

М. О. Волк, В. С. Курочкін, А. П. Запорожченко, П. А. Паронікян

Харківський національний університет радіоелектроніки, Харків, Україна

## ГІБРИДНИЙ МЕТОД РОЗПОДІЛУ РЕСУРСІВ В ХМАРНИХ СИСТЕМАХ

**Анотація.** Об'єктом досліджень є процес управління ресурсами в хмарних системах. Мета статті полягає у підвищенні ефективності хмарних систем шляхом розробки метода розподілу віртуальних машин в системах хмарних обчислень. Предметом статті виступають методи розподілу віртуальних машин та завдань в системах хмарних обчислень. У роботі розглядаються різні методи розподілу ресурсів в системах хмарних обчислень. На основі проведеного аналізу зроблено висновки про їх недоліки та переваги. В результаті дослідження отримано гібридний метод розподілу віртуальних машин за комп'ютерними ресурсами та пакетів завдань за віртуальними машинами. Експериментальні результати підтверджують підвищення ефективності запропонованого методу у порівнянні з існуючими за рахунок зниження енергоспоживання та часу виконання і збільшення коефіцієнту використання процесорних модулів.

**Ключові слова:** хмарні системи, хмарні обчислення, управління, метод, розподіл ресурсів, планування, віртуальні машини, віртуалізація, застосунок, програмне завдання

### Вступ

Хмарні обчислення сьогодні пропонують нові методи динамічного аналізу та підтримки великих і віртуалізованих даних та послуг в Інтернеті. Хмарні системи – це структури, які забезпечують доступ за вимогою до розосередженої колекції ресурсів, таких як хости, пристрої зберігання даних, процесорні елементи, спеціалізовані пристрої та інше [1].

Хмарні обчислення пропонують передові тенденції в генерації інформації та знань, які включають інформацію, отриману починаючи з робочого столу користувача до великих статистичних систем обробки та інформації. Часто доступ до цієї інформації реалізується через веб-додатки згідно хмарної моделі надання услуги.

Обчислювальна можливість хмарних систем забезпечуються датацентрами, які розкидані по всьому світу та об'єднані високошвидкісними комп'ютерними мережами. Хмара може бути ефективною розподіленою системою збору, обробки та зберігання інформації. Її послуги розповсюджуються між клієнтами з віддаленими комп'ютерними ресурсами в конкретний момент часу через Інтернет. В роботі [2] припускають, що підходи до керування ресурсами повинні синхронізувати контури управління хмарою як з боку споживачів хмари, так і з боку провайдерів.

Розподіл ресурсів (анг. Resource Allocation, RA) від постачальників до клієнтів дозволяє динамічно перерозподіляти програмні завдання та потоки даних в реальному часі, що надає користувачеві можливість кращого використання наявних потужностей. Розподіл ресурсів у хмарних системах – це процес розподілу віддалених ресурсів через комп'ютерну мережу для виконання відповідних хмарних програм. На високому рівні абстракції модель такого механізму розподілу ресурсів отримала назву Інфраструктура як послуга (IaaS) і надає комп'ютерні ресурси за запитами користувачів завдяки попередньо визначеному або динамічному механізму розподілу ресурсів. Для організації обчислень необхідно вирішити дві задачі: розподілити віртуальні машини за

ресурсами, а програмні завдання за віртуальними машинами.

### Аналіз публікацій та постановка проблеми

Було проведено аналіз методів та алгоритмів розподілу ресурсів на хмарних платформах, які наявні у відкритій літературі, хоча треба відмітити, що ряд комерційних планувальників ресурсів мають закритий код без інформації про реалізовані методи розподілу. Під час розподілу ресурсів у хмарних системах враховуються численні фактори, такі як висока продуктивність, вартість, максимальна пропускна здатність, якість обслуговування SLA (Service-level agreement). Нижче наведено огляд деяких досліджень, що безпосередньо стосуються задачі розподілу ресурсів.

В роботі [3] для оптимізації хмарного планування пропонується техніка, заснована на ретельному дослідженні хмарних систем на рівні IaaS. Її особливістю є гнучкий метод поширення віртуальних машин для найкращого використання фізичних ресурсів з дотриманням політики автоматизованого розвитку.

Необхідність постійного контролю порушення SLA для мінімізації вартості розглядається у [4] на базі множини критеріїв SLA, які слід перевірити у реальному часі під час розподілу послуг. В даній роботі було оцінено механізм планування з кількома параметрами SLA.

Одною з проблем хмарних обчислень є бажання клієнтів отримувати максимум ефективності від ресурсів, які вони використовують. Якщо ця проблема не вирішується провайдером, хмарну систему в цілому можна вважати неефективною. Рішення проблеми можливо тільки динамічними методами, тому що ситуація в хмарі та між клієнтами постійно змінюється. Динамічний розподіл ресурсів і найкраще їх використання протягом встановленого періоду обговорюється у [5]. В роботі обговорюється модель розподілу ресурсів на основі правил, яка проводить розподіл ресурсів відповідно до важливості (пріоритетності) завдання для уникнення неефективного використання ресурсів. Але додаткові дії

обмежують терміни виконання проєктів Real-Time.

В роботі [6] запропонована ідея федеративної обчислювальної та мережевої система (FCN) для обробки спільного розподілу обчислювальних та комунікаційних ресурсів. Обговорюється проблема перевантаження у випадках, коли запит користувача вимагає виділення різноманітних ресурсів (наприклад, процесорів, пам'яті, серверу зберігання та інше). Дослідження сценаріїв, що призводять до перевантажень, відмов та збоїв проведено в роботі [7] привило до розроблення механізму контролю перевантаження. Різні ресурси зазвичай виділяються одночасно для обслуговування запиту користувача, тому метод підвищує ефективне управління ресурсами. В роботі [8] ресурси об'єднуються в систему хмарних обчислень з урахуванням обмежень на спожиту електроенергію. Також враховуються можливості автоматизованого адміністрування та пропускна здатність комунікаційного обладнання. Ще одним плюсом є оцінка здатності забезпечення SLA.

Робота [9] описує модель розподілу ресурсів з використанням SLA на прикладі постачальників хмарних послуг. В моделі використовується поняття рівноваги Неша з теорії ігор щодо розподілу ресурсів між користувачами. Постачальники ресурсів та споживачі розглядаються в рамках гри, де об'єктами виступають мультихмарні послуги.

Існують ряд досліджень щодо оптимізації хмарних ресурсів, засновані на евристичних алгоритмах, підказаних природою. Метод мурашиних колоній [10] використано для призначення ресурсів споживачам. Пошук їжі мурахами визначає ідею даного алгоритму. За допомогою цього підходу шляхи пошуку кращого рішення можуть бути коротшими. Оптимізація методом плодової мушки було запропоновано в роботі [11] для розподілу ресурсів і управління завданнями у хмарі.

Удосконалений метод k-середніх на основі пріоритетів запропоновано у роботі [12]. Завдяки цьому методу розподіл ресурсів можна зробити більш ефективним, тим самим зменшивши витрати. Авторами роботи представлено оптимізаційний алгоритм для зменшення витрат на реалізацію, з виконанням вимог клієнтів. Методика має специфічну направленість на використання в хмарних датацентрах. Для доступу та управління використовується веб-інтерфейс. В основі лежить генетичний алгоритм, що використовується для прогнозування. На основі статистичної інформації прогнозується чи будуть доступні ресурси. Отримані дані прогнозування використовуються для розгортання віртуальних машин.

Незважаючи на те, що загалом запропоновано багато рішень щодо планування, непередбачуваності, споживання енергії, перевантаження, порушення SLA та інші проблеми приводять до ситуації, що жоден з пропонуєваних методів не може досягти оптимального розподілу ресурсів, особливо у реальному часі.

**Мета статті** полягає у підвищенні ефективності хмарних систем шляхом розробки метода розподілу віртуальних машин в системах хмарних обчислень.

## Гібридний метод розподілу ресурсів в хмарних системах

У статті пропонується метод розміщення віртуальних машин та програмних завдань на основі комбінації генетичного алгоритму [13] та алгоритму випадкового лісу (Random forest) [14]. На першому етапі методу генетичний алгоритм створює початкову схему призначення. Алгоритм випадкового лісу навчається за допомогою даних від генетичного алгоритму та реалізує техніку оптимізації. Віртуальні машини розподіляються між фізичними комп'ютерами на першому етапі, а завдання за віртуальними машинами – на другому. Таким самим пропонується гібридна модель, що послідовно використовує генетичний алгоритм та метод випадкового лісу для керування ресурсами хмари. Навчальний набір формується з множини ресурсів, віртуальних та фізичних машин. При застосуванні генетичного алгоритму отримується схема призначення – відображення множини віртуальних машин на множині наявних фізичних машин  $VM > PM$ . Алгоритм включає наступні кроки, що показано на рис. 1.



Рис. 1. Етапи генетичного алгоритму

Крок 1: ініціалізація множин завдань, ресурсів.

Крок 2: обчислення функції допасованості за одним з обраних критеріїв (енергоспоживання, час виконання, коефіцієнт використання ресурсу).

Крок 3: відбір варіантів з найліпшими функціями допасованості та виконаними угодами про рівень послуг (SLA).

Крок 4: схрещування. Проводиться для отримання нових потомків з новим правилом призначення віртуальної машини наявному ресурсу. Для алгоритму обрано 2-точкове схрещування.

Крок 5: мутація – передбачає перебалансування, заміну, переміщення за умов покращення оцінок розподілу.

Крок 6: отримання нового покоління (нової схеми призначення), перехід до етапу еволюції.

Експериментально встановлено кількість еволюцій за критеріями часу виконання та якості отриманої схеми розподілу. Рекомендується кількість еволюцій знаходиться у діапазоні 30-40.

Розподіл віртуальних машин між фізичними виконується наступним чином. Вхідними даними виступають двоє батьків із схеми розподілу. Кожен з батьків має хромосоми, які характеризують параметри віртуальної та фізичної машин (гени). Результатом є нащадки, кожен з яких прагне покращити співвідношення генів (відповідних характеристик ресурсу та віртуальної машини). Мутація відбувається тоді, коли один з потомків покращує популяцію (схему призначення). Критерієм завершення генетичного алгоритму встановлена фіксована кількість циклів. Подальше застосування алгоритму не дає значного росту ефективності, але призводить до збільшення часу планування.

На другому етапі методу застосовується алгоритм випадкового лісу. Формування більшої кількості дерев (можливих схем розподілу завдань за віртуальними машинами) означає більшу стійкість та гарантує близькість до оптимального результату. Кожен вузол випадкового лісу обирає невелику кількість параметрів для оцінки розподілу, для яких обчислюється покращення характеристик.

На кожній ітерації отримується прогноз з кожного дерева, а потім обирається найкращий варіант шляхом "голосуванням більшістю" (для одного критерію) або "голосуванням за ефективність" (у випадку сукупного критерію ефективності). Виконання алгоритму випадкового лісу починається після встановлення віртуальних машин на фізичні. Фізична машина може отримати одне, два або більше віртуальних машин для виконання, віртуальна машина може виконувати одне, два або більше програмних застосувань з завдань. При цьому можливе перенавантаження фізичної машини. Для вирішення цього типу сценарію використовується техніка міграції віртуальних машин та програмних застосувань. У випадку перевантаження фізичної машини, деякі з віртуальних машин переносяться на інші фізичні, які мають менше навантаження, забезпечуючи балансування навантаження на всі фізичні машини в центрі обробки даних [15–17]. Для перевірки запропонованого методу використано середовище моделювання CloudSim 4.0, що пропонує моделі надання хмарних послуг та віртуалізовані ресурси. Моделювання проводилося на 32 віртуальних машинах, які працювали на платформі Azure (фірми Microsoft) в пакеті Standard\_B4ms, що розгорталася на фізичних машинах з процесорами Intel® Xeon® Platinum 8370C (Ice Lake) та базовою продуктивністю процесора 22%.

В якості першого критерія ефективності використовувалось енергоспоживання. Фізичні машини споживають енергію, а уся спожита енергія дорівнює сумі усіх пристроїв споживання енергії, які обслуговують виконання обчислень. Наступним критерієм був час виконання завдань. З точки зору провайдера та користувачів хмарних послуг, дуже важливо виконати запити на обчислення якомога швидше. На рис. 2 показано порівняння енергоспоживання щодо розробленого та існуючих методів, а саме, на основі мурашиного алгоритму та оптимізації рою частинок (PSO).

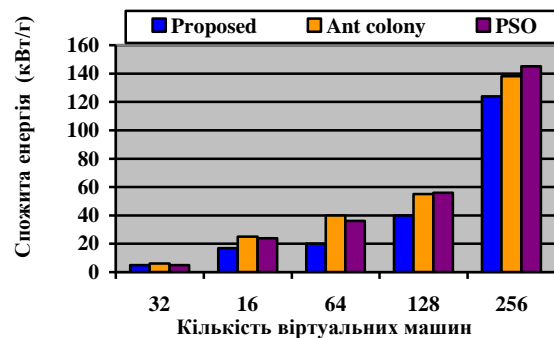


Рис. 2. Оцінка енергоспоживання

Кожен вузол має окремий процесор з чотирма ядрами, оперативну та зовнішню пам'ять. В експериментах для різних методів розподілу ресурсів використовувалась одна конфігурація комунікаційного обладнання для передачі даних та конфігурація пакета програмних завдань на обчислення. З отриманих даних можна зробити висновок, що енергоспоживання у випадку використання розробленого методу нижче за випадків використання існуючих методів. Зниження енергоспоживання досягає 6-8%.

На рис. 3. показана залежність часу виконання пакету завдань (містив 500 програмних компонентів) на віртуальних машинах.

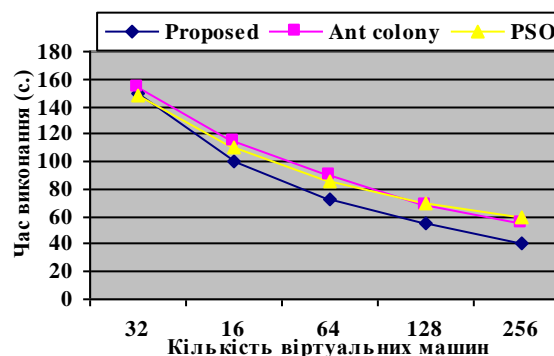


Рис. 3. Оцінка часу виконання

Аналіз результатів показує, що запропонований метод показує менший час виконання в порівнянні з існуючими алгоритми. Середній показник зменшення часу склав 10-12%.

## Висновки

Авторами запропоновано гібридний метод розподілу віртуальних машин в системах хмарних об-

числень. Серед найбільш важливих можливостей, що надає метод, є зменшення енергоспоживання та часу виконання завдань при встановленому рівні SLA. Отримані результати є важливими, тому що стандартні методи управління розподіленими хмарними системами є дуже ресурсозатратними. Якщо порівняти дані експериментів, можна зробити висновки, що час обчислень зменшується більшими темпами ніж енергоспоживання. Це можна поясни-

ти: обчислювальна складність завдань не змінюється, тому зменшення часу виконання відбувається шляхом оптимального використання процесорів, які збільшують кількість обчислень на одиницю часу, що призводить до зростання енергоспоживання.

Напрямок подальших досліджень методу є розроблення модифікацій методу з урахуванням таких критеріїв ефективності як вартість обчислень, забезпечення функціональної стійкості до відмов.

## СПИСОК ЛІТЕРАТУРИ

1. Kim W. Cloud computing architecture. *International Journal of Web and Grid Services*. Vol. 9, No.3. 2013. P. 287-303. DOI: <https://doi.org/10.1504/IJWGS.2013.055724>
2. Mary M., Mahalakshmi D. An extensive survey on resource allocation mechanisms in cloud computing. *PalArch's Journal of Archaeology of Egypt/Egyptology*. Vol. 17. No 9. 2020. P. 45–56.
3. Mohan N., Raj E. Resource allocation techniques in cloud computing—research challenges for applications. *Computational Intelligence and Communication Networks*, Vol. 6. 2021. P. 101–123. DOI: 10.1109/CICN.2012.177
4. Prodan R., Ostermann R. A survey and taxonomy of infrastructure as a service and web hosting cloud providers. *Grid Computing*. Vol. 10. 2019. P. 45–57. DOI: 10.1109/GRID.2009.5353074
5. Filimonchuk T., Volk M., Ruban I., Tkachov V. Development of information technology of tasks distribution for grid-systems using the GRASS simulation environment. *Eastern-European Journal of Enterprise Technologies. Information and controlling system*, 2016. Vol. 3/9 (81). pp. 45–53. DOI: <https://doi.org/10.15587/1729-4061.2016.71892>
6. Kumar K., Feng J., Nimmagadda Y., Lu Y. Resource allocation for real-time tasks using cloud computing. *Computer Communications and Networks*. Vol. 3. 2011. P. 21–30. DOI: 10.1109/ICCCN.2011.6006077
7. Гора М., Волк М. Моделі управління ресурсами для забезпечення функціональної стійкості процесу розподілених обчислень. *Вісник Херсонського національного технічного університету*. 2023. No 4(87). С. 244-251. DOI: <https://doi.org/10.35546/kntu2078-4481.2023.4.28>
8. Mamchych O., Volk M. Smartphone Based Computing Cloud and Energy Efficiency. 12th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece. 2022. P.1-5, DOI: 10.1109/DESSERT58054.2022.10018740
9. Chen J., Tsai C., Lu S. Resource reallocation based on SLA requirement in cloud environment. *IEEE Transactions on Services Computing*, Vol. 25. 2020. P. 89–102. DOI: 10.1109/ICEBE.2015.70
10. Yang Z., Liu M., Xiu J., Liu C. Study on cloud resource allocation strategy based on particle swarm ant colony optimization algorithm. *Cloud Computing and Intelligence Systems*. Vol. 2. 2012. P. 67–80. DOI: 10.1109/CCIS.2012.6664453
11. Zheng X. A Pareto-based fruit fly optimization algorithm for task scheduling and resource allocation in cloud computing environment. *IEEE Transactions on Services Computing*. Vol. 12. 2016. P. 112–121. DOI: 10.1109/CEC.2016.7744219
12. Goutam S., Yadav A. Preemptible priority-based dynamic resource allocation in cloud computing with fault tolerance. *International Journal of Communication Networks*. Vol. 12. No. 3. 2015. P.67–76. DOI: 10.1109/ICCN.2015.54
13. Tseng F., Wang X., Chou H., Leung V. Dynamic resource prediction and allocation for cloud data centre using the multi-objective genetic algorithm. *IEEE Systems Journal*. Vol. 12. 2018. P.1688–1699. DOI: 10.1109/JSYST.2017.2722476
14. Wei L. Genetic Algorithm Optimization of Concrete Frame Structure Based on Improved Random Forest. *International Conference on Electronics and Devices, Computational Science (ICEDCS)*. Marseille, France. 2023. P. 249-253, DOI: 10.1109/ICEDCS60513.2023.00051
15. Коваленко А. А., Кучук Г. А. Методи синтезу інформаційної та технічної структур системи управління об'єктом критичного застосування. *Сучасні інформаційні системи*. 2018. Т. 2, № 1. С. 22–27. DOI: <https://doi.org/10.20998/2522-9052.2018.1.04>
16. Свиридов А. С., Коваленко А. А., Кучук Г. А. Метод перерозподілу пропускну здатності критичної ділянки мережі на основі удосконалення ON/OFF-моделі трафіку. *Сучасні інформаційні системи*. 2018. Т. 2, № 2. С. 139–144. DOI: <https://doi.org/10.20998/2522-9052.2018.2.24>
17. Ivanisenko I.M., Volk M.O. Simulation methods for load balancing in distributed computing. *Proceedings of IEEE East-West Design & Test Symposium (EWDTS'2017)*, Novi Sad, Serbia, September 27 – October 2, 2017. P. 690-695. DOI: 10.1109/EWDTS.2017.8110078

Received (Надійшла) 27.02.2024

Accepted for publication (Прийнята до друку) 24.04.2024

### Hybrid Method of Recourse Allocation in Cloud Systems

Maksym Volk, Vladyslav Kurochkin, Anton Zaporozhchenko, Papin Paronikian

**Abstract.** The object of research is the process of resource management in cloud systems. The purpose of the article is to increase the efficiency of cloud systems by developing a method for distributing virtual machines in cloud computing systems. The subject of the article is the methods of distribution of virtual machines and tasks in cloud computing systems. The work considers various methods of resource allocation in cloud computing systems. Based on the analysis, conclusions were made about their disadvantages and advantages. As a result of the study, a hybrid method of distributing virtual machines by computer resources and task packages by virtual machines was obtained. The experimental results confirm the improvement of the efficiency of the proposed method in comparison with the existing ones due to the reduction of energy consumption and execution time and the increase of the utilization ratio of processor modules.

**Keywords:** cloud systems, cloud computing, management, method, resource allocation, planning, virtual machines, virtualization, application, software task.