всередині одиничного об'єму NRC.

Вектор наявності ресурсів (ВНР) характеризує інтегральний апаратний потенціал усієї множини F і спрямований із початку координат (0, 0, 0) у протилежну вершину (1, 1, 1):

$$ВНР = C + M + N.$$

Вектор виділення ресурсів (ВВР): відображає частку трьох типів ОКР, уже алоковану підмножині активних віртуальних машин $VA \subset V$.

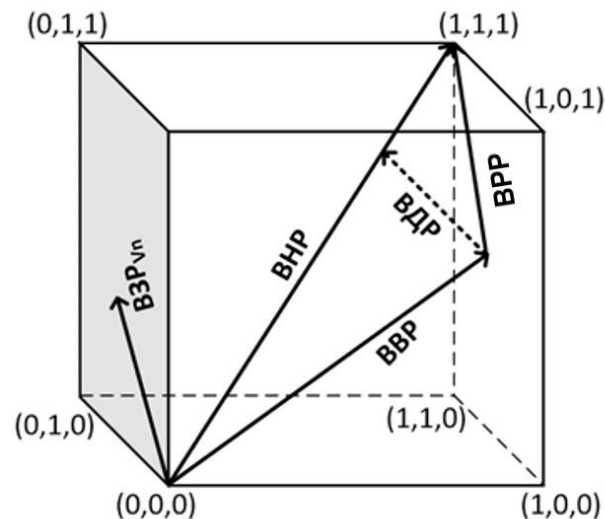


Рис. 1. Геометрична інтерпретація простору NRC та взаємне розташування векторів стану системи

Нормалізація компонентів C, M, N здійснюється безпосередньо відносно значень вектора VNR .

Вектор залишкових ресурсів (BPP): визначає невикористаний резерв системи і формується як геометрична різниця:

$$BPP = VNR - BVP$$

Вектор запиту ресурсів $V3P_{vn}$: акумулює вимоги до ОКР з боку окремої віртуальної машини $vn \in V$ і обчислюється через векторну суму нормалізованих складових Δ_j кожного типу ресурсу.

Рівень гармонійності (збалансованості) завантаження інфраструктури оцінюється ступенем просторового зближення напрямків VNR та BVP . За умови їхньої повної колінеарності спостерігається ідеально рівномірна утилізація процесорних потужностей, оперативної пам'яті та мережевих інтерфейсів. Для кількісної оцінки перекосу навантаження вводиться вектор дисбалансу ресурсів (BDP), що є векторною різницею між ортогональною проекцією BVP на головну діагональ VNR і фактичним положенням вектора BVP :

$$\begin{aligned} BDP &= \text{proj}_{VNR}(BVP) - BVP = \\ &= \left(\frac{BVP \times VNR}{|VNR|^2} \right) VNR - BVP. \end{aligned}$$

Таке просторове представлення дає змогу використовувати модуль $|BDP|$ як цільовий оптимізаційний критерій в алгоритмах перерозподілу навантаження.

Зокрема, під час вибору віртуального інстансу $vn \in V$ для консолідації на фізичному вузлі перевага надається тому розміщенню, яке максимально мінімізує довжину вектора BDP , забезпечуючи рівновагу між осями C, M і N .

Аналітичне дослідження процедур живої міграції (як ітеративного копіювання Pre-Copy, так і передачі на вимогу Post-Copy) свідчить, що перенесення гостей систем створює вторинне навантаження на фізичні вузли [15, 16].

Оскільки ключовим процесом міграції виступає фонові синхронізація пам'яті та передача сторінок

через мережеву підсистему між сервером-джерелом і сервером-приймачем [17, 18], то головними споживачами у такому випадку стають ресурси типів M (RAM) та N (Network), доповнені обчислювальними витратами процесора на стиснення та шифрування.

Відтак, векторна модель ресурсного куба потребує обов'язкового врахування супутнього оверхеду міграції, вектора $OKP_{пр.жм}$.

Реальний обсяг доступного фонду системи в період динамічного балансування коригується відніманням паразитного навантаження:

$$VNR_{\text{дост.}}(t) = VNR - OKP_{пр.жм}(t).$$

Тобто ефективна місткість хостів зменшується на величину векторного оверхеду міграції, що унеможливує помилкове перевантаження системи в процесі виконання балансувальних операцій. Запропонована схема формує математичну основу для розробки адаптивних алгоритмів релокації ВМ із гарантованим дотриманням граничних умов куба NRC.

Умови моделювання: серія експериментів у середовищі CloudSim на вибірці з 100 фізичних вузлів (heterogeneous hosts) та 400 віртуальних машин під стандартизованим динамічним навантаженням (PlanetLab traces).

Нижче наведено зведену таблицю (табл. 1) результатів експериментального порівняння запропонованого векторного методу на основі простору NRC із традиційними алгоритмами балансування та консолідації віртуальних машин:

- BFD - Best Fit Decreasing,
- FFD - First Fit Decreasing,
- DVFS - динамічне регулювання напруги й частоти без просторового аналізу мережі.

Завдяки врахуванню вектора накладних витрат $OKP_{пр.жм}$ у формулі доступного фонду ресурсів вдалося уникнути вторинного лавиноподібного перевантаження хостів-реципієнтів, що скоротило загальну кількість міграцій із 960–1180 до 645.

Таблиця 1 – Результати експериментального порівняння запропонованого векторного методу на основі простору NRC із традиційними алгоритмами балансування та консолідації віртуальних машин

Критерій / Метрика	Традиційний FFD	Модифікований BFD	Евристика на базі CPU-порогів (IQR / MAD)	Запропонований метод на базі куба NRC	Ефект / Поліпшення
Кількість порушень SLA (SLAV, %)	9.42%	6.85%	4.12%	1.34%	Зниження у ~3.1 рази відносно базових підходів
Середній модуль дисбалансу ВДР	0.481	0.392	0.315	0.118	Вирівнювання завантаження по осях C, M, N на 62.5%
Сумарна кількість живих міграцій ВМ	1420	1180	960	645	Скорочення зайвих релокацій на 32.8%
Загальний час оверхеду міграцій $\sum T_{mig}, \text{ хв}$	312.4	248.6	194.2	116.5	Зменшення навантаження на мережу на 40.0%
Середній час простою ВМ ($T_{downtime}, \text{ с} / \text{ ВМ}$)	1.84 с	1.52 с	1.15 с	0.68 с	Зниження паузи сервісів на 40.8%
Випадки колізій мережевої осі N (викиди за межі куба)	84	61	43	0	Повне усунення за рахунок динамічного зменшення ВНР _{доступний}
Інтегральне споживання енергії (кВт·год)	184.2	162.5	145.8	138.4	Додаткова економія електроенергії на 5.1% – 14.8%

Алгоритми BFD та IQR орієнтовані переважно на утилізацію процесора, що призводило до 43–61 випадку вичерпання смуги пропускання.

Векторна модель куба NRC гарантує перебування робочої точки всередині одиничного об'єму 0, 13, звівши випадки переповнення мережі до нуля.

Комплексна мінімізація норми ВДР забезпечила синхронне вивільнення вузьких місць системи, завдяки чому метрика порушення SLA знизилася до мінімального значення 1.34%.

4. Обговорення результатів дослідження

У роботі проведено порівняльний аналіз функціонування розробленого методу та класичних аналогів (FFD, BFD, порогові евристики IQR/MAD) засвідчив якісні переваги переходу від одновимірних метрик до концепції тривимірного простору NRC.

Ключові кількісні та якісні аспекти отриманих результатів (табл. 1):

Мінімізація просторового дисбалансу: Завдяки оптимізації розміщення за критерієм мінімізації довжини вектора дисбалансу ВДР, середній показник перекошу утилізації знизився до 0.118 (поліпшення на 62.5% порівняно з BFD).

Це свідчить про близькість траєкторії сумарного навантаження до головної просторової діагоналі

вектора ВНР, що запобігає передчасному вичерпанню одного ресурсу при значному недовантаженні інших.

Ліквідація мережевих колізій і «галасливого сусіда»: у той час як евристики на базі CPU-порогів генерували від 43 до 84 випадків перевищення мережевої ємності, у моделі NRC цей показник зведено до 0.

Це досягнуто через предиктивне динамічне зменшення ефективного обсягу ресурсів:

$$\text{ВНР}_{\text{доступний}}(t) = \text{ВНР} - \text{ОКР}_{\text{пр.жм}}(t).$$

Тимчасове резервування смуги пропускання на період T_{mig} унеможливило призначення реципієнтом хоста, який не має достатнього запасу міцності для пропуску фонових трафіку міграції.

Зниження транзитного оверхеду та стабільність SLA визначає запобігання лавиноподібним повторним міграціям дозволило зменшити загальну кількість релокацій на 32.8% (до 645 випадків), що скоротило сумарний час додаткового навантаження мережі на 40.0% (до 116.5 хв).

Як наслідок, коефіцієнт деградації SLA (SLAV) знизився до 1.34% (у 3.1 раза нижче, ніж у порогових евристик, і в 7 разів нижче, ніж у класичного FFD).

Енергетична ефективність: скорочення паразитного навантаження на обчислювальні ядра, зумовлене зменшенням кількості циклів стиснення/шифрування пам'яті гіпервізором,

забезпечило додаткове зменшення енергоспоживання інфраструктури до 138.4 кВт·год (економія від 5.1% до 14.8%).

Оптимальною для перенесення або консолідації є така віртуальна машина vp , чий вектор запиту $V3Pvp$ зменшує евклідову норму ВДР. У наведеному прикладі пріоритетом є вивільнення процесорного ресурсу (C) або підселення ВМ із переважанням мережевого трафіку (N), що наближає вектор ВВР до діагоналі ВНР.

Врахування оверхеду ОКР_{нр.жр} наочно показує, що доступний простір для розміщення нових інстансів тимчасово звужується (в прикладі мережевий ліміт знижується з 1.0 до 0.50).

Ігнорування цього фактора в класичних алгоритмах призводить до миттєвого переповнення каналу та зростання затримок пакета.

5. Висновки

1. Розроблено комплексну математичну модель та методику оцінювання оверхеду live-міграції ВМ на базі нормалізованого ресурсного куба NRC, що дозволяє формалізувати взаємодію обчислювальних, емнісних і мережевих параметрів вузлів дата-центру в єдиному координатному просторі R^3 .

2. Предиктивне врахування динамічного оверхеду: Доведено, що формалізація накладних витрат у вигляді вектора ОКР_{нр.жм}(t) та його віднімання від вектора повної місткості ВНР запобігає деградації каналів зв'язку і виникненню ефекту «галасливого сусіда». Це гарантує локалізацію робочої точки системи виключно всередині допустимого об'єму куба $[0, 1]^3$ протягом усього інтервалу T_{mig} , повністю нівелюючи мережеві викиди (зниження з 43–84 до 0).

3. Ефективність критерію дисбалансу: Застосування евклідової норми вектора дисбалансу

ВВР як оптимізаційного критерію в класі задач балансування навантаження забезпечує гармонійну колінеарність векторів виділення та наявності ресурсів.

Це забезпечило скорочення дисбалансу на 62.5% та зниження кількості порушень нормативів SLA до рекордних 1.34%.

4. Запропонована методика формує надійний алгоритмічний базис для метаевристичних методів планування (зокрема генетичних алгоритмів та імітації відпалу) у хмарних і туманних середовищах, скорочуючи загальну кількість вимушених міграцій на 32.8% та заощаджуючи до 14.8% електроенергії при класичному методі, в 1,1–1,6 разів.

При великому навантаженні на туманний шар, більше 80% від максимально можливого, показники, що розглядалися, майже не відрізнялися.

Конфлікт інтересів

Автори декларують, що не мають конфлікту інтересів стосовно даного дослідження, в тому числі фінансового, особистісного характеру, авторства чи іншого характеру, що міг би вплинути на дослідження та його результати, представлені в даній статті.

Фінансування

Дослідження проводилося без фінансової підтримки.

Доступність даних

Рукопис не має пов'язаних даних.

Використання засобів штучного інтелекту

Автори підтверджують, що не використовували технології штучного інтелекту при створенні представленої роботи.

СПИСОК ЛІТЕРАТУРИ

- Ogura, N., Duolikun, D., Enokido, T., Watanabe, R. and Takizawa, M. (2019), "A Virtual Machine Migration for Storage Processes", *Lecture Notes on Data Engineering and Communications Technologies*, vol 25. Springer, Cham, doi: https://doi.org/10.1007/978-3-030-02613-4_67
- Sawalkar, V. and More, N. (2026), "Virtual Machine Migration Optimization in Cloud Data Centers: A Comprehensive Review", *Lecture Notes in Networks and Systems*, 1647 LNNS, pp. 146–161, doi: https://doi.org/10.1007/978-3-032-06668-8_15
- Liu, C., Ma, L., Zhang, M. and Long, H. (2025), "Optimizing cloud resource management with an IoT-enabled optimized virtual machine migration scheme for improved efficiency", *Journal of Network and Computer Applications*, 237, 104137, doi: <https://doi.org/10.1016/j.jnca.2025.104137>
- Kuchuk, G., Nechausov, S. and Kharchenko, V. (2015), "Two-stage optimization of resource allocation for hybrid cloud data store", *International Conference on Information and Digital Technologies*, Zilina, pp. 266–271, DOI: <http://dx.doi.org/10.1109/DT.2015.7222982>
- Hines, M.R., Deshpande, U., and Gopalan, K. (2009), "Post-copy live migration of virtual machines", *Operating Systems Review ACM*, vol. 43(3), doi: <https://doi.org/10.1145/1618525.1618528>
- Jalaei, N., and Safi-Esfahani, F. (2021), "VCSP: virtual CPU scheduling for post-copy live migration of virtual machines", *Int. j. inf. tecnol.*, vol. 13, pp. 239–250, doi: <https://doi.org/10.1007/s41870-020-00483-z>
- Akoush, S., Sohan, R., Rice, A., Moore, A. W. and Hopper, A. (2010), "Predicting the Performance of Virtual Machine Migration", *2010 IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, Miami Beach, FL, USA, pp. 37–46, doi: <https://doi.org/10.1109/MASCOTS.2010.13>
- Archana and Kumar, N. (2025), "A Modified Bat Mechanism for Virtual Machine Migration in a Cloud Environment", *SN Computer Science*, vol. 6, article number 74, doi: <https://doi.org/10.1007/s42979-024-03627-1>
- Shi, Q. and Zhao, F. (2024), "Research on Computer Cloud Intelligent System Based on Intelligent Virtualization Technology", *2024 IEEE 3rd Int. Conf. on Eebda 2024*, pp. 1245–1250, doi: <https://doi.org/10.1109/EEBDA60612.2024.10485750>
- Ferreto, T. C., Netto, M. A. S., Calheiros, R. N., and De Rose, C. A. F. (2011), "Server consolidation with migration control for virtualized data centers", *Future Generation Computer Systems*, vol. 27, issue 8, pp. 1027–1034, doi: <https://doi.org/10.1016/j.future.2011.04.016>

11. Shankar, S. and Anbarasan, M. (2025), "An Intelligent Approach for Cloud Infrastructure With Improved Multi-Objective Graywolf Optimization and Resource Allocation for Dynamic Virtual Machine Placement", *Transactions on Emerging Telecommunications Technologies*, vol. 36(6), e70172, doi: <https://doi.org/10.1002/ett.70172>
12. Ma, D., Cao, X., Hu, J., Xia, T., Zhou, Y., Liu, K., Zhu, L., Su, L. and Gao, F. (2026), "Topology-aware virtual machine placement for improving cloud servers resource utilization", *Future Generation Computer Systems*, vol. 179, article number 108361, doi: <https://doi.org/10.1016/j.future.2025.108361>
13. Zhu, X., Griffith, R., Shah, P., Smirni, E. (2014), "Application-driven dynamic vertical scaling of virtual machines in resource pools", *2014 IEEE Network Operations and Management Symp.*, doi: <https://doi.org/10.1109/NOMS.2014.6838238>
14. Wood, T., Shenoy, P., Venkataramani, A. and Yousif, M. (2009), "Sandpiper: Black-box and gray-box resource management for virtual machines", *Computer Networks*, vol. 53(17), pp. 2923–2938, doi: <https://doi.org/10.1016/j.comnet.2009.04>
15. Meng, X., Pappas, V. and Zhang, L. (2010), "Improving the scalability of data center networks with traffic-aware virtual machine placement", *2010 Proc. IEEE INFOCOM*, doi: <https://doi.org/10.1109/INFOCOM.2010.5461930>
16. Petrovska, I., Kuchuk, H. and Mozhaiev, M. (2022), "Features of the distribution of computing resources in cloud systems", *2022 IEEE 4th KhPI Week on Advanced Technology*, KhPI Week 2022 – Conference Proceedings, 03–07 October 2022, Code 183771, doi: <https://doi.org/10.1109/KhPIWeek57572.2022.9916459>
17. Thorpe, J., Swiler, L.P., Hanson, S., Cruz, G., Tarman, T. Rollins, T. and Debusschere, B.J. (2022), "Verification of Cyber Emulation Experiments Through Virtual Machine and Host Metrics", *ACM International Conference Proc. Series*, pp. 71–80, doi: <https://doi.org/10.1145/3546096.3546115>
18. Kortas, N. and Youssef, H. (2025), "A Bayesian neural network study for virtual machine migration within cloud environment", *Journal of Supercomputing*, vol. 81(16), 1505, doi: <https://doi.org/10.1007/s11227-025-07974-5>

Received (Надійшла) 29.06.2026

Accepted for publication (Прийнята до друку) 02.09.2026

Publication date (Дата публікації) 25.09.2026

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Кучук Ніна Георгіївна – доктор технічних наук, професор, професорка кафедри комп'ютерної інженерії та програмування, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;
Nina Kuchuk – Doctor of Technical Sciences, Professor, Professor of Computer Engineering and Programming Department, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
 e-mail: nina_kuchuk@ukr.net; ORCID Author ID: <http://orcid.org/0000-0002-0784-1465>;
 Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57196006131>.

Пироженко Сергій Станіславович – аспірант кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Serhii Pyrozhenko – PhD candidate of Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
 e-mail: serhii.pyrozhenko@nure.ua; ORCID Author ID: <http://orcid.org/0009-0006-4209-2144>;
 Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=59675302400>.

**Methodology for calculating the overhead of live migration of virtual machines
 based on the NRC resource cube concept**

Nina Kuchuk, Serhii Pyrozhenko

Abstract. Relevance. Current requirements for the energy-efficient and adaptive operation of modern data centers are driven by the rapid expansion of the cloud computing segment. Under these conditions, live virtual machine (VM) migration tools have become a fundamental component of flexible hardware resource allocation. The ability to relocate virtual machines while they are running, without critical interruptions in application availability, enables administrators to address power optimization issues through workload consolidation, eliminate local host overloads, and perform infrastructure maintenance without interrupting business processes. On the other hand, continuous migration inevitably generates resource overhead. The operation of iterative schemes (Pre-Copy) requires continuous tracking of modified memory blocks, their repeated transmission, and final synchronization of CPU contexts, which significantly distorts the resource profile of servers. As a result, sharp increases in hypervisor utilization of computational cores occur, data transmission channel capacity is exhausted, and response time increases. Chaotic or inefficient initiation of such procedures creates risks of mutual interference among neighboring tasks, violations of SLA requirements, and even the risk of transmission loops, when memory pages are rewritten faster than they can be transmitted over the network. **Subject of the study:** processes of generating and methods for calculating overhead (additional load on computational cores, memory, and network channels) during live migration of virtual machines, as well as mathematical models and algorithms for their spatial assessment based on the vector concept of the NRC resource cube. **Objective of the study:** to improve the efficiency of dynamic resource management in a cloud data center and prevent SLA violations by developing a model and method for assessing the overhead of live virtual machine migration based on the vector space of the NRC resource cube. **Research tasks.** To formulate a load-minimization criterion that takes into account server capacity within the NRC model and to implement an algorithm for selecting target hosts in the Load Balancing class. **Research methods.** Methods of vector and multidimensional geometric analysis, methods of systems analysis and systems theory, graph theory, and tree-based data structures. **Results of the study.** It has been demonstrated that the effective capacity of hosts decreases by an amount corresponding to the vector migration overhead, which prevents false system overload during load-balancing operations. The proposed framework provides a mathematical basis for developing adaptive VM relocation algorithms with guaranteed compliance with the boundary conditions of the NRC cube.

Keywords: virtual machine migration; data transmission; simulated annealing algorithm; genetic algorithm; Internet of Things ecosystem.