

## МОДЕЛЮВАННЯ ПЕРЕДУМАННИХ ОБЧИСЛЕНЬ ДЛЯ ТАКТИЛЬНОГО ІНТЕРНЕТУ

<sup>1</sup>Харківський національний університет радіоелектроніки

*В сучасному світі інформаційних технологій та в умовах зростаючого навантаження на комп'ютерні мережі актуальним завданням є їхня оптимізація та поліпшення продуктивності, шляхом ефективного керування ресурсами та зниження затримки. Побудова архітектури інформаційних технологій здатна зменшити затримку шляхом переміщення хмарних структур на кордон мереж радіодоступу. Передуманні обчислення, які передбачають обробку даних на межі комп'ютерної мережі, зменшують затримку та підвищують швидкість відповіді, а використання пропускну здатності на периферії допомагає зменшити навантаження на смугу пропускання. Передуманні обчислення є важливою стратегією для поліпшення комп'ютерних мереж в середовищі Тактильного Інтернету. Метою роботи є моделювання передуманних обчислень ієрархічної мережі граничних хмар, спрямованих для визначення затримки під час передачі трафіку, оптимізації продуктивності та управління ресурсами. Об'єктом дослідження є модель ієрархічної мережі граничних хмар, включно з граничними та передуманними кластерами, туманними та хмарними обчисленнями. Предметом дослідження є модель передачі трафіку в ієрархічній хмарній мережі, для забезпечення оптимального управління ресурсами та передачі трафіку, враховуючи вимоги до затримки. У роботі змодельовано ієрархічну мережу граничних хмар, розроблено модель передачі трафіку та проведено аналіз затримки кластерів першого та другого рівнів ієрархічної хмарної мережі. Ієрархічна мережа граничних хмар розроблена для оптимізації передачі даних і управління ресурсами. Граничні кластери мають обмежені обчислювальні можливості, тому з'єднуються з потужнішими передуманними кластерами. До того ж, туманні обчислення забезпечують узгоджену взаємодію передуманних кластерів в рамках всієї комп'ютерної мережі. Модель передачі трафіку дає змогу досягти необхідної затримки, ефективності, гарантує безпеку та високу доступність, що надає їй актуальності та користі для середовища Тактильного Інтернету. Переваги змодельованої комп'ютерної мережі полягають у скороченні затримки від джерела даних до користувачів і зниженні ризику перевантаження мережі. При цьому забезпечується гнучкість у побудові мережі та підвищується її доступність, що задовольняє вимоги Тактильного Інтернету.*

**Ключові слова:** Тактильний Інтернет, комп'ютерна мережа, передуманні обчислення, хмарна мережа, граничний кластер, передуманий кластер, модель передачі трафіку.

### Вступ

Ефективне управління ресурсами та оптимізація затримки стають дедалі важливішими в умовах зростаючого навантаження на комп'ютерні мережі. Розвиток передуманних обчислень для Тактильного Інтернету є ключовим напрямком сучасних мережних технологій. Передуманні обчислення — це розподілена архітектура інформаційних технологій, яка передбачає аналіз і оброблення даних на границі комп'ютерної мережі. Попит на передуманні обчислення зазвичай пов'язаний з необхідністю скорочення часу відгуку додатків, а затримка має на це прямий вплив [1]. Передуманні обчислення пропонують низку переваг порівняно з моделями централізованих обчислень, як-от зниження затримки і збільшення швидкості відповіді комп'ютерної мережі, адже дані не потрібно передавати на великі відстані до центрального сервера або хмари. До того ж, передуманні обчислення сприяють збереженню приватності та безпеки інформації, оскільки важливі дані можуть не залишати локальну мережу. Розглядаючи цю концепцію в контексті Тактильного Інтернету, можна побачити її особливу цінність: вона дає змогу швидко реагувати на зміни в навколишньому середовищі без необхідності постійного зв'язку з центральним сервером. Серед інших переваг є

ефективне використання пропускної здатності — збільшення обсягу обробки даних на периферії знижує навантаження на смугу пропускання комп'ютерної мережі, звільняючи ресурси для необхідних робочих навантажень. Сьогоднішня побудова архітектури інформаційних технологій здатна зменшити затримку шляхом переміщення хмарних структур на кордон мереж радіодоступу RAN (Radio Access Network) [2]. Багато дослідників пропонують розміщувати невеликі хмарні структури на кордоні мобільної мережі. Більше того, термін «передуманні обчислення», знайшов широке застосування не тільки для кордонів безпроводних мереж, а й для стільникових мереж [3]. Системи, в яких хмару використовують в кожній базовій станції стільникової мережі, стали називати «зеленою» хмарною мережею [4]. «Зелена» хмарна мережа є гетерогенною і дає змогу в деяких зонах комп'ютерної мережі під'єднувати кожен базову станцію до хмари, забезпечуючи на невеликій місцевості обслуговування за допомогою однієї хмари групи базових станцій. Таке рішення базується на показнику щільності розміщення сот на місцевості.

*Мета дослідження* — моделювання передуманних обчислень ієрархічної мережі граничних хмар, спрямованих для визначення затримки під час передачі трафіку, оптимізацію продуктивності та доступності. Дослідження та оптимізація ієрархічної мережі граничних хмар, що розглядається як розвиток концепції передуманних обчислень містить такі задачі:

- 1) розробка ієрархічної мережі граничних хмар з чотирма рівнями обчислень;
- 2) розгляд характеристик кожного рівня та їхній вплив на ефективність мережі;
- 3) оцінка та порівняння затримок передачі даних від граничних кластерів до передуманних кластерів;
- 4) визначення параметрів, які сприятимуть впровадженню Тактильного Інтернету в ієрархічній хмарній мережі.

### Моделювання ієрархічної хмарної мережі

У рамках переходу від централізованих хмарних обчислень до гетерогенних розподілених хмарних структур, змодельовано ієрархічну мережу граничних хмар, яка є інноваційним підходом до побудови хмарної мережі, спрямованим на оптимізацію її продуктивності та доступності. Для виконання граничних обчислень малі стільники взаємодіють з граничними кластерами. Граничні кластери мають обмежені обчислювальні можливості, тому з'єднуються з потужнішими передуманними кластерами. До того ж, туманні обчислення забезпечують узгоджену взаємодію передуманних кластерів в рамках всієї комп'ютерної мережі [5]. На основі розподілених хмарних структур змодельовано чотирирівневу ієрархічну хмарну мережу передуманних обчислень (рис. 1), яка взаємодіє через високошвидкісні оптичні кабелі та безпроводні оптичні мережі зв'язку (Wi-Fi). Така ієрархічна мережа граничних хмар містить рівні:

- 1) граничний кластер; 2) передуманний кластер;
- 3) туманні обчислення; 4) хмарні обчислення.

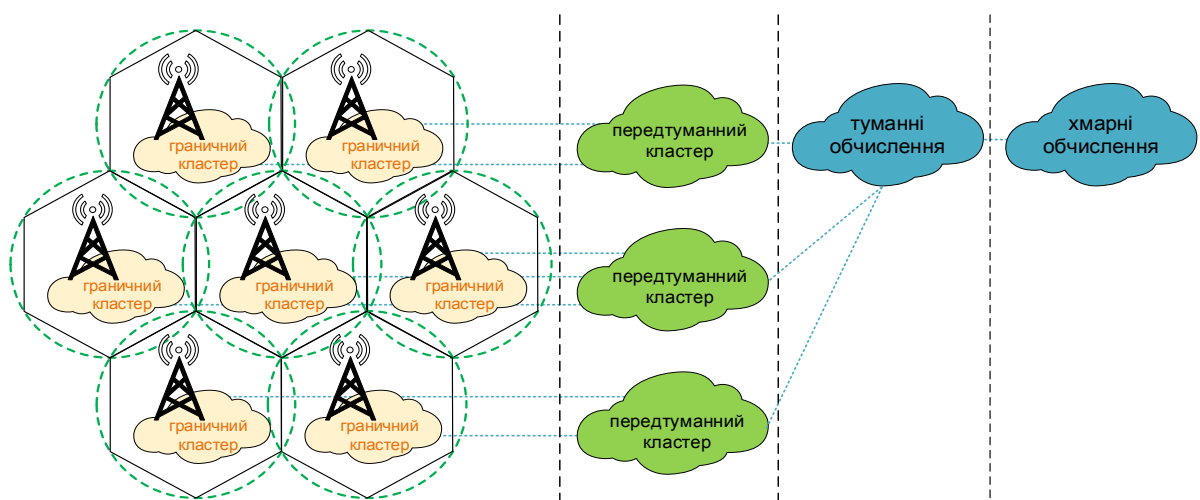


Рис. 1. Модель ієрархічної хмарної мережі для оптимізації Тактильного Інтернету

Граничний кластер — перший рівень ієрархічної хмарної мережі, безпосередньо інтегрований з базовою станцією. Може обслуговуватися, наприклад, як пікосотами, так і фемтосотами — мало-

потужними безпроводними точками доступу, керованими оператором. Пікосота — це невеликий виносний блок для приймання та передавання сигналу, який під'єднується до операторського контролера базової станції через інтернет-мережу. Фемтосота — це малопотужна мініатюрна базова станція стільникового зв'язку. Робота і взаємодія граничних кластерів контролюється передуманним кластером. У разі обмежених ресурсів або ризику перевантаження комп'ютерної мережі, запит на обслуговування може бути переданий на другий рівень.

Передуманий кластер — центр обробки та зберігання даних, який взаємодіє з кількома граничними кластерами, використовуючи високошвидкісний оптичний зв'язок. Побудований у вигляді гетерогенної комп'ютерної мережі, оскільки може містити різну кількість кластерів першого рівня, які працюють у різних стандартах та за різними технологіями. При цьому всі вони утворюють єдине інтегроване середовище, що працює як єдина мережа.

Туманні обчислення — третій рівень ієрархічної хмарної мережі, що володіє значно більшими обчислювальними ресурсами і можливостями зберігання даних порівняно з передуманими кластерами і відповідно керує ними. Туманні обчислення — це децентралізована обчислювальна модель, розміщена між пристроями, комп'ютерними мережами, центрами обробки даних або хмарою, що допомагає задовольнити вимоги ТІ. Разом з тим, забезпечується підтримка мобільності користувачів, покращена масштабованість, розуміння місцезнаходження, зв'язок в реальному часі та покращена здатність системи до механізмів хмарних обчислень. Туманні обчислення зазвичай включають в себе туманні вузли, такі як проксі-сервери, комутатори, маршрутизатори, шлюзи або різні туманні пристрої [6].

Хмарні обчислення відіграють ключову роль у забезпеченні ефективної та безпечної роботи всієї хмарної інфраструктури і узгодженні функціонування всіх рівнів. Під час хмарних обчислень дані збираються і передаються у хмару для подальшого аналізу та обробки. Цей процес забезпечується потужними обчислювальними ресурсами та масштабованістю. Виконують як керування ресурсами та комп'ютерною мережею, так і забезпеченням безпеки та високої доступності. Переваги змодельованої комп'ютерної мережі полягають у скороченні затримки від джерела даних до користувачів і зниженні ризику перевантаження мережі [7]. При цьому забезпечується гнучкість у побудові мережі та підвищується її доступність, що задовольняє вимоги Тактильного Інтернету.

### Аналіз затримки кластерів першого та другого рівнів ієрархічної хмарної мережі

Одним з найважливіших факторів, що впливають на значення затримки, є кількість мережних вузлів, що беруть участь у процесі забезпечення взаємодії між користувачами. У змодельованій ієрархічній хмарній мережі одну з ключових ролей для зменшення затримки відіграє передуманий кластер, який безпосередньо взаємодіє з кластером першого рівня (рис. 2).

Передача даних від граничного до передуманного кластера містить три аспекти затримки: затримку передачі по висхідних і низхідних лініях, затримку розповсюдження і затримку оброблення даних, і може бути визначена як сума цих трьох затримок

$$T = (T_B + T_H) + T_{3P} + T_{3O}, \quad (1)$$

де  $T_B + T_H$  — затримка передачі по висхідних і низхідних лініях відповідно;  $T_{3P}$  — затримка розповсюдження;  $T_{3O}$  — затримка обробки даних. Розглянемо кожен з аспектів затримки. Сукупна затримка для передачі по висхідних і низхідних лініях ( $T_B, T_H$ ) може бути обчислена, за використання формули [8]

$$T_B = (1 + V_B)(K_B / S_B); \quad (2)$$

$$T_H = (1 + V_H)(K_H / S_H), \quad (3)$$

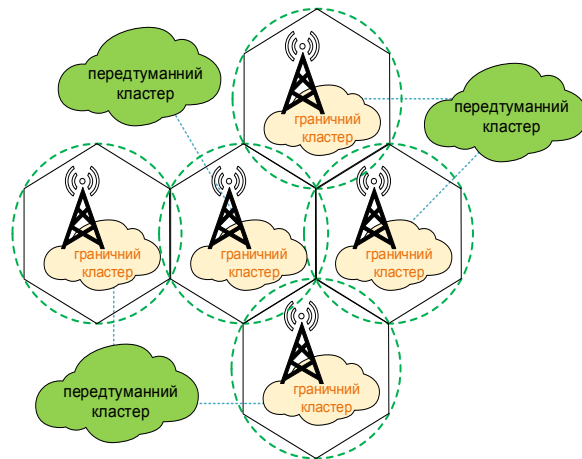


Рис. 2. Схема взаємодії кластерів першого та другого рівнів

де  $V_B$  та  $V_H$  — швидкості передачі для висхідної та низхідної ліній, при перенавантаженні;  $K_B$  та  $K_H$  — загальне число переданих біт по висхідній і низхідній лініях;  $S_B$  та  $S_H$  — швидкості передачі даних для висхідної та низхідної ліній відповідно.

Затримка розповсюдження  $T_{3P}$  є функцією відстані і може бути розрахована у такий спосіб:

$$T_{3P} = L_{KL} / U_{3P}, \tag{4}$$

де  $L_{KL}$  — відстань між користувачем і кластером;  $U_{3P}$  — швидкість розповсюдження.

Затримка через обробку даних  $T_{3O}$  ґрунтується на кількості необхідних операцій і швидкості обробки інформації процесором у кластері та обчислюється таким чином:

$$T_{3O} = K_1 / U_{KL}, \tag{5}$$

де  $K_1$  — кількість операцій;  $U_{KL}$  — швидкість процесора кластера.

### Визначення затримки в моделі ієрархічної хмарної мережі

Щоб досягти малої затримки в мережі Тактильного Інтернету з необхідною продуктивністю потрібно визначити всі параметри, що мають значення. Для цього створимо модель передачі даних в середовищі Тактильного Інтернету, на основі розглянутої вище ієрархічної хмарної мережі.

На рис. 3 показана модель, в якій кожна базова станція взаємодіє з граничним кластером  $H_{ГК}(i)$ , де  $i \in \{1, 2, \dots, M\}$ , де  $M$  — кількість граничних кластерів. Граничні кластери в свою чергу пов'язані з передтуманними кластерами  $H_{ГК}(i)$  де  $j \in \{1, 2, \dots, N\}$ , де  $N$  — кількість передтуманних кластерів у мережі. Кожен передтуманий кластер також містить контролер для управління граничними кластерами та надає шлюзи для взаємодії граничних кластерів з туманними обчисленнями за необхідності. У цій моделі передбачається, що кожен передтуманий кластер має з'єднання з фіксованою кількістю граничних кластерів  $K$ .



Рис. 3. Модель передачі даних в середовищі Тактильного Інтернету на основі ієрархічної хмарної мережі

Швидкість оброблення даних, переданих у граничний кластер, залежить від їхньої кількості. Якщо допустити, що надходження даних розподіляється за законом Пуассона з інтенсивністю  $\lambda_i$ , то кожна комірка створює навантаження  $w_i$  для граничного кластера ( $H_{ГК}$ ) з інтенсивністю пуассонівського процесу  $\lambda_i$ . Граничний кластер здатний обробити дані від відповідних базових станцій, але якщо необхідні ресурси дорівнюють або перевищують максимальне значення  $w_{h \max}(i)$ , то нові запити будуть спрямовані у передтуманий кластер, поки ресурси граничного кластера зайняті. У випадку, якщо запити обробки даних потребують більших ресурсів ніж наявні наразі в передтуманному кластері, то нові запити відповідно будуть спрямовуватись на наступний рівень

ієрархічної хмарної мережі.

Загальна затримка в цьому випадку складається з часу відгуку на запит і часу взаємодії. Середній час відгуку на запит для граничних кластерів і переддуманих кластерів складається з очікування в черзі та часу обробки запитів. Середній час опрацювання запитів у граничних кластерах і переддуманих кластерах можна визначити як функцію інтенсивності запитів  $\lambda$  для граничних кластерів і переддуманих кластерів та формули Ерланга відповідно [9].

$$T_{ГК(i)}(\lambda) = \frac{H\left(K_i, \frac{\lambda_i}{\mu_i}\right)}{K_i \mu_i - \lambda_i} + \frac{1}{\mu_i}, \quad (6)$$

де  $T_{ГК(i)}$  — середня тривалість обробки запитів у граничному кластері  $i$ ,  $K_i$  — загальна кількість серверів у граничному кластері  $i$ ,  $\lambda_i$  — інтенсивності надходження заявок на обслуговування у граничному кластері  $i$ ,  $\mu_i$  — інтенсивності обслуговування заявок у граничному кластері  $i$ .

$$T_{ПК(j)}(\lambda) = \frac{H\left(K_j, \frac{\lambda_j}{\mu_j}\right)}{K_j \mu_j - \lambda_j} + \frac{1}{\mu_j}; \quad (7)$$

$$H(n, \rho) = \frac{\left(\frac{(K\rho)^H}{n!}\right)\left(\frac{1}{1-\rho}\right)}{\sum_{k=0}^{n-1} \frac{(n\rho)^k}{k!} + \left(\frac{(n\rho)^H}{n!}\right)\left(\frac{1}{1-\rho}\right)}. \quad (8)$$

Для спрощення обчислення загальної затримки припускаємо, що функція затримки має лінійний характер, що є допустимим і перевіреним припущенням, як зазначено в дослідженні [10]. Таким чином, загальна затримка для граничних кластерів і переддуманих кластерів може бути визначена як

$$T_{ГК(i)}(w_i) = f_H(w) + d = \left[ a(w_{d_{ГК(i)}}) + \beta \right] + d, \quad (9)$$

де  $T_{ГК(i)}$  — загальна затримка при вивантаженні запиту у граничному кластері  $i$ ,  $f_H$  — лінійна функція для визначення затримки обслуговування запиту для поточного навантаження,  $d$  — затримка при з'єднанні всередині стільника,  $w_{d_{ГК(i)}}$  — робоче навантаження, що надходить.

$$T_{ПК(j)}(w_j) = f_H(w) + d = \left[ a(w_{d_{ГК(j)}}) + \beta \right] + d + d_{H_{ГК(i)}, H_{ПК(j)}}, \quad (10)$$

де  $T_{ПК(j)}(w_j)$  — загальна затримка під час вивантаження запиту у переддуманному кластері  $j$ ;  $d_{H_{ГК(i)}, H_{ПК(j)}}$  — затримка під час взаємодії між граничним і переддуманим кластером.

### Модель передачі трафіку в ієрархічній хмарній мережі

Візьмемо до уваги два рівні передачі трафіку для мережних додатків. Перший рівень передачі трафіку передбачає передачу на граничний кластер, а другий — на переддуманий кластер. Можливість швидкого реагування може бути забезпечена на нульовому рівні, коли не виникає потреба в передачі запитів на інші рівні. Цей рівень використовується у випадках, коли мережний пристрій має достатні ресурси для вирішення необхідних завдань з необхідною якістю обслуговування (QoS) та з меншими енергетичними витратами ніж ці запити можуть бути виконані на граничному кластері, і водночас рівень заряду батареї дозволяє їхнє вирішення. На рис. 4 показані різні рівні передачі трафіку та відповідні запити, які вирішуються на кожному рівні.

Мережний пристрій, так само як і граничні сервери, використовують контролер прийняття рішень, який визначає, чи варто обслуговувати трафік локально або передавати його на відповідний сервер. Система враховує всі необхідні фактори для прийняття кінцевого рішення.



Підхід мережного пристрою до ухвалення рішення про передачу трафіку від самого початку передбачає використання контролера прийняття рішень. Визначається загальна довжина вхідних

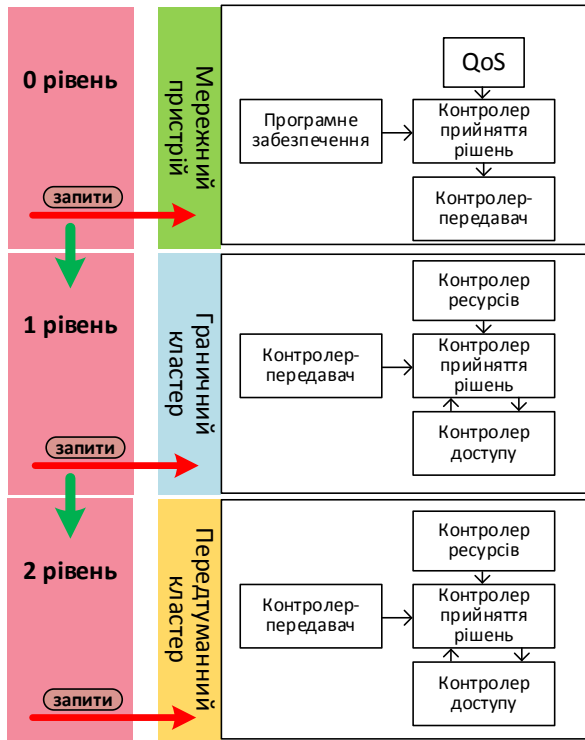


Рис. 4. Структурна схема рівнів передачі трафіку

даних запиту  $L$  і загальну кількість циклів  $N$  CPU (Central Processing Unit), необхідних для виконання запиту з даними довжини  $L$ . Мережні пристрої використовують цю інформацію для визначення, чи слід обслуговувати запит або передавати його далі. Це залежить від поточних ресурсів мережного пристрою і необхідної затримки обслуговування ( $\tau$ ), необхідної для забезпечення потрібного рівня якості обслуговування.

Мережний пристрій визначає ймовірну тривалість (мс) обслуговування запиту ( $T_{МП}$ ), ґрунтуючись на доступних ресурсах,

$$T_{МП} = \frac{N}{R_{МП}}, R_{МП} \in f_{МП}, \quad (11)$$

де  $R_{МП}$  — мережні ресурси, необхідні для обслуговування запиту;  $f_{МП}$  — загальні ресурси мережних пристроїв;  $N$  — загальна кількість циклів CPU, необхідних для обслуговування запиту з даними довжини  $L$ .

Далі мережний пристрій обчислює бінарне значення часу ухвалення рішення ( $D_{T_{МП}}$ ), порівнюючи  $T_{МП}$  і необхідну затримку обслуговування ( $\tau$ )

$$D_{T_{МП}} = I(T_{МП}, \tau) = \begin{cases} 1 & \text{IF } (T_{МП} \leq \tau), \\ 0 & \text{IF } (T_{МП} > \tau), \end{cases} \quad (12)$$

де  $I$  — змінна індикатора режиму.

Кожне бінарне значення часу ухвалення рішення розглядається як конкретний сценарій.

*Перше бінарне рішення.* Коли тривалість обслуговування запиту менша за максимально допустиму затримку, обслуговування запиту може бути виконано локально самим мережним пристроєм, і бінарне рішення в цьому разі дорівнює 1. Мережний пристрій обчислює споживання енергії для локального обслуговування заявки ( $E_{МП}$ ), і відправляє запит до відповідного граничного серверу.

$$E_{МП} = N\delta_{МП}, \quad (13)$$

де  $\delta_{МП}$  — споживання енергії на один цикл CPU мережним пристроєм.

На основі повідомлення відповіді від граничного кластера, мережний пристрій ухвалює рішення про передачу даних або виконання обробки локально. Граничний кластер оцінює можливість обробки завдання мережним пристроєм на основі майбутніх процедур, які мають бути виконані. Граничний кластер передає повідомлення відповіді мережному пристрою, який, виходячи з отриманої інформації, вирішує обслуговувати заявку локально або передати її далі.

*Нульове бінарне рішення.* Якщо необхідна тривалість обслуговування в мережному пристрої перевищує максимально допустиму затримку ( $\tau$ ), то локальне обслуговування неможливе, і ухвалюється рішення про передачу трафіку. У цьому випадку значення змінної дорівнює 0, і мережний пристрій надсилає запит до відповідного граничного серверу. Граничний кластер отримує запит мережного пристрою і реагує відповідним повідомленням. Якщо граничний кластер здатний опрацювати запит, то він надсилає повідомлення, погоджуючись на передачу даних і надання ресурсів для обслуговування; в іншому разі він перенаправляє повідомлення до відповідного переддуманого кластера та очікує на відповідь. Спершу граничний кластер оцінює параметри затримки, отримані в запиті від мережного пристрою а потім обчислює тривалість обслуговування запиту ( $t_{ГК}$ ), ґрунтуючись на доступних ресурсах кластера.

$$t_{ГК} = \frac{N}{R_{ГК}}, R_{ГК} \in f_{ГК}, \quad (14)$$

де  $R_{ГК}$  — ресурси граничного кластера, які необхідні для отримання запиту мережного пристрою;  $f_{ГК}$  — загальні ресурси граничного кластера.

Загальну затримку ( $T_{ГК}$ ) в граничному кластері обчислюють шляхом підсумовування обох розглянутих затримок: затримки оброблення (тривалість обслуговування) і затримки взаємодії (тривалість передавання даних висхідною лінією і тривалість відповіді на обчислення).

$$T_{ГК} = t_{ГК} + t_{ВХ} + t_{ВИХ}; \quad (15)$$

$$t_{ВХ} = \frac{L}{S_{ВХ}}, \quad (16)$$

де  $t_{ГК}$  — загальний час обслуговування запиту в граничному кластері;  $t_{ВХ}$  — час передачі вхідних даних запиту;  $t_{ВИХ}$  — час зворотного зв'язку передачі результатів обчислень;  $S_{ВХ}$  — допустима швидкість передачі даних по висхідній лінії.

Допустима швидкість передачі даних по висхідній лінії обчислюється як функція від потужності передачі, а потужність посилення в каналі визначається з використанням формули Шеннона–Хартлі [12]

$$S_{ВХ} = \omega \log_2 \left( 1 + \frac{hp}{\sigma} \right), \quad (17)$$

де  $\omega$  — смуга пропускання;  $h$  — підсилення каналу;  $\sigma$  — потужність шуму на приймачі.

Механізм прийняття рішення в граничному кластері обчислює бінарне значення часу прийняття рішення ( $D_{ТГК}$ ), порівнюючи затримку в обслуговуванні  $T_{ГК}$  з максимально допустимою затримкою

$$D_{ТГК} = I(T_{ГК}, \tau) = \begin{cases} 1 & \text{IF}(T_{ГК} \leq \tau), \\ 0 & \text{IF}(T_{ГК} > \tau). \end{cases} \quad (18)$$

Якщо рішення про передачу даних позитивне ( $D_{ТГК} = 1$ ), то обслуговування запиту може бути виконано в граничному кластері, якщо рішення нульове — граничний кластер надсилає запит у відповідний передуманний кластер. Наступний етап — це перевірка енергетичних характеристик. Граничний кластер обчислює споживання енергії для обслуговування запиту на кластері  $E_{ГК}$  за формулою

$$E_{ГК} = N\delta_{ГК} + t_{ВХ}P\eta_T, \quad (19)$$

де  $\delta_{ГК}$  — споживання енергії на один цикл CPU в граничному кластері;  $P$  — потужність передачі мережних пристроїв;  $\eta_T$  — ефективність каналу.

Ухвалюється рішення про локальну обробку запиту або передачу даних у граничний кластер.

$$D_{ТГК} = I(E_{МП}, E_{ГК}) = \begin{cases} 1 & \text{IF}(E_{МП} \leq E_{ГК}), \\ 0 & \text{IF}(E_{МП} > E_{ГК}). \end{cases} \quad (20)$$

У разі негативних рішень на рівнях мережного пристрою та граничного кластера останній надсилає запит відповідному передуманному кластерові. Передуманий кластер обчислює загальну затримку обслуговування заявки  $T_{ПК}$ , ґрунтуючись на доступних йому ресурсах

$$T_{ПК} = t_{ПК} + t'_{ВХ} + t'_{ВИХ}; \quad (21)$$

$$t_{ПК} = \frac{N}{R_{ПК}}, R_{ПК} \in f_{ПК}, \quad (22)$$

де  $t_{ПК}$  — загальний час обслуговування заявки у передуманному кластерові;  $R_{ПК}$  — ресурси передуманного кластера, що виділяються для отримання запиту з граничного кластера;  $f_{ПК}$  — за-

льні ресурси переддуманого кластера.

Переддуманний кластер обчислює значення бінарного часу прийняття рішення  $D_{T_{ПК}}$  шляхом порівняння загальної тривалості обслуговування запиту на рівні переддуманого кластера і максимально допустимої тривалості затримки ( $\tau$ ).

$$D_{T_{ПК}} = I(T_{ПК}, \tau) = \begin{cases} 1 & \text{IF } (T_{ПК} \leq \tau), \\ 0 & \text{IF } (T_{ПК} > \tau). \end{cases} \quad (23)$$

Передача даних в переддуманному кластері буде виконуватися за умови, якщо  $D_{T_{ПК}}$  та  $D_{T_{МП}} = 0$ .

### Висновки

В статті змодельовано ієрархічну мережу граничних хмар, яка є розвитком концепції переддуманних обчислень, переходячи від централізованих систем до гетерогенних розподілених структур. У мережі задіяні чотири рівні обчислень з різними характеристиками. Основну увагу в статті приділено першим двом рівням. Перший рівень становлять граничні кластери з обмеженими обчислювальними та пам'ятними ресурсами, безпосередньо пов'язаними з базовими станціями. Другий рівень — переддуманний кластер, представлений потужнішими обчислювальними ресурсами і пам'яттю, керує групами граничних кластерів через високошвидкісні оптичні з'єднання.

Проаналізовано затримки кластерів першого та другого рівнів ієрархічної хмарної мережі. Під час передачі даних від граничного до переддуманого кластера спостерігається три аспекти затримки: затримку передачі по висхідних і низхідних лініях, затримку розповсюдження і затримку обробки даних, яка може бути визначена як сума цих трьох затримок.

Також розроблена модель передачі трафіку в ієрархічній хмарній мережі, яка забезпечує необхідну затримку та енергоефективність, знижує навантаження на переддуманні обчислення хмарної мережі та сприяє реалізації Тактильного Інтернету.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] K. Kaur, A. Singh, and A. Sharma, "A systematic review on resource provisioning in fog computing. Transactions on Emerging Telecommunications Technologies", vol. 34, no. 2, pp. 1-36, January 2023. <https://doi.org/10.1002/ett.4731>.
- [2] A. Aijaz, "Toward human-in-the-loop mobile networks: A radio resource allocation perspective on haptic communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4493-4508, July 2018. <https://doi.org/10.1109/twc.2018.2825985>.
- [3] Fog computing and Mobile Edge Cloud Gain Momentum Open Fog Consortium ETSI MEC and Cloudlets, 2015. [Electronic resource]. Available: [https://yucianga.info/wp-content/uploads/2015/11/15\\_11\\_22\\_Fog\\_computing\\_and\\_mobile\\_edge\\_cloud\\_gain\\_momentum\\_Open\\_Fog\\_Consortium-ETSI\\_MEC-Cloudlets\\_v1\\_1.pdf](https://yucianga.info/wp-content/uploads/2015/11/15_11_22_Fog_computing_and_mobile_edge_cloud_gain_momentum_Open_Fog_Consortium-ETSI_MEC-Cloudlets_v1_1.pdf). Accessed on: Feb 08, 2024.
- [4] X. Sun, and N. Ansari, "Green Cloudlet Network: A Distributed Green Mobile Cloud Network," *IEEE Network*, vol. 31, no. 1, pp. 64-70, January 2017. <http://dx.doi.org/10.1109/mnet.2017.1500293nm>.
- [5] R. Yaroshevych, V. Tkachov, A. Kovalenko, and D. Rosinskyi "Modelling the Domain Architecture of the Tactile Internet Using a Foggy Infrastructure," *IEEE 9th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)*. October 2022, pp. 512-516. <https://doi.org/10.1109/PICST57299.2022.10238653>.
- [6] Aazam, Mohammad et al. "PRE-Fog: IoT trace based probabilistic resource estimation at Fog." *13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, January 2016, pp. 12-17.
- [7] A. Muthanna et al. "Secure and Reliable IoT Networks Using Fog Computing with Software-Defined Networking and Blockchain," *Journal of Sensor and Actuator Networks*, vol. 8, no. 1, pp. 15, February 2019. <https://doi.org/10.3390/jsan8010015>.
- [8] A. Mukherjee, D. De, and D. G. Roy, "Power and Latency Aware Cloudlet Selection Strategy for Multi-Cloudlet Environment," *IEEE Transactions on Cloud Computing*, vol. 7, no. 1, pp. 141-154, July 2019. <http://dx.doi.org/10.1109/tcc.2016.2586061>.
- [9] M. Jia, W. Liang, Z. Xu, M. Huang, and Y. Ma, "QoS-Aware Cloudlet Load Balancing in Wireless Metropolitan Area Networks," *IEEE Transactions on Cloud Computing*, vol. 8, no. 2, pp. 623-634, April-June 2020. <http://dx.doi.org/10.1109/tcc.2017.2786738>.
- [10] K. Intharawijitr, K. Iida, and H. Koga, "Analysis of fog model considering computing and communication latency in 5G cellular networks," *IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, March 2016, pp. 1-4. <https://doi.org/10.1109/PERCOMW.2016.7457059>.
- [11] K. Gai, M. Qiu, H. Zhao, L. Tao, and Z. Zong, "Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing," *Journal of Network and Computer Applications*, vol. 59, pp. 46-54, January 2016. <http://dx.doi.org/10.1016/j.jnca.2015.05.016>.
- [12] A. Kartun-Giles, S. Jayaprakasam, and S. Kim, "Euclidean Matchings in Ultra-Dense Networks," *IEEE Communications Letters*, vol. 22, no. 6, pp. 1216-1219, January 2018. <http://dx.doi.org/10.1109/lcomm.2018.2799207>.



**Коваленко Андрій Анатолійович** — д-р техн. наук, професор, завідувач кафедри електронних обчислювальних машин, e-mail: Andriy\_kovalenko@yahoo.com ;

**Ярошевич Роман Олександрович** — асистент кафедри електронних обчислювальних машин, e-mail: roman.yaroshevych@nure.ua .

Харківський національний університет радіоелектроніки, Харків

**A. A. Kovalenko<sup>1</sup>**  
**R. O. Yaroshevych<sup>1</sup>**

## Modeling of Pre-Fog Computing for the Tactile Internet

<sup>1</sup>Kharkiv National University of Radio Electronics

*In today's world of information technologies and in the conditions of growing load on computer networks, it is an urgent task to optimize and improve their performance by effective management of resources and latency reduction. Construction of information technology architecture can reduce latency by moving cloud structures to the edge of radio access networks. Edge computing, which involves processing data at the edge of the computer network, reduces latency and improves response time, and using bandwidth at the edge helps reduce bandwidth usage. Pre-fog computing is an important strategy for improving computer networks in the Tactile Internet environment. The aim of this paper is to model the pre-fog computing of a hierarchical edge cloud network, aimed at determining the delay in traffic transmission, optimizing performance, and managing resources. The object of research is a model of a hierarchical network of edge clouds, including edge and pre-fog clusters, fog and cloud computing. The subject of the study is a model of traffic transmission in a hierarchical cloud network to ensure optimal resource management and traffic transmission, taking into account the delay requirements. In this paper, we model a hierarchical edge cloud network, develop a traffic transfer model, and analyze the delay of clusters of the first and second levels of a hierarchical cloud network. The hierarchical edge cloud network is designed to optimize data transmission and resource management. Edge clusters have limited computing capabilities, so they are connected to more powerful pre-fog clusters. In addition, fog computing ensures coordinated interaction between edge clusters across the entire computer network. The traffic transmission model allows achieving the required latency, efficiency, security, and high availability, making it relevant and useful for the Tactile Internet environment. The advantages of the modeled computer network are reduced latency from the data source to the users and reduced risk of network congestion. This provides flexibility in building the network and increases its availability, which meets the requirements of the Tactile Internet.*

**Keywords:** Tactile Internet, computer network, fog computing, cloud network, edge cluster, pre-fog cluster, traffic transfer model.

**Kovalenko Andrii A.** — Dr. Sc. (Eng.), Professor, Head of the Chair of Electronic Computers, e-mail: Andriy\_kovalenko@yahoo.com ;

**Yaroshevych Roman O.** — Assistant of the Chair of Electronic Computers, e-mail: roman.yaroshevych@nure.ua