

Н. Г. Кучук, І. С. Зиков, В. І. Панченко

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

МЕТОД РОЗГАЛУЖЕННЯ ЗАПИТІВ ДО СХОВИЩ ДАНИХ СИСТЕМ, ЩО МАЮТЬ ГІПЕРКОНВЕРГЕНТНУ ІНФРАСТРУКТУРУ

Актуальність дослідження. В процесі функціонування комп'ютерної мережі гіперконвергентної архітектури за рахунок централізованого управління збільшується час доступу до сховищ даних. Але для деяких підсистем, особливо для тих, що повинні функціонувати у режимі, наближеному до режиму реального часу, дані показники є дуже суттєвими. **Метою статті** є розробка підходу до зменшення часу доступу до сховищ даних у гіперконвергентному середовищі шляхом розгалуження інформаційних потоків. **Результати дослідження.** Запропонований метод є двохетапним та складається із послідовності кроків знаходження можливих маршрутів та знаходження оптимального із знайдених варіантів розподілу каналів. В результаті знайдено найкращий варіант розподілу пакетів по каналах зв'язку серед таких, що задовольняють сформульованим умовам. Зокрема, максимальний час передачі за допомогою одного з каналів обраного варіанту розподілу не перевищує час, виділений для передачі необхідного обсягу даних та виконані обмеження за продуктивністю та вартістю. **Висновки.** Застосування підходу дозволяє істотно зменшити час передачі даних і витрати на оренду каналів, а також поліпшити живучість мережі.

Ключові слова: гіперконвергентна архітектура; сховище даних, час виконання запиту.

Вступ

Актуальність завдання. Розробники великих систем в даний час все більше схиляються до конвергентних і гіперконвергентних рішень для базових комп'ютерних мереж, які за рахунок централізації управління суттєво зменшують витрати на обслуговування [1]. При такому технологічному рішенні передбачається об'єднання пам'яті, обчислювальних, програмних і мережевих ресурсів в пул, заздалегідь об'єднаних для роботи в дата-центрі [2]. Управління відбувається через загальну консоль адміністрування [3]. Перевагами гіперконвергентної інфраструктури є такі:

- 1) спрощення інфраструктури управління – централізація управління сервером, мережевих ресурсів і сховищами даних, що дозволяє оптимізувати повсякденне обслуговування;
- 2) масштабна ємність сховищ даних – усі поширені комутаційні мережі і протоколи вбудовані в гіперконвергентну інфраструктуру, що дозволяє додати в неї додаткові гігабайти простіше і швидше;
- 3) швидка підготовка і виділення ресурсів – скорочення часу підготовки і виділення ресурсів;
- 4) більш швидке реагування інформаційної технології забезпечує гнучкість для реагування на зміни на ринку і зміни пріоритетів у бізнесі;
- 5) спрощений перехід в хмару спрощує впровадження приватних або гібридних хмар;
- 6) підвищений контроль забезпечує одночасне керування кількома функціями і пристроями.

Постановка завдання. В процесі функціонування комп'ютерної мережі (КМ) гіперконвергентної архітектури за рахунок централізованого управління збільшується час доступу до сховищ даних (СД) Але для деяких підсистем, особливо для тих, що повинні функціонувати у режимі, наближеному до режиму реального часу, дані показники є дуже суттєвими. Враховуючи особливості гіперконвергентних систем серед багатьох методів, призначених для зменшення часу доставки інформації обрано

методи, що пов'язані із розгалуженням потоку пакетів даних. Таким чином, завдання підвищення оперативності процесу доступу до сховищ даних може бути сформульованим таким чином: необхідно запропонувати метод розгалуження потоку даних по каналах зв'язку системи, котрий враховує особливості гіперконвергентної системи.

Аналіз літератури. Для вирішення поставленого завдання пропонується багато різних методів [4–19]. Однак, всі перераховані методи мають порівняно невисоку обчислювальну ефективність, що обмежує рішення поставленого завдання у гіперконвергентному середовищі. У зв'язку з цим, виникає необхідність в розробці відповідного методу.

Метою статті є розробка підходу до зменшення часу доступу до сховищ даних у гіперконвергентному середовищі шляхом розгалуження інформаційних потоків.

Результати досліджень

Систему доступу до сховищ даних (СДСД) гіперконвергентного середовища можна представити у вигляді ієрархічного неорієнтованого графа G (рис. 1).

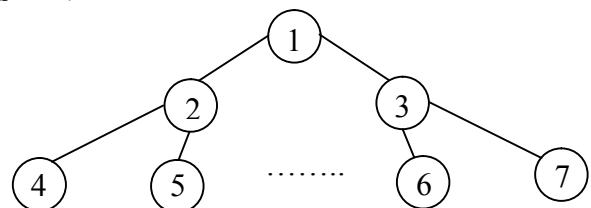


Рис. 1. G – граф системи доступу до сховищ даних гіперконвергентного середовища

Кожне ребро цього графа є певною підмережею. Дійсно, якщо вузли СДСД знаходяться на досить великій відстані один від одного, магістральний тракт між ними є деякою підмережею, що складається із множини вузлів (комутаторів) і трактів між ними. Кожну із таких підмереж можна представити у вигляді неорієнтованого графа (наприклад,

граф G_{13} на рис. 2 відповідає ребру між вершинами 1 і 3 графа G , це потокова мережа із джерелом s , що відповідає вершині 1 та стоком t , що відповідає вершині 3). Вимоги, що пред'являються до СДСД, також пред'являються і до таких підмереж.

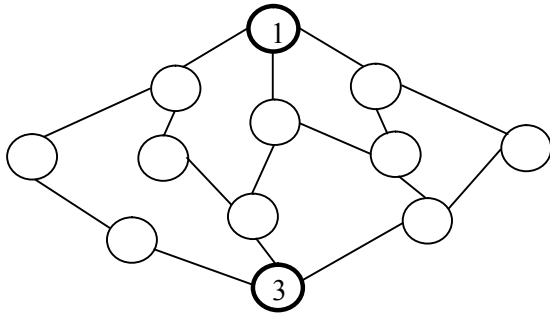


Рис. 2. G_{13} – граф підмережі

На етапі формування трафіку окремо взятої підмережі зазвичай стоять два завдання, які виходять із вимог оперативності та живучості:

1) необхідно знайти оптимальний маршрут передачі даних при заданій пропускній здатності тракту і вузлів підмережі так, щоб час передачі був не більше заданої величини;

2) підмережа повинна бути m -зв'язковою (в найпростішому випадку $m = 2$), тобто має бути мінімум m маршрутів, які відповідають попередній умові і не мають спільних вузлів і тракту.

Для вирішення першого завдання існує багато різних методів пошуку маршруту, наприклад, методи Дейкстри або Флойда [5, 6]. Для вирішення другого завдання можна використовувати запропонований надалі метод пошуку m -найкоротших шляхів з вихідної вершини джерела s до вершини стоку t .

Розглянемо покроково алгоритм цього методу.

Крок 0. Покладемо $k = 1$, тобто будемо шукати найкоротший шлях із джерела s до стоку t . Перед початком виконання алгоритму всі вершини і дуги графа визначаємо як невиділені. Кожній вершині x надається вага $d(x)$ таким чином: $d(s) = 0$ і $d(x) = \infty$ для всіх x , відмінних від джерела s (надалі для виділених вершин вага буде дорівнює довжині найкоротшого шляху із s до x , що включає тільки виділені вершини).

Крок 1. Виділимо вершину s і покладемо $y = s$ (y – остання з виділених вершин).

Крок 2. Для всіх невиділених вершин x , інцидентних з вершиною y , перерахуємо вагу $d(x)$ у таким чином:

$$d(x) = \min(d(x), d(y) + r(x, y)), \quad (1)$$

де $r(x, y)$ – вага відповідної дуги.

Виділяємо ту із вершин x , для якої величина $d(x)$ є найменшою і покладемо $y = x$. Якщо ж для всіх розглянутих вершин x вага $d(x) = \infty$, то подальший пошук неможливий і алгоритм закінчує роботу.

Крок 3. Якщо $y \neq t$, то повертаємося до виконання кроку 2, в іншому випадку k -й найкоротший шлях з вершини s до вершини t знайдений.

Крок 4. Якщо $k < m$, то покладемо $k = k + 1$, з графа G видалимо вершини, що входять в найко-

ротший маршрут (крім вершин s і t) і переходимо до кроку 1. При $k = m$ необхідний результат досягнутий і алгоритм закінчує роботу.

Зауважимо, що кожен тракт (i, j) між двома окремо взятими вузлами i і j є сукупністю декількох підтрактів – фізично незалежних ліній зв'язку, таких як повітряні лінії зв'язку, підземний броньований кабель тощо. З метою підвищення живучості та зменшення часу, що витрачається на передачу заданого обсягу даних, пакети можна відправляти не по одному, а по декількох каналах незалежних ліній зв'язку. В основу реалізації даного методу покладений розподіл пакетів між декількома каналами зв'язку. Відправник посилає пакети в канали по черзі: перший пакет в перший канал, другий – у другий і т.д. Одержувач приймає пакети з різних каналів і, використовуючи відповідний механізм, вибудовує їх у порядку відправлення [11]. Тому необхідна чітка синхронізація приймальних і передавальних пристроїв. Для цього в потоки пакетів з призначеною для користувача інформацією вставляються спеціальні пакети, як, наприклад, в технології ATM ICP (IMA Control Protocol, керуючий протокол IMA). Ці пакети нарівні з іншою, призначеною для користувача інформацією, утворюють кадри IMA. Хоча швидкості каналів, що входять в одну групу, є рівними, не існує гарантії, що вставлені пакети будуть отримані одночасно. Тому в більш швидкі канали для вирівнювання швидкостей надсилаються спеціальні порожні клітинки. Мультиплексори одержувача видаляють їх, перш ніж передавати в мережу. Розподіл потоків пакетів по різномітних різношвидкісних каналах незалежних ліній зв'язку дає можливість підвищити живучість мережі, зменшити час доставки пакетів і плату за оренду каналів.

Розглянемо процес передачі даних із використанням наведеного вище алгоритму. Для цього введемо такі позначення: V_{IU} – обсяг передаваних із сховища даних; V_{SU1} – службова інформація для формування комірок; V_{SU2} – службова інформація для формування кадрів; V_{SU} – потрібний обсяг службової інформації (рис. 3, $V_{SU} = V_{SU1} + V_{SU2}$).

Множину каналів, наявних в даний момент і маючих можливість для отримання даних із сховища, позначимо як

$$I = \{i \mid i = \overline{1, n}\}, \quad (2)$$

а всі можливі варіанти розподілу каналів (I_k), отримані при виконанні наведеного алгоритму, є підмножинами цієї множини, тобто

$$I_k \subset I. \quad (3)$$

Нехай також відомі такі значення: V_i – обсяг даних, які будуть передані по i -му каналу, $i \in I_k$; T_i – час передачі відповідного обсягу V_i i -м каналом; S_i – швидкість передачі даних i -м каналом; C_i – вартість оренди каналів; T_{IU} – час, виділений для передачі обсягу V_{IU} ; C_{IU} – кошти (фінансування), виділені для оренди каналів.

Зазвичай, значення T_i відрізняються одне від одного. Визначимо мінімальне серед них

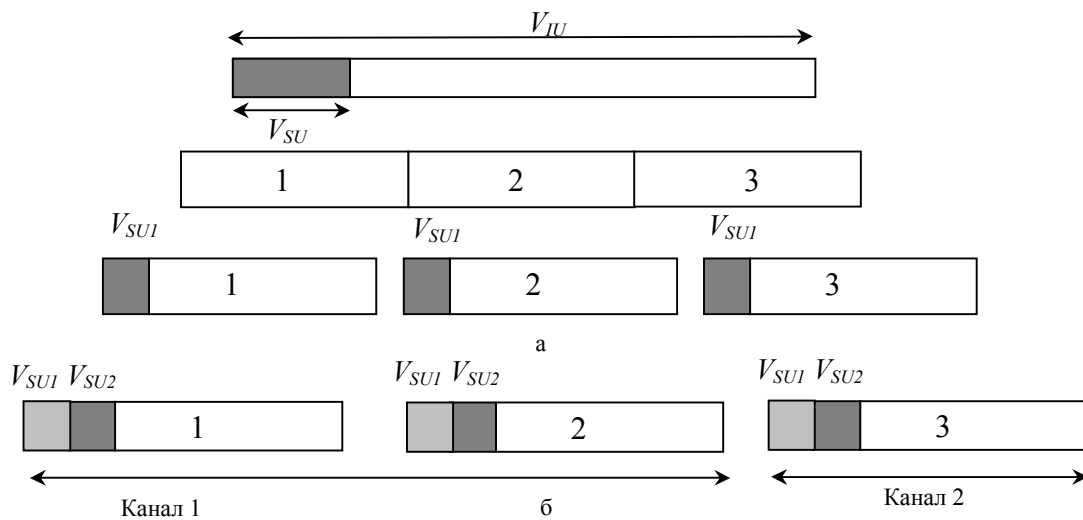


Рис. 3. Формування комірок (а) та кадрів (б)

$$T_{min} = \min_{i \in I_k} (T_i). \quad (4)$$

Це надає можливість отримати значення затримки при передачі пакетів кожним i -м каналом:

$$\Delta T_i = T_i - T_{min}. \quad (5)$$

Отже необхідно вирішити таку оптимізаційну задачу: знайти найкращий варіант розподілу пакетів по каналах зв'язку серед таких, що задовольняють умові про те, що максимальний час передачі за допомогою одного з каналів обраного варіанту розподілу I_k не перевищує час, виділений для передачі необхідного обсягу даних V_{IU} , тобто

$$\max_{i \in I_k} T_i \leq T_{IU}. \quad (6)$$

Крім того, затримки по цих каналах повинні бути зведені до мінімуму, тобто ($\Delta T_i \rightarrow 0$).

$$\max_{i \in I_k} \Delta T_i \rightarrow 0 \quad \forall i \in I_k. \quad (7)$$

Ще одно обмеження: кошти, що витрачаються на оренду каналів не перевищували виділеного фінансування

$$\sum_{j=1}^n C_j \leq C_{IU}, \quad j \in \overline{1, n}. \quad (8)$$

В ході вирішення отриманої оптимізаційної задачі потрібно, щоб відібрані варіанти задовольняли такій умові:

$$\sum_{i=1}^n (V_i + V_{SUI}(i)) - V_{IU} \leq S_{min} \times \Delta t, \quad i \in I_k, \quad (9)$$

де Δt – елементарний проміжок часу, а найменша швидкість передачі серед відібраних каналів є такою:

$$S_{min} = \min_{i \in I_k} S_i. \quad (10)$$

Висновки

Таким чином, запропонований метод зменшення часу доступу до сховищ даних у гіперконвергентному середовищі шляхом розгалуження інформаційних потоків. Метод є двохетапним та складається із послідовності кроків знаходження можливих маршрутів та знаходження оптимального із знайдених варіантів розподілу каналів.

Відзначимо, що при застосуванні такого підходу виникає ряд труднощів з перерозподілом потоку комірок в таких випадках:

- виродження підмережі (відмова каналів зв'язку і вузлів);
- необхідність повторної передачі, якщо комірка отримана з помилкою;
- будь-який канал почав працювати швидше або повільніше.

Напрямок подальших досліджень: розглянутий підхід представляє безперечний інтерес для подальшого вивчення, так як дозволяє істотно зменшити час передачі даних і витрати на оренду каналів, а також поліпшити живучість мережі.

СПИСОК ЛІТЕРАТУРИ

1. White Paper: Riverbed Hyper-converged Edge, available at: <https://www.riverbed.com/document-repository/white-paper--riverbed-hyper-converged-edge.html>.
2. Черняк, Л. (2012), “Время конвергентных инфраструктур”, Открытые системы. СУБД, № 4, available at: <https://www.osp.ru/os/2012/04/13015754/>.
3. Ганьжа, Д. (2016), “Гиперконвергенция: ИТ-инфраструктура на раз, два, три”, /Журнал сетевых решений., № 5, available at: www.osp.ru/lan/2016/05/13049349.
4. Кучук Г. А. Концептуальный подход до синтезу структуры информационно-телекоммуникационной сети / Г. А. Кучук, I. В. Рубан, О. П. Давікоза // Системы обработки информации. – Х.: ХУ ПС, 2013. – Вып. 7 (114). – С. 106 – 112.
5. Сергиенко И.В. Математические модели и методы решения задач дискретной оптимизации. – К.: Наук. думка, 1985. – 520 с.
6. Gelenbe E. Analysis and synthesis of computer systems (2nd Edition) / E. Gelenbe, G. Pujolle // Advances in Computer Science and Engineering : Texts – Vol.4 – 2010. – 309 p.
7. Whitt W. The Queuing Network Analyzes / W. Whitt // Bell System Tech. I. – 1983. – Vol. 62, № 9. – P. 2779 – 2815.

8. 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>
9. Свиридов А. С., Коваленко А. А., Кучук Г. А. Метод перерозподілу пропускної 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>
10. Кучук Г.А. Управління трафіком мультисервісної розподіленої телекомунікаційної мережі / Г.А. Кучук // Системи управління, навігації та зв'язку. – К.: ЦНДІ НiУ, 2007. – Вип. 2. – С. 18-27.
11. Saravana, Balaji B., Mohamed, Uvaze Ahamed, Eswaran C. and Kannan R., (2019), “Prediction-based Lossless Image Compression”, *Lecture Notes in Computational Vision and Biomechanics* (Springer), Vol. 30, No 1, pp.1749 – 17961, DOI: https://doi.org/10.1007/978-3-030-00665-5_161
12. Sivaram, M., Batri, K., Amin Salih, Mohammed and Porkodi V. (2019), “Exploiting the Local Optima in Genetic Algorithm using Tabu Search”, *Indian Journal of Science and Technology*, Volume 12, Issue 1, doi: <http://doi.org/10.17485/ijst/2019/v12i1/139577>
13. Ruban, I. Redistribution of base stations load in mobile communication networks / I. Ruban, H. Kuchuk, A. Kovalenko // *Innovative technologies and scientific solutions for industries*. – 2017. – No 1 (1) – P. 75-81.
14. Кучук Г. А. Метод параметрического управления передачей данных для модификации транспортных протоколов беспроводных сетей / Г.А. Кучук, А.С. Мохаммад, А.А. Коваленко // *Системи обробки інформації*. – 2011. – № 8(98). – С. 211-218.
15. Кучук, Г.А. Метод уменьшения времени передачи данных в беспроводной сети / Г.А. Кучук, А.С. Мохаммад, А.А. Коваленко // *Системи управління, навігації та зв'язку*. – К.: ЦНДІ НiУ, 2011. – Вип. 3 (19). – С. 209–213.
16. Кучук Г.А. Метод мінімізації середньої затримки пакетів у віртуальних з'єднаннях мережі підтримки хмарного сервісу / Г.А. Кучук, А.А. Коваленко, Н.В. ЛуковаЧуйко // *Системи управління, навігації та зв'язку*. – Полтава . ПНТУ, 2017. – Вип. 2(42). – С. 117-120.
17. Sivaram, M., Yuvraj, D., Amin Salih, Mohammed, Porkodi, V. and Manikandan V. (2018), “The Real Problem Through a Selection Making an Algorithm that Minimizes the Computational Complexity”, *International Journal of Engineering and Advanced Technology*, Vol. 8, iss. 2, 2018, pp. 95-100.
18. Svyrydov, A., Kuchuk, H., Tsiapa, O. (2018), “Improving efficiency of image recognition process: Approach and case study”, *Proceedings of 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies, DESSERT 2018*, pp. 593-597, DOI: <http://dx.doi.org/10.1109/DESSERT.2018.8409201>
19. Kosenko V. Mathematical model of optimal distribution of applied problems of safety-critical systems over the nodes of the information and telecommunication network. *Сучасні інформаційні системи (Advanced Information Systems)*. 2017. Т. 1, № 2. С. 4-9. DOI: <https://doi.org/10.20998/2522-9052.2017.2.01>

Рецензент: д-р техн. наук, доц. А. А. Коваленко,

Харківський національний університет радіоелектроніки, Харків

Received (Надійшла) 16.08.2019

Accepted for publication (Прийнята до друку) 25.09.2019

Метод разветвления запросов к хранилищам данных систем с гиперконвергентной инфраструктурой

Н. Г. Кучук, И. С. Зыков, В. И. Панченко

Актуальность исследования. В процессе функционирования компьютерной сети гиперконвергентной архитектуры за счет централизованного управления увеличивается время доступа к хранилищам данных. Но для некоторых подсистем, особенно для тех, которые должны функционировать в режиме, приближенном к режиму реального времени, данные показатели являются очень существенными. **Целью статьи** является разработка подхода к уменьшению времени доступа к хранилищам данных в гиперконвергентной среде путем разветвления информационных потоков. **Результаты исследования.** Предложенный метод является двухэтапным и состоит из последовательности шагов нахождения возможных маршрутов и нахождения оптимального из найденных вариантов распределения каналов. В результате найден лучший вариант распределения пакетов по каналам связи среди тех, что удовлетворяют сформулированным условиям. В частности, максимальное время передачи с помощью одного из каналов выбранного варианта распределения не превышает время, выделенное для передачи необходимого объема данных и выполнены ограничения по производительности и стоимости. **Выводы.** Применение подхода позволяет существенно уменьшить время передачи данных и расходы на аренду каналов, а также улучшить живучесть сети.

Ключевые слова: гиперконвергентная архитектура, хранилище данных, время выполнения запроса.

Method of branching requests to data system stores with hyperconvergent infrastructure

N. Kuchuk, I. Zikov, V. Panchenko

The relevance of research. In the process of functioning of the computer network of the hyper-converged architecture due to the centralized management, the access time to the data warehouses is increased. But for some subsystems, especially those that are supposed to function in real-time mode, these figures are very significant. **The purpose of the article** is to develop an approach to reducing access time to data warehouses in a hyper-converged environment by branching information flows. **Results of the research.** The proposed method is two-step and consists of a sequence of steps to find possible routes and find the best of the found channel distribution options. As a result, we found the best variant of packet distribution over the communication channels among those that satisfy the formulated conditions. In particular, the maximum transmission time by one of the channels of the selected distribution option does not exceed the time allocated to transmit the required amount of data and the performance and cost constraints are met. **Conclusions.** The approach can significantly reduce data transmission time and channel rental costs, as well as improve network viability.

Keywords: hyperconvergent architecture, data warehouse, query time.