

КОМПЛЕКСНИЙ МЕТОД ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ОБРОБКИ НАВАНТАЖЕННЯ У РОЗПОДІЛЕНИХ ЦОД ОПЕРАТОРІВ ЗВ'ЯЗКУ

Глоба Л.С., Райчук А.А., Прокопець Н.А.
*Навчально-науковий Інститут телекомунікаційних
систем КПІ ім. Ігоря Сікорського, Україна
E-mail: raaukr@gmail.com*

A COMPLEX METHOD OF INCREASING THE EFFICIENCY OF LOAD PROCESSING IN DISTRIBUTED DATA CENTERS OF COMMUNICATION OPERATORS

The article proposes a solution to the problem of preventing a rapid increase in electricity consumption in remote data centers by increasing the energy efficiency and productivity of information and communication network (ICM) load maintenance while meeting the requirements for the availability of its maintenance system by applying a comprehensive method of energy-efficient load maintenance.

The demo aims to analyze the proposed comprehensive method, which allows us to claim that it meets the requirements for the availability of the service system and gives an efficiency gain of 15.722% compared to Backfill and 88.887% compared to Round Robin.

Розвиток сучасних інформаційно-комунікаційних технологій, зокрема технологій віртуалізації, зумовив появу тенденції програмної реалізації апаратних функцій, яка сприяла розвитку технологій і концепцій: NFV, SDN, Network Slicing, Edge Computing, bDDN. Ці технології вимагають значного обсягу обчислень, які виконуються у віддалених Дата Центрах (у Cloud середовищі) і, відповідно, споживають значну кількість енергії під час обробки навантаження. Функціонування інфраструктури територіально розподілених ЦОД є енерговитратним та вносить суттєвий вклад у загальне енергоспоживання сучасних інформаційно-комунікаційних мереж (ІКМ). У зв'язку з цим особливої актуальності набуває вирішення проблемних питань із забезпечення енергоефективності розподілених ЦОД. За даними досліджень [1, 2], частка споживання електроенергії такими системами станом на 2020 рік становила близько 1% усієї електроенергії, що виробляється у світі.

Згідно глобальної оцінки споживання енергії центрами обробки даних на 2030 рік загальне середнє енергоспоживання дата центрів у світі складатиме 848 TWh, а максимальне значення може скласти близько 1929 TWh [3].

Існує велика кількість підходів з підвищення енергоефективності роботи дата центрів, але вони не враховують одночасного забезпечення трьох основних показників: продуктивності, енергоефективності та дотримання вимог SLA.

У статті пропонується вирішення задачі запобігання стрімкому збільшенню споживання електроренергії у віддалених дата центрах за рахунок підвищення енергоефективності та продуктивності обслуговування навантаження інформаційно-комунікаційної мережі (ІКМ) при виконанні вимог щодо доступності системи його обслуговування за рахунок застосування комплексного методу енергоефективного обслуговування навантаження.

Демо націлено на аналіз запропонованого комплексного методу, який дозволяє стверджувати, що він забезпечує виконання вимог щодо доступності системи обслуговування та дає вииграш за критерієм ефективності на 15,722% у порівнянні із Backfill та на 88,887% у порівнянні з Round Robin.

Огляд підходів щодо енергоефективного комп'ютерингу. Згідно з рекомендаціями Міжнародного союзу електрозв'язку ITU-T L.1300 [4], загальні підходи до підвищення енергоефективності обробки навантаження ІКМ такі:

- підбір енергоефективного комп'ютерного та телекомунікаційного обладнання, його

регулярна атестація (аудит);

- розгортання сервісів з використанням підходів віртуалізації;
- розробка ефективного програмного забезпечення;
- послуги з виведення з експлуатації невикористаного обладнання та зупинки обладнання у разі простою;
- консолідація поточних послуг;
- постійний моніторинг енергоспоживання системи.

В роботі [5] запропоновано розподілити підходи щодо енергоефективної обробки навантаження на статичні та динамічні підходи. Статичні підходи не враховують динамічну змінюваність навантаження. Вдосконалення енергоефективності в динамічних підходах на рівні апаратного забезпечення – є складним та водночас дорогим для реалізації процесом, тому є доцільним вдосконалювати енергоспоживання на рівні програмного забезпечення.

Проведений огляд підходів щодо енергоефективного розподілу навантаження в ІКМ показав, що алгоритм енергоефективного планування навантаження Backfill більше підходить для вирішення задачі підвищення енергоефективності обробки даних з огляду на концепцію енерго-пропорційних розрахунків. Однак існує проблема вдосконалення Backfill з метою покращення його роботи в умовах недостатнього навантаження на систему, а також пошуку балансу між енергоефективністю та продуктивністю обробки даних в умовах помірного навантаження. Крім того, пропонується провести попередню сертифікацію кожного сервера кластера для отримання реальної залежності енергоспоживання від навантаження для кожного обчислювального вузла для індивідуального визначення моделі енергоспоживання кожної машини окремо. Однак ефективність цього підходу зменшується у випадку невисокої інтенсивності вхідного навантаження, оскільки він не забезпечує відсутність простою в умовах збитковості обчислювальних ресурсів.

Окремим виявленим недоліком існуючих методів є те, що жоден із них не вирішує задачу підвищення ефективності обслуговування навантаження з урахуванням усіх трьох виділених показників ефективності (енергоефективності, продуктивності та доступності системи), що є ключовими для обслуговування навантаження ІКМ.

Визначення проблеми та цілі дослідження. У дата центрах надається велика кількість послуг, що мають різні витрати енергоспоживання. Для обробки навантаження використовуються кластери. У залежності від задач вони можуть складатись з різних компонент, таких, як процесор, оперативна пам'ять, графічний процесор та жорсткий диск. Всі вони так, чи інакше використовуються електроенергію для їх роботи.

Тому основною метою є розробка програмне забезпечення, яке б дозволило:

знизити загальне споживання електроенергії, забезпечити дотримання вимог щодо SLA та збільшити енергоефективність обробки навантаження за рахунок збільшення кількості оброблених завдань, використовуючи ту саму кількість спожитої потужності.

Критерії ефективності запропонованого комплексного методу задаються специфікаціями та рекомендаціями Міжнародної спілки електрозв'язку та Європейського інституту телекомунікаційних стандартів, які визначають основні задачі, що формують обчислювальне навантаження сучасної ІКМ. Аналіз вимог щодо обслуговування цього навантаження дозволив виділити основні показники ефективності процесу обслуговування навантаження ІКМ: енергоефективність та продуктивність обслуговування, а також коефіцієнт готовності системи як показник її доступності. На основі цих показників сформульовано критерій ефективності обробки обчислювального навантаження в системі [6].

Енергоефективність обробки даних E_{Σ} – обсяг корисного навантаження ω обробленого системою (серверним кластером) з N вузлів при споживанні обсягу енергії W . При обробці навантаження в N вузлах сумарний обсяг енергії розраховується як $W = \sum_{j=1}^N W_j$, а енергоефективність обробки даних відповідно визначається як:

$$E_{\Sigma} = \frac{\omega}{\sum_{j=1}^N W_j} \left[\frac{FLOP}{\text{Вт*год}} \right]. \quad (1)$$

Чим більший цей показник, тим більш енергоефективною є система, оскільки вона може

обробити більшу кількість навантаження при меншому обсязі спожитої енергії.

Продуктивність обробки даних C_Σ – це обсяг навантаження, що обробляється системою за одиницю часу:

$$C_\Sigma = \frac{\omega}{T} \left[\frac{FLOP}{c} \right], \quad (2)$$

де ω – обсяг корисного навантаження, що обробляється системою; T – час обробки.

Показник продуктивності фактично визначає швидкість обслуговування навантаження обчислювальною системою.

Коефіцієнт готовності системи p_{avail} як показник її доступності.

Коефіцієнт готовності в імовірнісному визначенні (тобто у випадку визначення його через математичні сподівання часу роботи та простою), пов'язаний з імовірністю втрати запиту згідно з виразом:

$$p_{avail} = 1 - p_{loss}. \quad (3)$$

Якщо з певних причин (наприклад, через відсутність вільних ресурсів обробки даних) запит було втрачено, він потребує повторної обробки, що негативно впливає на QoE [7], тобто якість клієнтського досвіду.

Сформулюємо **критерій ефективності** для досліджуваної системи. У теорії прийняття рішень критерій ефективності – це правило, що формується на основі вибраних показників ефективності та дозволяє зробити цілеспрямований вибір з множини альтернатив [8]. При цьому виділяють критерії оптимальності (що використовується для прийняття оптимальних рішень згідно з концепцією максимізації корисності) та критерій придатності (що використовується для прийняття задовільних рішень згідно з концепцією обмеженої раціональності).

З огляду на вищесказане, критерій оптимальності для досліджуваного процесу обслуговування навантаження в інформаційно-комунікаційній мережі визначимо у вигляді:

$$K_{opt} = E_\Sigma * C_\Sigma \rightarrow \max, \quad (4)$$

при $p_{avail} \geq p_{avail_{SLA}}$.

Тобто у множині можливих конфігурацій системи обслуговування навантаження найбільш ефективною є така, що дозволяє отримати максимальне значення добутку показників енергоефективності E_Σ та продуктивності C_Σ при задовільному значенні коефіцієнта готовності системи p_{avail} . Задоволення критерію ефективності K_{opt} для кожного серверного кластера, що бере участь в обслуговуванні навантаження ІКМ (а саме, задач SDN та NVF, Network Slicing, Edge Computing та bDDN) позитивно впливає на продуктивність та енергоефективність інформаційно-комунікаційної мережі в цілому.

Короткий опис запропонованого підходу. Запропонований алгоритм обробки робочого навантаження складається з 4 основних кроків:

- 1) атестація та сортування вузлів;
- 2) визначення шаблонів масштабування;
- 3) горизонтальне масштабування та енергоефективне планування;
- 4) інтелектуальне прогнозування відхилень робочого навантаження та перебудова патернів.

На цих етапах ми прагнемо відповісти на такі питання відповідно:

- 1) Які обчислювальні вузли слід використовувати більш інтенсивно?
- 2) Скільки активних вузлів потрібно для обслуговування статистичного навантаження?
- 3) Як розподілити навантаження між поточними активними вузлами?
- 4) Як динамічно впоратися з відхиленнями робочого навантаження від статистичного?

Експериментальний аналіз. Для визначення ефективності та працездатності розробленого програмного забезпечення було проведено натурний експеримент в одному з дата центрів НАН України.

Запропонований підхід було випробувано під час проведення натурального експерименту

в одному з дата центрів України. Були змодельовані умови, під час яких надходить на обробку непередбачуваний вхідний потік заявок. Для опрацювання потоку використано 6 вузлів, а саме 3 фізичні та 3 віртуальні. Віртуальні вузли було створено з використанням VirtualBox. На всіх вузлах встановлені дистрибутиви Linux. Один з комп'ютерів працював як головний вузол, що керує процесами розподілу та їх координуванням. Під час процесу підготовки відбувається встановлення необхідних програм, бібліотек та утиліт. Далі виконується налаштування мережі та розмежування прав доступу. Під час експерименту збирались необхідні для аналізу дані та заносились у файли формату .csv.

При аналізі результатів експерименту було визначено, що комплексний підхід у порівнянні з кожним показником по підвищенню ефективності дійсно забезпечує підвищення енергоефективності та демонструє гарний показник збереження потужності. Результати порівняння приведені в таблиці.

Таблиця 1. Результати порівняння.

Виграш за показником енергоефективності	9,953% порівняно з Backfill 26,382% порівняно з Round Robin
Виграш за показником продуктивності	5,593% порівняно з Backfill 49,458% порівняно з Round Robin
Виграш за показником доступності	-2,49% порівняно з Backfill 17,709% порівняно з Round Robin
Виграш за критерієм ефективності	15,722% порівняно з Backfill 88,887% порівняно з Round Robin при дотриманні вимог щодо доступності системи

Висновки. У даному дослідженні представлено ефективність комплексного методу підвищення енергоефективності та продуктивності програмного забезпечення розподілу навантаження у вузлах ІКМ при виконанні вимог щодо доступності системи обслуговування навантаження.

Література

1. Повторне калібрування оцінок використання енергії глобального центру обробки даних / E. Masanet et al. Наука. 2020. Vol. 367, no. 6481. P. 984–986. – Режим доступу до ресурсу: <https://doi.org/10.1126/science.aba3758> (date of access: 03.07.2022).
2. Ринок ЦОД і технологічні тенденції. Презентація Power Electronics [Електронний ресурс] // Yole Développement, APEC. – 2016. – Режим доступу до ресурсу: https://www.slideshare.net/Yole_Developpement/data-center-market-and-technology-trends-power-electronics-presentation-held-at-apec-2016-from-yole-dveloppement.
3. Проблема з потужністю центрів обробки даних – Режим доступу до ресурсу: <https://www.datacenterdynamics.com/es/features/el-problema-con-las-cifras-sobre-la-energ%C3%ADa-del-centro-de-datos/>
4. Міжнародна спілка електрозв'язку. ITU-T L.1300 - Best practices for green data centres [Електронний ресурс] / Міжнародна спілка електрозв'язку. – 2014. – Режим доступу до ресурсу: <https://www.itu.int/rec/T-REC-L.1300/en>.
5. Надійність та енергоефективність у системах хмарних обчислень: Огляд і таксономія / Ю. Шарма та ін. Журнал мережевих і комп'ютерних програм. 2016. Vol. 74. P. 66–85. – Режим доступу до ресурсу: <https://doi.org/10.1016/j.jnca.2016.08.010>.
6. Оптимізація обчислень у дата-центрі за критерієм енергоефективності – Режим доступу до ресурсу: <https://ela.kpi.ua/handle/123456789/39712>
7. J. Berger QoS та QoE у мережах 5G Розвиток додатків і вимірювань : Семінар QSDG // Тестування мобільної мережі Rhode & Schwarz, Сінгапур, 19-21 August 2019.
8. Чижова, Є. Н. Управлінські рішення: навч. посібник/Є.М. Чижова, Д.Г. Сазонов, М.С. Чижов. - Білгород: Вид-во БДТУ, 2009. - 202 с.