

А. А. Коваленко, Г. А. Кучук, О. С. Ляшенко

<sup>1</sup> Харківський національний університет радіоелектроніки, Харків, Україна

<sup>2</sup> Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

## РОЗПОДІЛ РЕСУРСІВ БАГАТОФАЗНОЇ СИСТЕМИ ОБРОБКИ ВЕЛИКИХ ДАНИХ ПРИ ВИСОКОІНТЕНСИВНОМУ ВХІДНОМУ ПОТОЦІ

Завдання оптимального проектування багатофазної системи обробки великих даних (БФСОВД) при високоінтенсивному вхідному потоці потребують застосування сучасних математичних і комп'ютерних методів та засобів. **Предметом дослідження** є ресурси багатофазної системи обробки великих даних. Мета дослідження – розроблення методу оптимального розподілу ресурсів БФСОВД при високоінтенсивному вхідному потоці, котрий враховує як можливість проведення аналізу даних у режимі реального часу, так і особливостей високоінтенсивного потоку вхідних даних із багатьох джерел. **Результати.** Визначено особливості високоінтенсивних потоків вхідних даних та показана можливість апроксимації його з достатньою точністю гауссівським розподілом. Також визначені особливості функціонування багатофазних систем та запропонована схема функціонування багатофазної системи обробки великих даних. Базуючись на отриманих результатах розроблена математична модель процесу оптимізації розподілу ресурсів багатофазної системи обробки великих даних при високоінтенсивному вхідному потоці. Для цього використано метод динамічного програмування. **Висновок.** В результаті отримана мінімальна кількість каналів в кожній фазі БФСОВД, при якій середній час аналізу даних в системі буде мінімальним.

**Ключові слова:** великі дані, багатофазна система масового обслуговування, обчислювальний ресурс, високоінтенсивний потік.

### Вступ

**Аналіз проблеми.** Зростаючий розвиток засобів обчислювальної техніки і інфокомунікаційних технологій привели до створення систем, які призначені для обробки даних і розподілені на великій території. У вузлах таких систем зосереджені потужні обчислювальні ресурси, а самі вузли пов'язані між собою швидкодіючими каналами. Такого роду системи відносяться до класу систем розподіленої обробки даних (СРОД), аналіз і оптимізація функціонування яких потребують застосування сучасних математичних і комп'ютерних методів та засобів. Зокрема, в окремий клас можна виділити завдання оптимального проектування багатофазної системи обробки великих даних при високоінтенсивному вхідному потоці. Такі завдання виникають при вирішенні задач управління і контролю в умовах великого числа керованих об'єктів і контролюємих параметрів, вони вимагають створення систем обробки великих даних (СОВД), що дозволяють користувачам оперативно працювати в режимі реального часу (РРЧ), виконуючи функції додаткового аналізу з урахуванням ряду факторів, що не піддаються формалізації. Регулярний обмін інформацією між системою розподіленої обробки даних і інформаційними джерелами великих даних здійснюється з використанням множини вхідних інформаційних джерел систем відображення інформації (ВІД), та системи програмно-апаратних засобів, що забезпечують одночасну обробку вхідної інформації в режимі реального часу, працюють незалежно один від одного, формування звичного характеру пропонованої інформації; збереження інформації при відмовах і збоях тощо [1–6].

**Аналіз наукових публікацій.** При роботі з розподіленими даними в середовищі СРОД, при високоінтенсивному вхідному потоці великих даних, що потребують декілька кроків обробки в режимі реального часу, зазвичай використовуються багатофазні

СОВД (БФСОВД). БФСОВД характеризується наявністю декількох апаратно-технічних засобів (АТЗ), які послідовно обслуговують запити системи та виконують попередню обробку інформації. Аналіз таких систем є складним завданням. Розроблені методи аналізу засновані на зведенні системи до набору еквівалентних однофазних систем, для кожної з яких отримують імовірнісні стани системи в сталому режимі шляхом вирішення рівнянь Колмогорова. Однак при наявності в деяких фазах системи великої кількості альтернативних каналів обслуговування моделювання процесу функціонування системи при такому підході викликає значні труднощі. На сьогодні існує велика кількість всіляких підходів для моделювання таких або подібних систем [7–22]. Але в усіх роботах, що розглянуті, не приділено достатньої уваги випадку знаходження мінімальних витрат обчислювального ресурсу СОВД при врахуванні як можливості проведення аналізу даних у режимі реального часу, так і особливостей високоінтенсивного потоку вхідних даних із багатьох джерел.

**Постановка задачі.** Метою статті є розроблення методу оптимального розподілу ресурсів багатофазної системи обробки великих даних при високоінтенсивному вхідному потоці, котрий враховує як можливість проведення аналізу даних у режимі реального часу, так і особливостей високоінтенсивного потоку вхідних даних із багатьох джерел.

### 1. Особливості високоінтенсивних потоків вхідних даних

Випадковий потік однорідних подій є високоінтенсивним, якщо його інтенсивність має вигляд

$$\Lambda = \lambda \cdot N, \quad (1)$$

де  $\lambda > 0$  – деяка кінцева величина;  $N$  – значення великого параметра. У теоретичних дослідженнях передбачається, що  $N \rightarrow \infty$ , а в практичних реалізаціях його значення є достатньо великим (залежить

від системи, що розглядається, зазвичай, близько 100, або більше). Високоінтенсивний потік є гауссівським (ВІГ-поток), якщо розподіл ймовірностей кількості його подій, що наступили за час  $t$ , можна апроксимувати з достатньою точністю гауссівським розподілом з математичним очікуванням  $a$  і дисперсією  $\sigma^2$ :

$$a = \lambda \cdot N \cdot t, \quad \sigma^2 = (\lambda + \kappa) \cdot N \cdot t = \lambda \cdot N \cdot t + \kappa \cdot N \cdot t, \quad (2)$$

де  $\kappa$  – деякий параметр дисперсії).

У [23] показано що для широкого класу високоінтенсивних потоків (пуассонівських, рекурентних, ММРР, МАР і ін.) апроксимація за допомогою (1) є допустима, а величини  $a$  і  $\sigma^2$  можуть бути представлені в формі (2). Перелічимо основні властивості ВІГ-потоків:

– сума незалежних ВІГ-потоків із заданими параметрами інтенсивностей  $\lambda_k$  і параметрами дисперсій  $\kappa_k$  також є ВІГ-поток, параметри  $\lambda$  і  $\kappa$  якого дорівнюють сумам відповідних параметрів складаємих потоків, тобто

$$\lambda = \sum_k \lambda_k; \quad \kappa = \sum_k \kappa_k; \quad (3)$$

– просіяний з ймовірністю  $\tau$  ВІГ-потік з параметрами  $\lambda$  і  $\kappa$  є також ВІГ-поток з параметрами  $r \cdot \lambda$  і  $r^2 \cdot \kappa$ .

Відзначимо важливу особливість просіяного ВІГ-поток, яка полягає в тому, що дисперсія просіяного потоку змінюється більш значно, ніж його середнє значення. Такі зміни параметрів ВІГ-поток при досить малих значеннях  $r$  дозволяють апроксимувати просіяний ВІГ-потік високоінтенсивним пуассонівським потоком.

Розподілений за поліноміальною схемою з ймовірностями  $p_1, p_2, \dots, p_k, \dots, p_K$  ВІГ-потік з параметрами  $\lambda$  і  $\kappa$  є  $K$ -мірним корельованим ВІГ-поток, багатовимірний розподіл ймовірностей якого є  $K$ -мірним гауссівським, визначається багатовимірною характеристичною функцією  $h(\mathbf{u})$  векторного аргументу  $\mathbf{u} = \{u_1, u_2, \dots, u_k, \dots, u_K\}$  такого вигляду:

$$h(\mathbf{u}) = \exp \left( \begin{array}{l} \sum_{\ell=1}^K (i \cdot u_\ell \cdot p_\ell \cdot \lambda \cdot N \cdot t) + \\ + \sum_{\ell=1}^K (i \cdot u_\ell)^2 / 2 - \\ - \sum_{\ell=1}^K \sum_{v=1}^K (u_\ell \cdot p_\ell \cdot u_v \cdot p_v \cdot \lambda \cdot N \cdot t \cdot \kappa / 2) \end{array} \right), \quad (4)$$

тут  $i = \sqrt{-1}$  – уявна одиниця.

Таким чином, кожна  $k$ -та компонента розділеного потоку є ВІГ-поток з параметрами  $p_k \cdot \lambda$  і  $p_k^2 \cdot \kappa$ , але потоки є стохастично залежними. Елементи коваріаційної матриці для них мають вигляд

$$C_{\ell v} = p_\ell \cdot p_v \cdot \lambda \cdot N \cdot t \cdot \kappa \quad (5)$$

для  $\ell$ -го і  $v$ -го з розділених потоків.

## 2. Особливості функціонування багатofазних систем

Розглянемо найпростішу систему масового обслуговування (СМО) без черги з підготовкою каналів. Нехай на  $n$ -каналну СМО надходить найпрос-

тіший потік запитів з інтенсивністю  $\lambda$ , а час обслуговування заявок розподілено за показовим законом з параметром  $\mu$ . До початку обслуговування канал повинен бути підготовлений. Час підготовки  $t_{\text{подг}}$  має показовий розподіл з параметром  $\phi$  і не залежить від того, як давно канал припинив роботу. При надходженні запиту спочатку виконується операція підготовки, а потім він надходить на обслуговування. Запит, який застав всі канали зайнятими, на обслуговування не приймається.

В даному випадку обслуговування запиту складається з двох фаз: підготовки з часом підготовки  $t_{\text{подг}}$  і самого обслуговування протягом часу  $t_{\text{обсл}}$ . В результаті загальний середній час обслуговування складає

$$\tilde{t}_{\text{обсл}} = \bar{t}_{\text{подг}} + \bar{t}_{\text{обсл}}. \quad (6)$$

Випадкова величина  $\bar{t}_{\text{обсл}}$  зазвичай розподілена за узагальненим законом Ерланга другого порядку [9] з параметрами  $\mu$  і  $\phi$ .

Відомо, що формули Ерланга є вірними не тільки для показового, а й для будь-якого іншого розподілу часу обслуговування. Для вирішення розрахунків при розподілі ресурсів нам необхідно знайти величину  $\tilde{m}$ :

$$t_{\text{обсл}} = 1/\mu + 1/\phi = (m + \phi) / (m \cdot \phi), \quad (7)$$

$$\text{отже,} \quad \tilde{m} = m \cdot \phi / (m + \phi). \quad (8)$$

Визначивши  $\tilde{c} = \lambda / \tilde{m}$  і підставивши його у формули Ерланга, отримаємо:

$$p_0 = \left( 1 + \frac{\tilde{c}}{1!} + \dots + \frac{\tilde{c}^k}{k!} + \dots + \frac{\tilde{c}^n}{n!} \right)^{-1}; \quad (9)$$

$$p_k = \frac{\tilde{c}^k}{k!} p_0, \quad (1 \leq k < n); \quad (10)$$

$$p_{\text{отк}} = \tilde{p}_n = \left( \tilde{c}^n / n! \right) \cdot p_0. \quad (11)$$

Показники ефективності СМО визначаються за відомими формулами

$$Q = 1 - \frac{\tilde{c}^n}{n!} p_0; \quad A = \lambda Q = \lambda \left( 1 - \left( \tilde{c}^n / n! \right) \cdot p_0 \right). \quad (12)$$

Для визначення середньої кількості зайнятих каналів потрібно поділити  $A$  на  $\tilde{m}$ , тоді

$$k = (\lambda / \tilde{m}) \cdot \left( 1 - \left( \tilde{c}^n / n! \right) \cdot p_0 \right) = \tilde{c} \left( 1 - \left( \tilde{c}^n / n! \right) \cdot p_0 \right). \quad (13)$$

Багатофазна система масового обслуговування (multistage queueing system, MQS) – система, в якій вимога, що надійшла, проходить послідовно кілька етапів обробки. Головна відмінність MQS від однофазних систем – неоднорідність каналів обслуговування. Для аналізу систем MQS необхідно знати не тільки довжину черги, час очікування обслуговування, навантаження кожного послідовного ланки системи, а й характеристики виходить потоку на кожному етапі обслуговування. У більшості випад-

ків MQS може бути представлена лінійним замкнутим стохастичним ланцюгом, що дозволяє сформулювати завдання її оптимізації в термінах теорії масового обслуговування таким чином.

Нехай система містить  $N$  фаз, причому кожна  $j$ -та фаза складається з  $n_j$  каналів масового обслуговування.

Фази системи з'єднані послідовно, а канали в фазах - паралельно (рис. 1). Відома інтенсивність  $\mu_j$  обслуговування запиту одним каналом в  $j$ -й фазі. Система обслуговує роботу  $m$  джерел вхідних даних, причому з  $k$ -го джерела інформаційні запити щодо аналізу даних надходять з інтенсивністю  $\lambda_k$ .

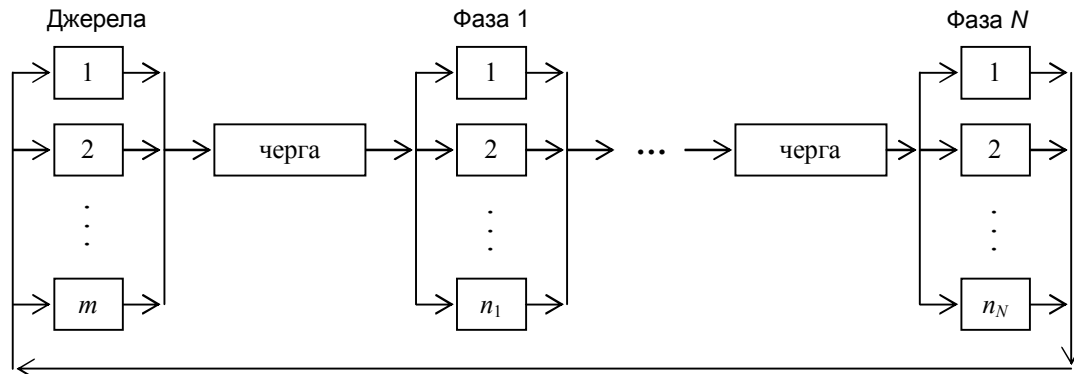


Рис. 1. Схема функціонування багатофазної СОВД

### 3. Оптимізація розподілу ресурсів

Для експлуатації СОВД передбачається виділити обчислювальний ресурс в обсязі  $W$  умовних одиниць, обслуговування одного каналу  $j$ -ої фази дорівнює  $\omega_j$ . Визначимо кількість каналів кожної фази, при якому середній час перебування запиту в системі ( $t_{cep}$ ) буде мінімальним.

Розглянемо роботу системи в стаціонарному режимі, причому всі запити повинні бути обслуговані, а ймовірність надходження заявки із  $i$ -ої фази до  $(i + 1)$ -ої, дорівнює  $P_{i,i+1}$ . Якщо  $t_j^{(cep)}$  – середня тривалість перебування заявки в  $j$ -й фазі, то

$$t_{cep} = \sum_{j=1}^N P_{j-1,j} \cdot t_j^{(cep)}. \quad (14)$$

Для виконання вимоги обов'язкового обслуговування всіх запитів, що надходять до СОВД, необхідно дотримуватися умови ненасиченості, тобто сумарна інтенсивність обслуговування всіх каналів  $j$ -ої фази, повинна бути більше, ніж інтенсивність вхідного потоку заявок цієї фази. Виходячи з цього кількість каналів  $n_j$  має бути не менше фіксованого мінімуму, тобто

$$n_j \geq n_j^{(\phi)}. \quad (15)$$

При виконанні цієї умови для розподілу залишається такий обсяг обчислювального ресурсу:

$$W^{(ocm)} = W - \sum_{j=1}^N \omega_j \cdot n_j^{(\phi)}. \quad (16)$$

Для вирішення завдання методом динамічного програмування необхідно провести поетапну оптимізацію розглянутої СМО, причому етапом оптимізації в даному випадку буде оптимізація фаз системи, починаючи з останньої. Введемо такі позначення:

$x_j$  – обсяг ресурсу, виділеного для обслуговування  $j$ -ої фази (понад обов'язково виділеної суми),

причому в  $\ell$ -му варіанті побудови ( $\ell = \overline{1, L}$ ) системи змінна  $x_j$  приймає фіксоване значення  $x_j^{(\ell)}$ ;

$y_j^{(\ell)}$  – частина ресурсу із  $x_j^{(\ell)}$ , що виділяється безпосередньо на обслуговування фази  $j$ ;

$f_{j,N}(x_j^{(\ell)})$  – мінімальна середня тривалість

перебування заявки у фазах від  $j$ -ої до  $N$ -ої включно при оптимальному розподілі ресурсів між цими фазами в варіанті з номером  $\ell$ .

При оптимізації кожної  $j$ -ої фази будемо розглядати гіпотези про обсяг обчислювального ресурсу  $x_j^{(\ell)}$ , який виділено на обслуговування ланцюга фаз від  $j$ -ої до  $N$ -ої, причому будь-яке припущення повинно відповідати умові

$$0 \leq y_j \leq x_j^{(\ell)} \quad (17)$$

і для кожного значення  $\ell$  має вирішуватися функціональне рівняння

$$f_{j,N}(x_j^{(\ell)}) = \min_{0 \leq y_j \leq x_j^{(\ell)}} \left( P_{j-1,j} \cdot t_j^{(cep)} + f_{j+1,N}(x_j^{(\ell)} - y_j) \right), \quad (18)$$

тобто мінімальне середнє час проходження запиту в мережі фаз (від  $j$  до  $N$ ) дорівнює мінімуму з суми часу перебування заявки в  $j$ -й фазі (з урахуванням виділення на її обслуговування обчислювального ресурсу в обсязі  $y_j$ ) і умовно мінімального часу перебування запиту в фазах від  $(j + 1)$ -ої до  $N$ -ої.

Для кожного етапу оптимізації функціональне рівняння (18) вирішується  $L$  разів, формуючи вектор з  $L$  умовних мінімальних середніх часів і вектор найбільш вигідних значень  $y_j$  при заданих  $x_j^{(\ell)}$ .

На першому кроці процесу оптимізації є вірним  $0 \leq y_N \leq W^{(ocm)}$ , кількість каналів  $N$ -ї фази дорівнює  $n_N = n_N^{(\phi)} + \lceil y_N / \omega_N \rceil$ , а функціональне рівняння (18) приймає вигляд:

$$f_{N,N}(x_N^{(\ell)}) = P_{N-1,N} \cdot t_N^{(cep)}(x_N^{(\ell)}). \quad (19)$$

На другому кроці оптимізації функціональне рівняння (18) з урахуванням (21) є таким:

$$f_{N-1,N}(x_{N-1}^{(\ell)}) = \min_{0 \leq y_{N-1} \leq x_{N-1}^{(\ell)}} \left( P_{N-2,N-1} \cdot t_{N-1}^{(cep)}(x_{N-1}^{(\ell)}) + f_{N,N}(x_{N-1}^{(\ell)} - y_{N-1}) \right). \quad (20)$$

Далі, на кожному k-му кроці значення умовно мінімальної середньої тривалості перебування запиту в фазах від k до N –  $f_{N-k+1,N}(x_{N-k+1}^{(\ell)})$ , знаходиться, виходячи з результатів рішення функціонального рівняння для значення  $f_{N-k+2,N}(x_{N-k+2}^{(\ell)})$ .

Описаний ітераційний процес дозволяє на N-му кроці оптимізації знайти безумовно оптимальний обсяг обчислювального ресурсу  $y_1^{(0)}$  для побудови цієї фази з функціонального рівняння:

$$f_{1,N}(W^{(ocm)}) = \min_{0 \leq y_1 \leq L} \left( P_{0,1} \cdot t_1^{(cep)}(y_1) + f_{2,N}(L - y_1) \right), \quad (21)$$

причому величина  $f_{1,N}(W^{(ocm)})$  є середнім часом перебування запиту в системі, тобто  $t_{cep}$ .

Розгортаючи процес в зворотному напрямку від першої до останньої фази можна розрахувати безумовно оптимальні розміри обчислювального ресурсу, що виділяються на обслуговування кожної фази, використовуючи такий ітераційний процес:

$$f_{j,N}(W^{(ocm)} - \sum_{i=1}^{j-1} y_i^{(0)}) =$$

$$= \min_{0 \leq y_j \leq W - \sum_{i=1}^{j-1} y_i^{(0)}} \left( P_{i-1,i} \cdot t_j^{(cep)}(y_j) + f_{(j+1),N} \left( L - \sum_{i=1}^{j-1} y_i^{(0)} - x_j \right) \right). \quad (22)$$

При цьому необхідна кількість каналів в кожній фазі розраховується як

$$n_j = n_j^{(\phi)} + \left[ y_j^{(0)} / \omega_j \right]. \quad (23)$$

## Висновки

У статті запропонований метод оптимального розподілу ресурсів багатофазної системи обробки великих даних при високоінтенсивному вхідному потоці, котрий враховує як можливість проведення аналізу даних у режимі реального часу, так і особливостей високоінтенсивного потоку вхідних даних із багатьох джерел.

Визначено особливості високоінтенсивних потоків вхідних даних та показана можливість апроксимації його з достатньою точністю гауссівським розподілом. Також визначені особливості функціонування багатофазних систем та запропонована схема функціонування багатофазної системи обробки великих даних.

Базуючись на отриманих результатах розроблена математична модель процесу оптимізації розподілу ресурсів багатофазної системи обробки великих даних при високоінтенсивному вхідному потоці. Для цього використано метод динамічного програмування. В результаті отримана мінімальна кількість каналів в кожній фазі, при якій середній час аналізу даних в системі буде мінімальним.

**Напрямок подальших досліджень** – розробка методу оптимального розподілу ресурсів, що враховує вартісні витрати при обслуговуванні багатофазної системи обробки великих даних.

## СПИСОК ЛІТЕРАТУРИ

1. Кучук Г.А. Метод оценки характеристик АТМ-трафика / Г.А. Кучук // Информационно-керуючі системи на залізничному транспорті, – 2003. – № 6. – С. 44–48.
2. Кучук Г. А., Можаяев О. О., Воробйов О. В. Метод агрегування фрактального трафіка. Радіоелектронні та комп'ютерні системи. 2006. № 6 (18). С. 181–188.
3. 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.
4. 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](https://doi.org/10.1007/978-3-030-00665-5_161)
5. Кучук Г. А. Фрактальный гауссовский шум в трафиковых трассах / Г.А. Кучук // Системи обробки інформації. – 2004. – № 3(31). – С. 91-100.
6. Кучук Г.А. Анализ та моделі самоподібного трафіка / Г.А. Кучук, О.О. Можаяев, Воробйов // Авиационно-космическая техника и технология. – 2006. – Вып. 9 (35). – С. 173-180.
7. Amin Salih M., Potrus M.Y. A Method for Compensation of TCP Throughput Degrading During Movement Of Mobile Node, *ZANCO Journal of Pure and Applied Sciences*. 2015. Vol. 27, No 6. P. 59–68.
8. 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>
9. 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.
10. 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.
11. Kuchuk G., Nechausov S., Kharchenko, V. Two-stage optimization of resource allocation for hybrid cloud data store. *International Conference on Information and Digital Technologies*. 2015. P. 266-271.

12. 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.
13. 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>
14. 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.
15. 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.
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., 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.
18. Sivaram, M., Porkodi, V., Mohammed, A.S., Manikandan V. Detection of Accurate Facial Detection Using Hybrid Deep Convolutional Recurrent Neural Network. *ICTACT Journal on Soft Computing*. 2019. Vol. 09, Issue 02. pp. 1844-1850.
19. Mohammed, A. S. Optimal Forecast Model for Erbil Traffic Road Data. *ZANCO Journal of Pure and Applied Sciences*. 2017. [Vol. 29, No 5. P. 137-145. DOI: https://doi.org/10.21271/ZJPAS.29.5.15](https://doi.org/10.21271/ZJPAS.29.5.15)
20. Amin Salih M., Yuvaraj D., Sivaram M., Porkodi V. Detection And Removal Of Black Hole Attack In Mobile Ad Hoc Networks Using Grp Protocol. *International Journal of Advanced Research in Computer Science*. Vol. 9, No 6. P. 1-6.
21. 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>
22. Свиридов А. С., Коваленко А. А., Кучук Г. А. Метод перерозподілу пропускну 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>
23. Moiseev A. Investigation of High Intensive General Flow / A. Moiseev, A. Nazarov // Proc. of the IV International Conference «Problems of Cybernetics and Informatics» (PCI'2012), September 12-14, 2012. Baku, Azerbaijan. – Baku: ANAS, 2012. – P. 161-163.

**Рецензент:** д-р техн. наук, проф. І. В. Рубан,  
Харківський національний університет радіоелектроніки, м. Харків  
Received (Надійшла) 09.04.2019  
Accepted for publication (Прийнята до друку) 23.05.2019

#### **Распределение ресурсов многофазной системы обработки больших данных при высокоинтенсивном входном потоке**

А. А. Коваленко, Г. А. Кучук, А. С. Ляшенко

Задача оптимального проектирования многофазной системы обработки больших данных (МФСОБД) при высокоинтенсивном входном потоке требует применения современных математических и компьютерных методов и средств. **Предметом исследования** является ресурсы многофазной системы обработки больших данных. **Цель исследования** - разработка метода оптимального распределения ресурсов МФСОБД при высокоинтенсивном входном потоке, который учитывает как возможность проведения анализа данных в режиме реального времени, так и особенностей высокоинтенсивного потока входных данных из многих источников. **Результаты.** Определены особенности высокоинтенсивных потоков входных данных и показана возможность аппроксимации его с достаточной точностью гауссовским распределением. Также определены особенности функционирования многофазных систем и предложена схема функционирования многофазной системы обработки больших данных. Основываясь на полученных результатах, разработана математическая модель процесса оптимизации распределения ресурсов многофазной системы обработки больших данных при высокоинтенсивном входном потоке. Для этого использован метод динамического программирования. **Вывод.** В результате получено минимальное количество каналов в каждой фазе МФСОБД, при которой среднее время анализа данных в системе будет минимальным.

**Ключевые слова:** большие данные, многофазная система массового обслуживания, вычислительный ресурс, высокоинтенсивный поток.

#### **Distribution of resources of the largest phase system of processing big data for a high-intensive input power**

A. Kovalenko, H. Kuchuk, O. Lyashenko

The tasks of optimal design of a multi-phase system for processing big data at high-intensity input flow require the use of modern mathematical and computer methods and tools. **The subject of the study** is the resources of a multi-phase system for processing large data. **The purpose of the research** is to develop a method for optimal allocation of resources of multi-phase system for processing big data at a high-intensity input stream, which takes into account both the possibility of real-time data analysis and the high-intensity flow of input data from many sources. **Results.** The features of high-intensity streams of input data are determined and the possibility of approximating it with a sufficient accuracy by the Gaussian distribution is shown. The peculiarities of the operation of multiphase systems are also determined, and the scheme of operation of the multiphase system of processing large data is proposed. Based on the obtained results a mathematical model of the process of optimization of the resource allocation of a multi-phase system for processing large data at a high-intensity input flow is developed. This is done using the dynamic programming method. **Conclusion.** As a result, we obtain the minimum number of channels in each phase of the multi-phase system for processing big data, in which the average time of data analysis in the system will be minimal.

**Keywords:** large data, multiphase mass service system, computing resource, high intensity flow.