

УДК 396.42:004.7 (045)

[0000-0002-3268-342X] **П. С. Усік**¹, аспірант,

e-mail: mr.usik@ukr.net

[00000-0001-9543-874X] **О. А. Смірнов**¹, д.т.н., професор,

e-mail: dr.SmirnovOA@gmail.com

[0000-0003-2007-9943] **І. В. Миронець**², к.т.н., доцент,

e-mail: i.myronets@chdtu.edu.ua

[0000-0001-6195-7533] **К. О. Буравченко**¹, к.т.н.,

e-mail: buravchenkok@gmail.com

[0000-0002-4498-0093] **Н. М. Якименко**¹, к.ф.-м.н., доцент,

доцент кафедри кібербезпеки та програмного забезпечення

e-mail: yakimenko_n_m@ukr.net

¹Центральноукраїнський національний технічний університет
просп. Університетський, 8, м. Кропивницький, 25006, Україна²Черкаський державний технологічний університет
б-р Шевченка, 460, м. Черкаси, 18006, Україна

МЕТОД ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ РОЗПОДІЛЕНОЇ ОБРОБКИ ДАНИХ У КОМП'ЮТЕРНИХ СИСТЕМАХ ОПЕРАТОРІВ СТІЛЬНИКОВОГО ЗВ'ЯЗКУ

За аналізом якості обслуговування абонентів у реалізованих проектах мереж 5G в світі встановлено, що заявлені вимоги до мереж п'ятого покоління не досягаються в жодній із реалізованих мереж, а фактична якість обслуговування абонентів перебуває на не досить високому рівні, що свідчить про низьку ефективність наявних методів керування мережами стільникового оператора та розподіленої обробки даних у них. Тому було розроблено метод оптимізації розміщення масштабованих послуг на розподілених обчислювальних ресурсах мережі стільникового оператора. Також було проведено імітаційне моделювання для оцінювання ефективності розробленого у роботі методу, результати якого підтвердили його ефективність.

Ключові слова: стільниковий зв'язок, передача даних, розподілені обчислення, мережеві канали, генетичний алгоритм.

Вступ. Дві основні тенденції спонукають індустрію безпроводного зв'язку розвивати мережі стільникового зв'язку п'ятого покоління: стрімке збільшення попиту на безпроводні широкосмугові послуги, які потребують значно більших швидкостей передачі даних, і мережі значно більшої ємності, які можуть надавати відео- й інші ресурсоємні послуги та послуги Інтернету речей (IoT), що спонукають до необхідності масового підключення пристроїв, а також до необхідності наднадійного зв'язку з наднизькою затримкою [1, 2].

Можна визначити низку різних сфер застосування, де будуть використовуватися ці мережі: сюди входять V2X комунікація (комунікація транспортних засобів між собою та з іншими об'єктами інфраструктури); промислова автоматизація та комунальні програми; безпроводні медичні послуги; споживчі та умови роботи локальної мережі розподілених мобільних пристроїв [3-6].

З метою вирішення цих проблем, пов'язаних з хмарами, нещодавні дослідження

представили аналогічну концепцію, що розширює можливості хмарних розрахунків, які ближчі до кінцевих користувачів (тобто на межі мережі) – Mobile Edge Computing or Multi-access Edge Computing (MEC) [7]. MEC надає новий рівень розподілених обчислювальних вузлів між пристроями кінцевих користувачів і хмарними центрами обробки даних. Тому додатки, що працюють на MEC, можуть виконувати дії, які близькі до їх користувачів, перед підключенням до хмари. Це значно знижує навантаження на мережу, забезпечує більш оперативну відповідь і дає змогу отримувати локальну контекстну інформацію найбільш ефективним способом [8-13].

Аналіз останніх досліджень та публікацій. Останнім часом з'явилося багато наукових праць вітчизняних та, здебільшого, закордонних дослідників (Климаш М. М., Лемешко О. В., Одарченко Р. С., Xiang Sun, Nirwan Ansari, Robert Schober, Albert Banches, Tuyen X. Tran, Luca Foschini та багато інших) [7, 14-15], присвячених дослідженням

оцінювання та підвищення ефективності використання МЕС для різних застосувань з метою підвищення надійності та динамічності використання їх у мережах 5G.

Проте ця технологія розподілених обчислень зовсім не позбавлена недоліків, які необхідно нівелювати для її повсюдного розгортання.

Мета дослідження полягає у підвищенні ефективності розподіленої обробки

даних в комп'ютерних системах операторів стільникового зв'язку. Досягнення мети спрямовано на вдосконалення якості обслуговування абонентів сучасних стільникових мереж і забезпечення вимог до мереж наступних поколінь.

Виклад основного матеріалу. Узагальнений випадок використання мобільної мережі з граничними обчислювальними можливостями (мережа 5G) зображено на рисунку 1 [14].

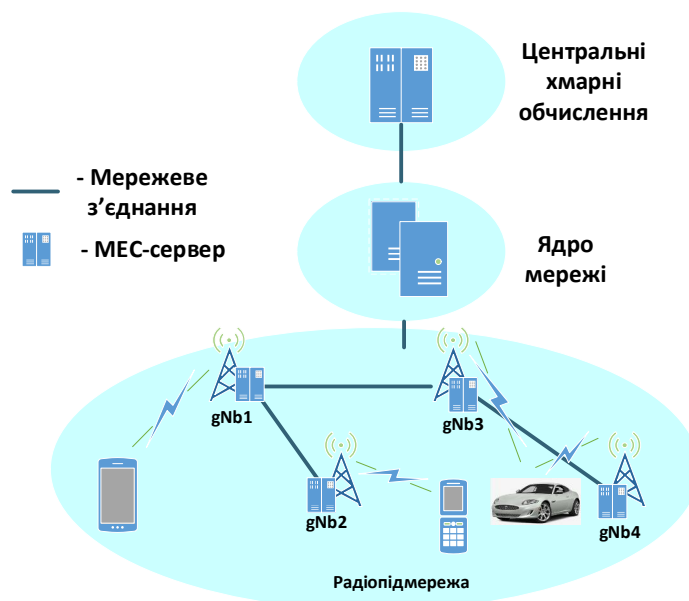


Рисунок 1 – Архітектура мережі 5G із використанням МЕС

У запропонованій схемі обчислювальні вузли можуть розміщуватися в різних частинах мережі, а не тільки на базових станціях мобільної мережі. В цьому випадку мережа ЕС складається з різних вузлів – граничних та хмарних вузлів, пристроїв кінцевого користувача і ліній зв'язку, які з'єднують ці вузли між собою.

Розглянута вище мережа може бути представлена у вигляді однонаправленого графа $G = (v, \varepsilon)$, де вершини v – вузли мережі, а ε – це мережеві канали між вузлами (рисунком 2).

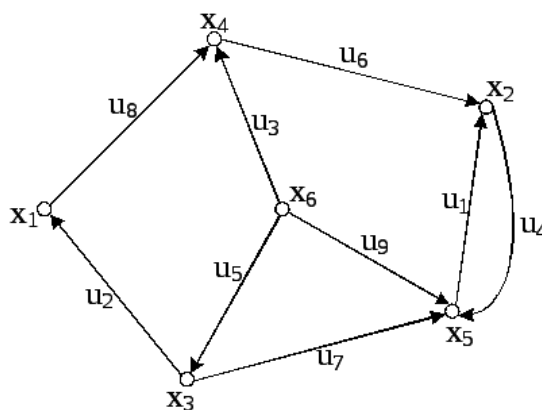


Рисунок 2 – Неорієнтований граф мережі з граничними обчисленнями

Припустимо, що кожна вершина доступна будь-якій іншій вершині в графі через декілька послідовних передач (хопів). Окрім того, пристрої кінцевих користувачів і їх з'єднання не представлені в G . Водночас кожне мережеве з'єднання $e \in \mathcal{E}$ має наступні властивості:

- *затримка передачі* $D_{net}^{a,e}$ – це час, потрібний для виконання запиту, що надійшов від додатка a , який буде переданий в канал e . Запропонована модель визначає різні типи ресурсів R . Наприклад, набір $R = \{CPU, RAM, Storage\}$ складається з обчислювальної потужності процесора, оперативної пам'яті (RAM) і ресурсів дискового простору.

Кожний вузол $v \in V$ має наступні характеристики:

- *ємність ресурсу* $C_{v,r}$ – число, що описує загальну ємність ресурсу $r \in R$ на вузлі v ; водночас приймаємо, що хмара має потенційно необмежені ресурси, $C_{cloud,r} = \infty$ через різницю в ємності між хмарою і граничним вузлом.

Кожний додаток $a \in A$ має наступні параметри:

- максимальна допустима затримка D_a ,
- максимальна кількість екземплярів додатка $N_a \in \mathbb{C}$,
- попит на ресурс $f_a^r(\lambda)$,
- необхідна обчислювальна потужність W_a ,
- частота запитів λ_a .

У практичному сценарії неможливо розмістити всі додатки на граничних вузлах з урахуванням обмеженості ресурсів вузлів МЕС. З цього випливає, що деякі додатки розгортаються додатково (тобто в базовій мережі або хмарі).

Ця значна відстань між вузлом і користувачем може призвести до збільшення затримки відгуку на запит. Більше того, перевантажений сервер також збільшує час відгуку, а розподілення навантаження між копіями додатків можуть зменшити цю проблему. Тому дослідимо сумісну проблему розміщення і завантаження сервісів розподілення для мінімізації можливих порушень QoS (максимальна допустима затримка).

Водночас мережева затримка обробки оцінюється наступним чином:

$$\bar{d}(F_a^{u,v}) = \bar{d}_{net}(F_a^{u,v}) + \bar{d}_{proc}(a, v). \quad (1)$$

Мережева затримка відгуку на запити включає:

- затримку між кінцевим пристроєм, що робить запит, та вузлом, до якого він приєднаний;

- затримку передачі з цього останнього вузла на сервер, де розміщено додаток, який прямує транзитним маршрутом.

Обробка запиту моделюється з використанням моделі масового обслуговування $M/M/1$ [16]. У цій моделі користувачів безперервно створюються запити щодо застосування відповідного пуассонівського процесу з відношенням A .

Варто відзначити, що вирішення проблеми обмеження автономного розміщення служби можливе тільки при виконанні всіх наступних вимог:

1) *кількість екземплярів*: вузол може розміщувати тільки одну репліку даного додатка. Окрім того, кількість екземплярів, які розгорнуті в системі, повинна відповідати обмеженням, які визначені додатками, і всі вони мають бути розміщені

$$1 \leq \sum_{v \in V} p_a^v \leq N_a \quad \forall a \in A; \quad (2)$$

2) *існування потоку запитів*: потік запитів існує тільки в тому випадку, якщо копія додатка a розміщена в v , і є користувачі, які підключені в u із запитом a . Нехай $\gamma_a^{u,v} \in \{0,1\}$ – двійкова змінна, яка виражає існування потоку $F_a^{u,v}$,

$$\gamma_a^{u,v} \leq p_a^v Q_a^u \quad \forall a \in A, \forall u, v \in V; \quad (3)$$

3) *розмір потоку запиту*: якщо потік існує, його розмір повинен бути, принаймні, одним і дорівнювати кількості запитів

$$\gamma_a^{u,v} \leq \delta_a^{u,v} \leq \gamma_a^{u,v} Q_a^u \quad \forall a \in A, \forall u, v \in V; \quad (4)$$

4) *збереження навантаження*: сукупний розмір усіх потоків запитів для застосування a від того ж вихідного вузла u дорівнює загальній кількості запитів від користувачів, пов'язаних цим вузлом,

$$\sum_{v \in V} \delta_a^{u,v} = Q_a^u \quad \forall_a \in A, \forall_u \in V; \quad (5)$$

5) *продуктивність вузла*: загальна сума ресурсів, що запитуються додатками, не повинна перевищувати продуктивності вузлів

$$\sum_{a \in A} p_a^v f_a^r(\lambda_a^v) \leq C_{v,r} \quad \forall_r \in R, \forall_v \in V; \quad (6)$$

6) *стабільність черги*: черга в моделі $M/M/1$ стабільна, тільки якщо середній темп обслуговування більший, ніж середній темп надходження запитів. Ця стабільність повинна бути гарантована для кожного додатка, поміщеного у вузол:

$$\lambda_a^v < \mu_a^v \quad \forall_{a,v} (p_a^v = 1), a \in A, v \in V; \quad (7)$$

7) *порушення QoS*: час відклику наявного потоку не повинен перевищувати крайній термін свого застосування плюс рівень порушення QoS системи:

$$\gamma_a^{u,v} \bar{d}(F_a^{u,v}) \leq D_a + \varepsilon \quad \forall_a \in A, \forall_{u,v} \in V. \quad (8)$$

Ми визначаємо рівень порушення QoS потоку запиту як відмінність між його середнім часом відгуку і крайнім терміном виконання.

Рівень порушення QoS системи – найвищий рівень порушення серед усіх потоків у системі $(\bar{d}(F_a^{u,v}) - D_a)$. Тому основна мета дослідження полягає в тому, щоб мінімізувати рівень порушення QoS. Отже, поставлена сервісна проблема розміщення може бути сформульована наступним чином:

$$\min(\varepsilon). \quad (9)$$

Результати дослідження. Для вирішення окресленої вище задачі в цій роботі

пропонується евристичне рішення, що ґрунтується на генетичних алгоритмах [17]. Запропонований генетичний алгоритм використовує упереджені хромосоми з випадковими ключами, що є масивом випадково генерованих реальних чисел в інтервалі $[0,1]$.

Таке представлення хромосом використовується не для створення нездійснених рішень, які можуть погіршити продуктивність генетичного алгоритму. Ця пропозиція використовує елітарну стратегію, зберігаючи елітарних індивідів, тобто тих, хто має найкращі цінності, наступному поколінню. Він також додає нові випадкові покоління індивідів, амутантів у наступному поколінні. Крім того, для поповнення популяції породження створюється параметризований рівномірний кросовер між елітарними та неелітарними особами.

В алгоритмах з випадковим ключем випадковий алгоритм детермінованого декодера приймає хромосому індивіда та обчислює її значення. Таким чином, представлення хромосоми та алгоритму декодера відіграє важливу роль у розробленій моделі. Ця пропозиція щодо кодування хромосоми та опис її частин наведені нижче:

$$C = [O_1^1, O_1^2, \dots, O_1^{|V|}, \dots, O_{|A|}^1, O_{|A|}^2, \dots, O_{|A|}^{|V|}, \\ M_1, M_2, \dots, M_{|A|}, \\ V_1^1, V_1^2, \dots, V_1^{|V|}, \dots, V_{|A|}^1, V_{|A|}^2, \dots, V_{|A|}^{|V|}] \quad (10)$$

1) $O_1^1, \dots, O_{|A|}^{|V|}$ – це порядок створення запитів потоку;

2) $M_1, \dots, M_{|A|}$ – описує вагу, яка використовується при виборі сервера для розміщення програми;

3) $V_1^1, \dots, V_{|A|}^{|V|}$ – параметри для обчислення пріоритету вузла, який буде обраний як місце для розгортання програми.

Цей пріоритет вузла v для програми a задається як:

$$M_a^v V_a^v + (1 - M_a^v) \frac{D_{net}^{a,u,cloud} - D_{net}^{a,u,v}}{D_{net}^{a,u,cloud}}. \quad (11)$$

Алгоритм роботи декодера зображено на рисунку 3.

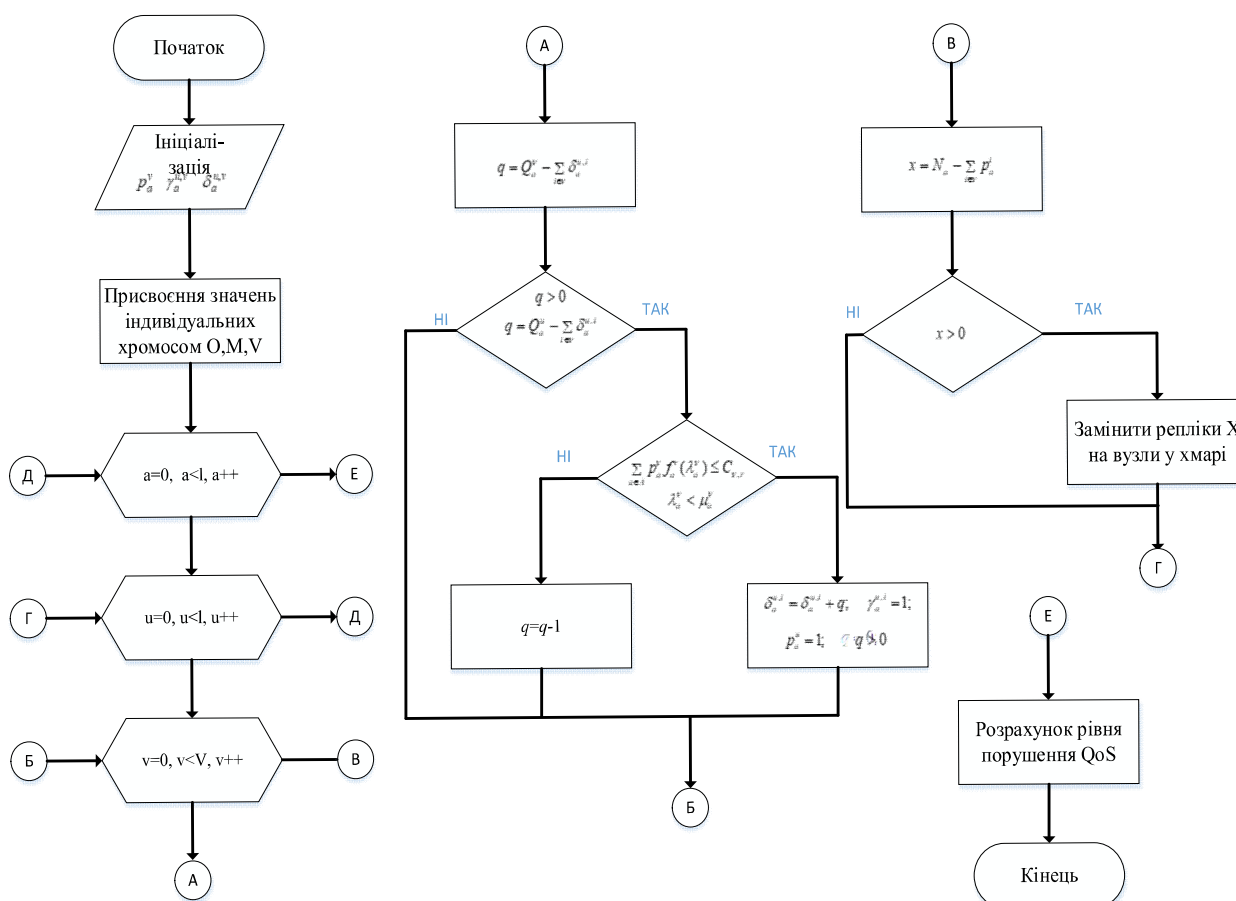


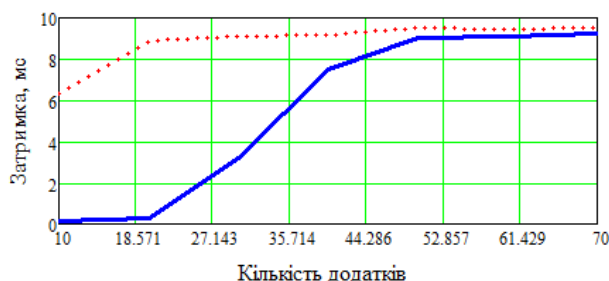
Рисунок 3 – Алгоритм роботи декодера

Просте рішення проблеми розміщення сервісу розгортає додаток на найближчих серверах до користувачів цього додатка, тобто серверів з меншою затримкою мережі для цих користувачів. Однак обмеження ємності вузлів заважає цій схемі розгортання оптимально працювати для великої кількості програм або користувачів.

Щоб покращити це рішення, пропонується включити ще один параметр у допов-

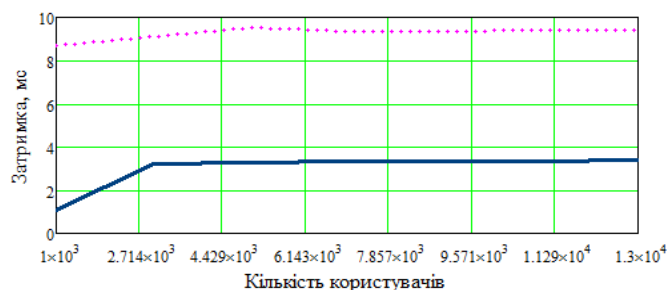
нення до затримки мережі в процедурі вибору вузла. Крім того, ваговий коефіцієнт (M) врівноважує ці два параметри в процесі прийняття рішення.

Оцінку розробленого методу та його порівняння із реалізаціями інших методів оптимізації, застосованих до вирішення задачі розміщення масштабованих послуг, зображено на рисунку 4.



(а) Кількість користувачів – 4000

для мережі із 10 базових станцій gNb



(б) Кількість додатків – 30

Рисунок 4 – Результати моделювання досліджуваних алгоритмів

Аналіз результатів моделювання, продемонстрований на рисунку 4, дає змогу впевнитись в ефективності розробленого методу.

Обговорення результатів. За аналізом якості обслуговування абонентів у реалізованих проектах мереж 5G в світі встановлено, що заявлені вимоги до мереж п'ятого покоління не досягаються в жодній із реалізованих мереж, а фактична якість обслуговування абонентів перебуває на не досить високому рівні, що свідчить про низьку ефективність наявних методів керування мережами стільникового оператора та розподіленої обробки даних у них. Тому було проаналізовано ефективність методів, моделей і технологій розподіленої обробки даних у комп'ютерних системах операторів стільникового зв'язку та, відповідно, встановлено їх недоліки. Після цього було розроблено метод оптимізації розміщення масштабованих послуг на розподілених обчислювальних ресурсах мережі стільникового оператора, що полягає у послідовному використанні моделі граничних обчислень, узагальненої моделі мережі стільникового оператора та евристичного рішення, що ґрунтується на використанні генетичних алгоритмів.

Висновки. Розроблений метод оптимізації розміщення масштабованих послуг на розподілених обчислювальних ресурсах мережі стільникового оператора дає можливість зменшити рівень деградації якості обслуговування кінцевих абонентів мережі стільникового оператора, зокрема, затримки на величину до 8 мс для великої кількості абонентів та, відповідно, великої кількості завдань.

Отримані наукові результати доцільно використовувати під час планування безпроводних мереж стандарту LTE та 5G, розробки обладнання та програмного забезпечення для систем передачі даних, а також у навчальному процесі.

Список використаних джерел

- [1] Cisco Annual Internet Report. [Online]. Available: <https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=2055169>
- [2] Ericsson Mobility Report. [Online]. Available: <https://www.ericsson.com/en/mobility-report>
- [3] М. А. Маколкина, А. А. Атея, А. С. А. Мутханна, и А. Е. Кучерявый, "Метод выгрузки трафика приложений дополненной реальности в многоуровневой системе граничных вычислений", *Электросвязь*, № 6, с. 36-42, 2019.
- [4] Д. В. Кашкаров, и А. Е. Кучерявый, "Анализ приложений и перспектив развития технологий граничных вычислений с множественным доступом в сетях связи", *Информационные технологии и телекоммуникации*, № 8.1, с. 28-33, 2020.
- [5] А. А. Атея, А. И. Выборнова, и А. Е. Кучерявый, "Многоуровневая облачная архитектура для услуг Тактильного Интернета", *Электросвязь*, № 2, с. 26-30, 2017.
- [6] А. С. Викторов, и В. Н. Шведенко, "Методика создания эффективного распределенного хранилища данных для периферийного сервиса обработки данных телеметрии БПЛА", *Информационно-экономические аспекты стандартизации и технического регулирования*, № 1, 2018.
- [7] N. Hassan, K. A. Yau, and C. Wu, "Edge computing in 5G: A review", in *IEEE Access*, vol. 7, pp. 127276-127289, 2019. DOI: 10.1109/ACCESS.2019.2938534
- [8] Shi, Weisong, et al., "Edge computing: Vision and challenges", *IEEE Internet of Things Journal*, no. 3.5, pp. 637-646, 2016.
- [9] Premsankar, Gopika, Mario Di Francesco, and Tarik Taleb, "Edge computing for the Internet of Things: A case study", *IEEE Internet of Things Journal*, no. 5.2, pp. 1275-1284, 2018.
- [10] Li, Hongxing, et al. "Mobile edge computing: Progress and challenges", in *4th IEEE Int. Conf. on mobile cloud computing, services, and engineering (MobileCloud)*, IEEE, 2016.
- [11] Wang, Shangguang, et al., "Edge server placement in mobile edge computing", *Journal of Parallel and Distributed Computing*, no. 127, pp. 160-168, 2019.
- [12] Wang, Shangguang, et al., "A survey on service migration in mobile edge computing", *IEEE Access*, no. 6 pp. 23511-23528, 2018.
- [13] Liu, Juan, et al. "Delay-optimal computation task scheduling for mobile-edge computing systems", in *IEEE Int. Symposium on Information Theory (ISIT)*, IEEE, 2016.
- [14] J. Hammer, P. Moll, and H. Hellwagner, "Transparent access to 5G edge computing services", in *IEEE Int. Parallel and Distributed Processing Symposium*

- Workshops (IPDPSW)*, Rio de Janeiro, Brazil, 2019, pp. 895-898. DOI: 10.1109/IPDPSW.2019.00147
- [15] Q. Pham et al., "A survey of multi-access edge computing in 5G and beyond: Fundamentals, technology integration, and state-of-the-art", *IEEE Access*, vol. 8, pp. 116974-117017, 2020. DOI: 10.1109/ACCESS.2020.3001277
- [16] L. Satish, and B. I. Gururaj, "Use of hidden Markov models for partial discharge pattern classification", *IEEE Transactions on Dielectrics and Electrical Insulation*, April 2003.
- [17] R. Poli, W. B. Langdon, and N. F. McPhee, *A Field Guide to Genetic Programming*, 2008. [Online]. Available: Lulu.com. ISBN 978-1-4092-0073-4.
- [7] N. Hassan, K. A. Yau, and C. Wu, "Edge computing in 5G: A review", in *IEEE Access*, vol. 7, pp. 127276-127289, 2019. DOI: 10.1109/ACCESS.2019.2938534
- [8] Shi, Weisong, et al., "Edge computing: Vision and challenges", *IEEE Internet of Things Journal*, no. 3.5, pp. 637-646, 2016.
- [9] PremSankar, Gopika, Mario Di Francesco, and Tarik Taleb, "Edge computing for the Internet of Things: A case study", *IEEE Internet of Things Journal*, no. 5.2, pp. 1275-1284, 2018.
- [10] Li, Hongxing, et al. "Mobile edge computing: Progress and challenges", in *4th IEEE Int. Conf. on mobile cloud computing, services, and engineering (MobileCloud)*, IEEE, 2016.
- [11] Wang, Shangguang, et al., "Edge server placement in mobile edge computing", *Journal of Parallel and Distributed Computing*, no. 127, pp. 160-168, 2019.
- [12] Wang, Shangguang, et al., "A survey on service migration in mobile edge computing", *IEEE Access*, no. 6 pp. 23511-23528, 2018.
- [13] Liu, Juan, et al. "Delay-optimal computation task scheduling for mobile-edge computing systems", in *IEEE Int. Symposium on Information Theory (ISIT)*, IEEE, 2016.
- [14] J. Hammer, P. Moll, and H. Hellwagner, "Transparent access to 5G edge computing services", in *IEEE Int. Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, Rio de Janeiro, Brazil, 2019, pp. 895-898. DOI: 10.1109/IPDPSW.2019.00147
- [15] Q. Pham et al., "A survey of multi-access edge computing in 5G and beyond: Fundamentals, technology integration, and state-of-the-art", *IEEE Access*, vol. 8, pp. 116974-117017, 2020. DOI: 10.1109/ACCESS.2020.3001277
- [16] L. Satish, and B. I. Gururaj, "Use of hidden Markov models for partial discharge pattern classification", *IEEE Transactions on Dielectrics and Electrical Insulation*, April 2003.
- [17] R. Poli, W. B. Langdon, and N. F. McPhee, *A Field Guide to Genetic Programming*, 2008. [Online]. Available: Lulu.com. ISBN 978-1-4092-0073-4.

References

- [1] Cisco Annual Internet Report. [Online]. Available: <https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=2055169>
- [2] Ericsson Mobility Report. [Online]. Available: <https://www.ericsson.com/en/mobility-report>
- [3] M. A. Makolkina, A. A. Atea, A. S. A. Muthanna, and A. E. Kucheryaviy, "The method of unloading augmented reality application traffic in a multilevel boundary computing system", *Elektrosvyaz*, no. 6, pp. 36-42, 2019. [in Russian].
- [4] D. V. Kashkarov, and A. E. Kucheryaviy, "Analysis of applications and prospects for the development of border computing technologies with multiple access in communication networks", *Informatsionnyye Tekhnologii i Telekommunikatsii*, no. 8.1, pp. 28-33, 2020. [in Russian].
- [5] A. A. Ateya, A. I. Vybornova, and A. E. Kucheryaviy, "Multi-tier cloud architecture for Tactile Internet services", *Elektrosvyaz*, no. 2, pp. 26-30, 2017. [in Russian].
- [6] A. S. Viktorov, and V. N. Shvedenko, "Methods for creating an efficient distributed data warehouse for a UAV telemetry peripheral data processing service", *Informatsionno-ekonomicheskiiye aspekty standartizatsii i tekhnicheskogo regulirovaniya*, no. 1, 2018. [in Russian].

P. S. Usik¹, graduate student,
e-mail: mr.usik@ukr.net

O. A. Smirnov¹, Dr.Sc., professor,
e-mail: dr.SmirnovOA@gmail.com

I. V. Myronets², Ph.D., associate professor,
e-mail: i.myronets@chdtu.edu.ua

K. O. Buravchenko¹, Ph.D.,
e-mail: buravchenkok@gmail.com

N. M. Yakymenko¹, Ph.D., associate professor
e-mail: yakimenko_n_m@ukr.net

¹Central Ukrainian National Technical University
University ave., 8, Kropyvnytskyi, 25006, Ukraine

²Cherkasy State Technological University
Shevchenko blvd, 460, Cherkasy, 18006, Ukraine

METHOD OF IMPROVING THE EFFICIENCY OF DISTRIBUTED DATA PROCESSING IN COMPUTER SYSTEMS OF CELLULAR OPERATORS

The creation of information society in Ukraine is one of the most urgent tasks today. In the Strategy for the Development of the Information Society in Ukraine, the priorities for the formation of the country's modern information infrastructure include the creation of high-speed mobile broadband access networks to Internet resources throughout Ukraine.

At the same time, with the development of cellular networks, new and more advanced network architectures for data transmission and management appear. However, there are still a number of unresolved issues and problem areas that need to be addressed accordingly. Therefore, the creation of high-speed networks of the fourth and fifth generations of broadband access to Internet resources and increase of the efficiency of their operation is an urgent and promising task.

The purpose of the study is to develop methods to improve the efficiency of distributed data processing in computer systems of cellular operators. Achieving the goal is aimed at improving the quality of customer service of modern cellular networks and meeting the requirements for next generation networks.

The paper analyzes the effectiveness of methods, models and technologies of distributed data processing in computer systems of cellular operators and identifies their shortcomings. After that, a method has been developed to optimize the placement of scalable services on the distributed computing resources of the cellular network, which consists in the sequential use of the boundary computing model, a generalized cellular network model and a heuristic solution based on genetic algorithms.

The method obtained as a result of this work allows to reduce the level of degradation of the quality of service of end subscribers of the cellular operator's network, in particular, delays of up to 8 ms for a large number of subscribers and, accordingly, a large number of tasks.

Keywords: cellular communication, data transmission, distributed computing, network channels, genetic algorithm.