УДК 681.324 **doi: 10.26906/SUNZ.2019.2.119**

Н. Г. Кучук[1], В. І. Панченко[2], А. М. Філоненко[2], І. Ю. Петровська[2]

[1] Харківський національний університет ім. В. Н. Каразіна, Харків, Україна
[2] Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

РІВНОМІРНИЙ РОЗПОДІЛ РЕСУРСІВ КОМП'ЮТЕРНИХ СИСТЕМ, ЩО МАЮТЬ ГІПЕРКОНВЕРГЕНТНУ ІНФРАСТРУКТУРУ

Актуальність дослідження. Інфраструктура, створювана на конвергентної платформі, передбачає об'єднання пам'яті, обчислювальні-них і мережевих ресурсів в єдиний пул, а при гіперконвергентной інфраструктури обчислювальні потужності, сховища, сервери, мережі об'єднуються в одне ціле за допомогою програмних засобів. **Метою статті** є розробка підходу до зменшення нерівномірних розподілу ресурсів в комп'ютерних мережах гіперконвергентной архітектури. **Результати дослідження**. Запропоновано метод мінімізації середньої затримки транзакцій в комп'ютерних мережах гіперконвергентной архітектури, що дозволяє побудувати рівномірний розподіл виділених обчислювальних ресурсів для обробки безлічі завдань системи по квантам заданого інтервалу часу. Цільова функція завдання пошуку оптимального розбиття множини запитів, що обробляються в обчислювальної мережі, на підмножини і їх розподілу по вузлах гіперконвергентной мережі, визначається за допомогою введення функції штрафів при виділенні кожного запиту одиниці обчислювального ресурсу в поточний квант часу. **Висновки**. Запропоновано підхід до формування рівномірного розподілу ресурсів в комп'ютерах терни мережах гіперконвергентной архітектури. Застосування підходу дозволяє збалансувати мережевt навантаження і досягти вимог до оперативності.

Ключові слова: гіперконвергентная архітектура; балансування мережі, метод штрафів.

Вступ

Актуальність завдання. В даний час на ринку IT-технологій розподілені хмарні платформи поступово витісняються конвергентними і гіперконвергентними [1]. Розробники систем, які передбачають централізоване управління обчислювальними ресурсами, в даний час все більше схиляються до конвергентних і гіперконвергентних рішень для базових комп'ютерних мереж. Інфраструктура, створювана на конвергентній платформі, передбачає об'єднання пам'яті, обчислювальних і мережевих ресурсів в пул, заздалегідь об’єднаних для роботи в дата-центрі [2], а при гіперконвергентній інфраструктурі обчислювальні потужності, сховища, сервери, мережі об'єднуються в одне ціле за допомогою програмних засобів. Управління відбувається через загальну консоль адміністрування [3]. Однак в більшості таких систем кінцеві абоненти використовують гетерогенне Software and Hardware.

Постановка задачі. В процесі функціонування гетерогенної комп'ютерної мережі (ГКМ) гіперконвергентної архітектури з гетерогенними компонентами виникає необхідність в забезпеченні распереділення обробки заявок віддалених абонентів в задані проміжки часу. Досягнення максимальної ефективності розподіленої обробки заявок віддалених абонентів ГКМ в задані проміжки часу, можливо, за рахунок забезпечення рівномірного розподілу обчислювальних ресурсів (ОР) ГКМ по квантам заданого інтервалу часу і знаходження розподілу заявок віддалених абонентів по вузлах ГКМ, що дозволяє мінімізувати середню затримку пакета даних в ній [4, 5]. Таким чином, завдання підвищення ефективності розподіленої обробки заявок віддалених абонентів ГКМ може бути сформульована так: необхідно побудувати рівномірний розподіл ОР ГКМ по квантах заданого інтервалу часу і знайти таке розбиття множини заявок віддалених абонентів ГКМ на підмножини і їх розподіл по вузлах ГКМ в процесі розподіленої обробки, щоб середня затримка пакета даних в мережі приймала мінімальне значення і забезпечувалася рівномірна завантаження ГКМ.

Аналіз літератури. Для вирішення поставленого завдання пропонується багато різних методів [6–23]. Однак в [4], всі перераховані методи мають порівняно невисоку обчислювальну ефективність, що обмежує рішення поставленого завдання числом змінних $h_z \times h_y$, де h_z – кількість незалежних заявок віддалених абонентів, що обробляються в мережі; h_y – кількість вузлів ГКМ. У зв'язку з цим, виникає необхідність в розробці математичної моделі, що дозволяє побудувати рівномірний розподіл обчислювальних ресурсів неоднорідною обчислювальної мережі по квантах заданого інтервалу часу і забезпечує мінімізацію середньої затримки пакета даних в мережі при розподіленої обробки заявок віддалених абонентів для числа змінних $h_z \times h_y > 300$.

Метою статті є розробка підходу до зменшення нерівномірності розподілення ресурсів в комп'ютерних мережах гіперконвергентної архітектури. Для цього пропонується розробити метод мінімізації середньої затримки транзакцій в комп'ютерних мережах гіперконвергентної архітектури, що дозволяє побудувати рівномірний розподіл виділених обчислювальних ресурсів для обробки безлічі запитів системи по квантам заданого інтервалу часу і знайти таке розбиття множини запитів на підмножини і їх розподіл по вузлах мережі в процесі розподілу обробки, щоб середня затримка пакета даних в мережі приймала мінімальне значення і забезпечувалася рівномірне завантаження мережі при великій кількості одночасно оброблюваних заявок.

Результати досліджень

Рішення завдання побудови рівномірного розподілу виділених обчислювальних ресурсів по квантах заданого часового інтервалу при обробці заявок

© Кучук Н. Г., Панченко В. І., Філоненко А. М., Петровська І. Ю., 2019

окремих абонентів обчислювальної мережі докладно розглядалося в [24]. Вихідними даними для вирішення даної задачі є: T_z – заданий інтервал часу, який представляє собою кінцевий набір квантів, рівних 1 с; Z – безліч транзакцій e-learning. При цьому часовий інтервал T_z представляється у вигляді відрізка натурального ряду $\{t_{z_1}, t_{z_2}, \dots, t_{z_i}, \dots, t_{z_{h_t}}\}$, де t_{z_i} – i-й квант часу, $1 \le i \le h_t$; h_t – число квантів тимчасового інтервалу T_z. Кожна заявка $z_b \in Z$, $1 \le b \le h_z$, характеризується параметрами ϕ_{z_b}, $M_{\phi}^{(\gamma)}$, де ϕ_{z_b} – необхідний ВР для обробки заявки z_b; $T_{z_b} = \{t_{z_{b_1}}, t_{z_{b_2}}\}$ – інтервал часу, протягом якого необхідно надати необхідний обчислювальний ресурс; $t_{z_{b_1}}$ – початковий квант часового інтервалу T_{z_b}; $t_{z_{b_2}}$ – кінцевий квант часового інтервалу T_{z_b}.

В результаті розподілу γ обчислювальних ресурсів ГКМ формується матриця $M_{\phi}^{(\gamma)}$, в якій кожній заявці $z_b \in Z$ зіставляється вектор-рядок $m_{\phi_b} = \left(m_{\phi_{b,1}}, \dots, m_{\phi_{b,h_t}}\right)$, що представляє собою розклад виділення обчислювальних ресурсів ГКМ для обробки заявки z_b, де компонент $m_{\phi_{b,i}}$ визначає виділений для заявки z_b обчислювальний ресурс в i-й квант часу.

Якість розподілу γ оцінюється за допомогою цільової функції $F^{(\gamma)}$ і величини максимального сумарного виділеного ВР, що припадає на квант заданого тимчасового інтервалу T_z в розподілі γ по всіх заявках безлічі Z [7]: $m_{\phi\max}^{(\gamma)} = \max\limits_{i=1,\dots,h_t} \sum\limits_{b=1}^{h_z} m_{\phi_{b,i}}$. За умови рівномірного розподілу по квантам заданого тимчасового інтервалу T_z сумарного обсягу обчислень, необхідного для обробки заявок безлічі Z, вираз для визначення величини мінімального сумарного необхідного ВР, що припадає на квант інтервалу, набуде вигляду [24]: $\phi_{z\min} = \frac{1}{h_t} \sum\limits_{b=1}^{h_z} \phi_{z_b}$. Основою визначення цільової функції $F^{(\gamma)}$ служить штраф при виділенні заявці $z_b \in Z$ одиниці обчислювального ресурсу в i-й квант часу. Якщо одиниця ВР для заявки z_b, яка характеризується інтервалом часу обробки $T_{z_b} = \{t_{z_{b_1}}, t_{z_{b_2}}\}$, виділена в i-й квант, то відповідний їй штраф визначається як [24]:

$$s_{t_{b,i}} = \begin{cases} 0, & \textit{если} \quad t_{z_{b_1}} \le t_{z_i} \le t_{z_{b_2}}; \\ \left(t_{z_{b_1}} - t_{z_i}\right) \Big/ \phi_{z_b}, & \textit{если} \quad t_{z_i} < t_{z_{b_1}}; \\ \left(t_{z_i} - t_{z_{b_2}}\right) \Big/ \phi_{z_b}, & \textit{если} \quad t_{z_i} > t_{z_{b_2}}, \end{cases}$$

Таким чином, для кожної заявки $z_b \in Z$ маємо вектор $s_{t_b} = \left(s_{t_{b_1}}, \dots, s_{t_{b_{h_t}}}\right)$, у якого компонент $s_{t_{b_i}}$, $1 \le i \le h_t$, визначає величину штрафу при виділені заявці z_b одиниці ВР в i-й квант часу.

Величина штрафу, що характеризує отриманий розподіл γ виділених ВР для обробки безлічі заявок Z, визначає цільову функцію [24]:

$$F^{(\gamma)} = \sum_{b=1}^{h_z} \sum_{i=1}^{h_t} m_{\phi_{b,i}} \cdot s_{t_{b,i}}.$$

При побудові розподілу γ ВР по квантам на заданому інтервалі часу T_z мінімізуються величини $F^{(\gamma)}$ і $m_{\phi\max}^{(\gamma)}$. При цьому для розподілу γ потрібно:

$$\forall z_b \in Z, \ \forall t_{z_i} \in T_{z_b}, \ m_{\phi_{b,i}} \ge 0, \ s_{t_{b,i}} \ge 0;$$

$$\forall z_b \in Z \quad \sum_{i=1}^{h_t} m_{\phi_{b,i}} \le \phi_{z_b}; \quad \forall t_{z_i} \in T_z \quad \sum_{b=1}^{h_z} m_{\phi_{b,i}} \le \phi_{t_i},$$

де ϕ_{t_i} – сумарний доступний обчислювальний ресурс ГКМ в i-й квант завданого проміжку часу T_z.

Отриманий розподіл γ описується за допомогою кортежу $\langle Z, \phi_z, T_z, \phi_t, M_{\phi}^{(\gamma)}, F^{(\gamma)}, m_{\phi\max}^{(\gamma)} \rangle$, де Z – множина заявок віддалених абонентів; $\phi_z : Z \to N_+$ – функція, яка вказує кожній заявці $z_b \in Z$ необхідний обчислювальний ресурс для її обробки; $T_z : Z \to N_+$ – функція, яка вказує кожній заявці $z_b \in Z$ інтервал часу для її обробки; $\phi_t : T_z \to N_+$ – функція, яка вказує кожному кванту часу $t_{z_i} \in T_z$ сумарний доступний обчислювальний ресурс ГКМ

Для вектор-стовпця $m_{\phi_i} = \left(m_{\phi_{1,i}}, \dots, m_{\phi_{h_z,i}}\right)$ матриці $M_{\phi}^{(\gamma)}$, що визначає виділений ВР ГКМ для обробки заявок множини Z в i-й квант заданого інтервалу часу T_Z необхідно знайти таке розбиття множини заявок Z на підмножини і їх розподіл по вузлах ГКМ, щоб середня затримка пакета даних в мережі приймала мінімальне значення. Цільова функція задачі пошуку оптимального розбиття множини завдань Z, оброблюваних в мережі, на підмножини і їх розподілу по вузлах $Y_A \in Y$, визначається так [4]:

$$F^{(\gamma)} = \frac{1}{u_{z\max}} \cdot \sum_{b=1}^{h_z} \sum_{a=1}^{h_y} m_{z_{b,a}} \cdot s_{y_{b,a}}, \qquad (1)$$

де Y – множина вузлів ГКМ; $u_{z\max}$ – незалежна від розподілу γ величина, яка визначає максимальну сумарну інтенсивність обміну заявок з вузлами обчислювальної мережі відповідно до вираження $u_{z\max} = \sum\limits_{b=1}^{h_z} \sum\limits_{i=1}^{h_y} u_{z_{b,i}}$; $u_{z_{b,i}}$ – інтенсивність обміну заявки $z_b \in Z$ з вузлом $y_i \in Y$; $m_{z_{b,a}}$ – ОР вузла y_a, необхідний для обробки заявки z_b; $s_{y_{b,a}}$ – штраф при розподілі заявки $z_b \in Z$ на узел $y_a \in Y$,, визначений вира-

зом $s_{y_{b,a}} = \sum_{i=1}^{h_y} \left(u_{z_{b,i}} \cdot h_{wa,i} \right) / \phi_{z_b}$; $h_{wa,i}$ – довжина найкоротшого маршруту між вузлами y_a і y_i, що визначається числом каналів ПД, що входять в цей маршрут. Отриманий розподіл γ відповідати вимогам:

$$\forall y_a \in Y \left| \sum_{b=1}^{h_z} m_{z_{b,a}} \le \phi_{y_a} ; \ \forall z_b \in Z \right| \sum_{a=1}^{h_y} m_{z_{b,a}} \le \phi_{z_b} ;$$

$$\sum_{a=1}^{h_y} \phi_{y_a} \ge \sum_{b=1}^{h_z} \phi_{z_b} ; \ s_{y_{b,a}} \ge 0, \ s_{z_{b,a}} \ge 0 ,$$

при $1 \le a \le h_y$, $1 \le b \le h_z$, де ϕ_{y_a} – доступний обчислювальний ресурс вузла $y_a \in Y$.

З урахуванням наведених умов, завдання пошуку оптимального розбиття множини заявок Z, оброблюваних в обчислювальної мережі, на підмножини і їх розподілу по вузлах $y_a \in Y$ може бути сформульована таким чином. Нехай задані безлічі заявок Z і вузлів Y обчислювальної мережі, які визначаються кортежами $\langle Z, \phi_z, U_z \rangle$ та $\langle Y, \phi_y, H_w \rangle$, де $\phi_z = \left(\phi_{z_1}, \ldots, \phi_{z_{h_z}} \right)$ – вектор необхідних ОР для обробки множини заяв Z; $U_z = \left\| u_{z_{b,i}} \right\|$ – матриця інтенсивності обміну заявок множини Z з вузлами множини Y; $\phi_y = \left(\phi_{y_1}, \ldots, \phi_{y_{h_y}} \right)$ — вектор доступних ВР множин вузлів Y обчислювальної мережі; $H_w = \left\| h_{w_{a,i}} \right\|$ – матриця довжин найкоротших маршрутів між кожною парою вузлів ГКМ y_a и y_i, $1 \le a \le h_y$ і $1 \le i \le h_y$. Потрібно знайти такий розподіл γ, який задовольняє умови 1 – 4, щоб вираз (1) приймало мінімальне значення.

Для вирішення зазначеного завдання доцільно використовувати метод потенціалів [4], що забезпечує послідовне виконання наступних операцій: побудова базового розподілу ОР мережі; побудова системи потенціалів; перевірку базового розподілу на раціональність; побудова замкнутого контуру і перерозподіл ОР по контуру з метою мінімізації цільової функції $F^{(\gamma)}$. В результаті розбиття множини заявок Z, оброблюваних в ГКМ, на підмножини і їх розподілу по вузлах безлічі Y, що забезпечують мінімізацію середньої затримки пакета даних в мережі, формується матриця $M_z^{(\gamma)}$, кожен елемент $m_{z_{b,a}}$ якої визначає ОР вузла $y_a \in Y$, виділений для обробки заявки $z_b \in Z$. Результуючий розподіл описується як $\langle Z, \phi_z, Y, \phi_y, M_z^{(\gamma)}, F^{(\gamma)} \rangle$.

Висновки

Таким чином, запропонований підхід до формування рівномірного розподіл ресурсів в комп'ютерних мережах гіперконвергентної архітектури. Підхід ґрунтується на запропонованому методі мінімізації середньої затримки, що враховує особливості гіперконвергентної архітектури. Застосування підходу дозволяє збалансувати мережеву навантаження при великому числі запитів і досягти вимог до оперативності їх обробки. **Напрямок подальших досліджень** – розробити метод, який зв'язує інформаційну структуру e-learning зі структурою базової гі-перконвергентної 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 (accessed 23 April 2017).
2. Черняк, Л. (2012), “Время конвергентных инфраструктур”, Открытые системы. СУБД, № 4, available at: https://www.osp.ru/os/2012/04/13015754/ (accessed 23 April 2017).
3. Ганьжа, Д. (2016), “Гиперконвергенция: ИТ-инфраструктура на раз, два, три”, /Журнал сетевых решений,. № 5, available at: www.osp.ru/lan/2016/05/13049349 (accessed 23 April 2017).
4. Кучук Г. А. Інформаційні технології управління інтегральними потоками даних в інформаційно-телекомунікаційних мережах систем критичного призначення / Г. А. Кучук. – Х.: ХУПС, 2013. – 264 с.
5. Кучук Г. А. Концептуальний підхід до синтезу структури інформаційно-телекомунікаційної мережі / Г. А. Кучук, І. В. Рубан, О. П. Давікоза // Системи обробки інформації. – Х.: ХУ ПС, 2013. – Вип. 7 (114). – С. 106 – 112.
6. Сергиенко И.В. Математические модели и методы решения задач дискретной оптимизации. – К.: Наук. думка, 1985. – 520 с.
7. 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.
8. Whitt W. The Queuing Network Analyzes / W. Whitt // Bell System Tech. I. – 1983. – Vol. 62, № 9. – P. 2779 – 2815.
9. Кучук Г.А. Метод оценки характеристик АТМ-трафика / Г.А. Кучук // Інформаційно-керуючі системи на залізничному транспорті, – 2003. – № 6. – С. 44–48.
10. Кучук Г. А., Можаєв О. О., Воробйов О. В. Метод агрегування фрактального трафіка. Радіоелектронні та комп'ютерні системи. 2006. № 6 (18). С. 181–188.
11. Saravana, Balaji B,, Karthikeyan, N.K. and Raj Kumar, R.S., (2018), “Fuzzy service conceptual ontology system for cloud service recommendation”, *Computers & Electrical Engineering*, Vol. 69, pp. 435–446.
12. 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
13. Кучук Г. А. Фрактальный гауссовский шум в трафиковых трассах / Г.А. Кучук // Системи обробки інформації. – 2004. – № 3(31). – С. 91-100.
14. Кучук Г.А. Аналіз та моделі самоподібного трафіка / Г.А. Кучук, О.О. О.В. Можаев, Воробйов // Авиационно-космическая техника и технология. – 2006. – Вып. 9 (35). – С. 173-180.

15. Gomathi, B, Karthikeyan, N.K. and Saravana, Balaji B., (2018), "Epsilon-Fuzzy Dominance Sort Based Composite Discrete Artificial Bee Colony optimization for Multi-Objective Cloud Task Scheduling Problem", *International Journal of Business Intelligence and Data Mining*, Vol. 13, Issue 1-3, pp. 247-266, **DOI:** https://doi.org/10.1504/IJBIDM.2018.088435
16. Dhivakar B., Saravanan S.V., Sivaram M., Krishnan R.A. Statistical Score Calculation of Information Retrieval Systems using Data Fusion Technique". *Computer Science and Engineering*. 2012. Vol. 2, Issue 5. pp.43-45. doi: http://doi.org/10.5923/j.computer.20120205.01
17. 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
18. 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.
19. Коваленко А.А. Сучасний стан та тенденції розвитку комп'ютерних систем об'єктів критичного застосування / А.А. Коваленко, Г.А. Кучук // Системи управління, навігації та зв'язку. – Полтава . ПНТУ, 2018. – Вип. 1(47). – С. 110-113. DOI : https://doi.org/10.26906/SUNZ.2018.1.110
20. Кучук Г. А. Модель процесса эволюции топологической структуры компьютерной сети системы управления объектом критического применения / Г.А. Кучук, А.А. Коваленко, А.А. Янковский // Системи обробки інформації. – 2014. – № 7(123). – С. 93-96.
21. 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.
22. 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.
23. Sivaram, M., Yuvaraj, 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.
24. 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, проф. І. В. Рубан,
Харківський національний університет радіоелектроніки, м. Харків
Received (Надійшла) 16.01.2019
Accepted for publication (Прийнята до друку) 27.02.2019

Равномерное распределение ресурсов компьютерных систем с гиперконвергентной инфраструктурой

Н. Г. Кучук, В. И. Панченко, А. М. Филоненко, И. Ю. Петровская

Актуальность исследования. Инфраструктура, создаваемая на конвергентной платформе, предполагает объединение памяти, вычислительных и сетевых ресурсов в единый пул, а при гиперконвергентной инфраструктуре вычислительные мощности, хранилища, серверы, сети объединяются в одно целое с помощью программных средств. **Целью статьи** является разработка подхода к уменьшению неравномерности распределения ресурсов в компьютерных сетях гиперконвергентной архитектуры. **Результаты исследования**. Предложен метод минимизации средней задержки транзакций в компьютерных сетях гиперконвергентной архитектуры, позволяющий построить равномерное распределение выделенных вычислительных ресурсов для обработки множества заданий системы по квантам заданного интервала времени. Целевая функция задачи поиска рационального разбиения множества запросов, обрабатываемых в вычислительной сети, на подмножества и их распределения по узлам гиперконвергентной сети, определяется при помощи ввода функции штрафов при выделении каждому запросу единицы вычислительного ресурса в текущий квант времени. **Выводы.** Предложен подход к формированию равномерного распределения ресурсов в компьютерных сетях гиперконвергентной архитектуры. Применение подхода позволяет сбалансировать сетевую нагрузку при большом числе запросов и достичь требований к оперативности их обработки.

Ключевые слова: гиперконвергентная архитектура; балансировка сети, метод штрафов.

Uniform resource allocation of computer systems with hyperconvergent infrastructure

N. Kuchuk, V. Panchenko, A. Filonenko, I. Petrovskaya

The relevance of research. The infrastructure created on a converged platform involves the integration of memory, computing and network resources into a single pool, and with a hyper-convergent infrastructure, computing power, storage, servers, networks are combined into one whole using software tools. **The purpose of the article** is to develop an approach to reducing the uneven distribution of resources in computer networks of hyperconvergent architecture. **The results of the study**. A method is proposed for minimizing the average transaction latency in computer networks of a hyperconvergent architecture, which makes it possible to construct a uniform distribution of the allocated computational resources for processing a set of system tasks using the quanta of a given time interval. The objective function of the task of finding a rational partition of the set of queries processed in the computer network into subsets and their distribution among the nodes of the hyperconvergent network is determined by entering the penalty function when each query unit is allocated a computing resource in the current time slot. **Conclusions.** An approach to the formation of a uniform distribution of resources in computer networks of hyperconvergent architecture is proposed. The application of the approach allows you to balance the network load with a large number of requests and to achieve the requirements for the efficiency of their processing.

Keywords: hyperconvergent architecture; network balancing, penalty method.