

Вывод

Была разработана система «Data Mining Repository», которая представляет собой сервис-ориентированную архитектуру. Она состоит из следующих уровней: презентационный уровень, сервис-уровень, уровень агентов, база данных.

Было создано несколько агентов из основных функций хранилища. Данная система актуальна и дает возможность взаимодействовать с различными системами с помощью интеллектуальных агентов.

1. *Web Ontology Language (OWL) OWL Web Ontology Language Overview. W3C Recommendation. – February 2004.* 2. *Resource Description Framework (RDF) Model and Syntax Specification. W3C Proposed Recommendation. – January 1999.* 3. *Dubes R.C. and Jain A.K. Algorithms for Clustering Data / Prentice Hall, 1988.* 4. *Blake, C.L. & Merz, C.J., 2000. UCI Repository of machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html].*

УДК 004.75+519.612.2

Д. Федасюк, П. Сердюк, Ю. Семчишин
Національний університет “Львівська політехніка”,
кафедра програмного забезпечення

АДАПТИВНИЙ АЛГОРИТМ ІЄРАРХІЧНО РОЗПОДІЛЕНОГО РОЗВ’ЯЗУВАННЯ СЛАР ВЕЛИКОЇ РОЗМІРНОСТІ

© Федасюк Д., Сердюк П., Семчишин Ю., 2009

У контексті систем розподілення обчислень адаптивністю називають здатність таких систем продовжувати працювати та повністю використовувати обчислювальні ресурси після зміни конфігурації мережі чи швидкодії окремих її вузлів. Предметом цього дослідження є реалізація адаптивного підходу до ієрархічно розподіленого розв’язування систем лінійних алгебраїчних рівнянь великої розмірності. У роботі сформульовано задачу адаптивного підходу, розроблено та досліджено алгоритм та програмну реалізацію адаптивного підходу. Результати серії експериментів підтвердили ефективність адаптивного підходу.

In the context of the distributed computing systems, adaptability is an ability of such systems to continue functioning and fully utilizing computing resources even after changing network configuration or performance of its nodes. The subject of this study is to implement an adaptive approach to hierarchically-distributed solving of high dimensional linear algebraic equations. The problem of an adaptive approach is stated, the algorithm and the software implementation of an adaptive approach are researched and developed in the work below. The experimental results proved the effectiveness of an adaptive approach.

Вступ

Значну кількість задач математичної фізики можна розв’язати за допомогою систем диференційних рівнянь у часткових похідних. Розв’язування таких систем після їхньої лінеаризації числовими методами зводиться до розв’язування систем лінійних алгебраїчних рівнянь (СЛАР).

Внаслідок складної природи багатьох фізичних об'єктів отримані матриці коефіцієнтів можуть мати надто велику розмірність, що ускладнює, а подекуди й унеможливорює розв'язання задачі без залучення потужних обчислювальних ресурсів [1].

Якщо взяти до уваги те, що, з одного боку, організація обчислювальних кластерів на основі існуючої інфраструктури (що, на жаль, часто означає гетерогенність обчислювальних ресурсів та денормалізованість локальних мереж) значно економить кошти, а, з іншого боку, матриці описаних СЛАР часто мають спеціальну (наприклад, блоково-стрічкову) структуру, що робить можливим порівняно легкий їх поділ на слабозв'язні частини, то актуальність систем розподілення обчислень стає очевидною [2, 3].

У роботах [4] та [5] авторами введено поняття неоднорідних систем розподіленої обробки даних, описано ключові особливості таких систем та здійснено їх класифікацію. Також в роботах запропоновано низку критеріїв для кількісної, якісної та комбінованої оцінок ступеня неоднорідності таких систем.

Автор роботи [6] продовжила дослідження попередніх авторів, значно розширивши набір доступних критеріїв для оцінки параметрів неоднорідних систем розподіленої обробки даних. Проте, на жаль, поза межами роботи залишилися питання можливості та доцільності використання пропонуваного критеріїв під час розроблення систем розподілення обчислень, для розв'язування реальних задач.

У роботах [7, 8] колектив авторів детально розглядає питання адаптивності у розподілених середовищах, проте метою цих досліджень є не мінімізація сумарного часу простою апаратних засобів, а забезпечення максимальної надійності та відмовостійкості обчислювального кластера програмними засобами.

Автори роботи [9] намагалися вирішити проблему ефективного поділу завдання на підзавдання з метою максимізації продуктивності гетерогенного розподіленого середовища, проте в роботі не розглядається можливість нелінійної залежності між обчислювальною потужністю вузла та розмірністю даних підзавдання.

Незважаючи на численні переваги розглянутих робіт, проблема реалізації адаптивного підходу до ієрархічно-розподіленого розв'язування СЛАР великої розмірності є актуальною та потребує детальнішого вивчення.

Постановка задачі

Задачею роботи є реалізація адаптивного підходу до ієрархічно-розподіленого розв'язування СЛАР великої розмірності, а також дослідження змін продуктивності системи розподілення обчислень, пов'язаних з реалізацією такого підходу. Розглянемо детальніше поняття адаптивного підходу та вимоги до нього.

Нехай задана СЛАР вигляду:

$$A \cdot \vec{x} = \vec{b}, \quad (1)$$

де $A \in R^{n \times n}$ – дійсна квадратна матриця розмірності $n \times n$ стрічкового вигляду:

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & 0 & 0 \\ A_{21} & A_{22} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & A_{(n-1)(n-1)} & A_{(n-1)n} \\ 0 & 0 & \dots & A_{n(n-1)} & A_{nn} \end{pmatrix}, \quad (2)$$

де A_{ij} – прямокутні матриці довільного розміру.

Нехай також для матриці A задане секціонування у вигляді натурального вектора $\vec{s} \in N^m$ розмірності m ,

$$\vec{s} = (\vec{s}_1 \quad \vec{s}_2 \quad \dots \quad \vec{s}_{m-1} \quad \vec{s}_m), \quad (3)$$

де $\vec{s}_1, \vec{s}_2, \dots, \vec{s}_m$ — розміри відповідних секцій. Таке секціонування за означенням задовольняє одночасно дві умови:

- сума розмірів секцій дорівнює розмірності матриці A :

$$\sum_{i=1}^m \vec{s}_i = n; \quad (4)$$

- будь-який ненульовий елемент матриці A знаходиться всередині квадратної області, утвореної двома послідовними секціями:

$$\forall i, j \in \overline{1, n} : a_{ij} \neq 0 \Rightarrow \exists k \in \overline{2, m} : \left(\sum_{i=1}^{k-1} \vec{s}_i < i \leq \sum_{i=1}^k \vec{s}_i \right) \wedge \left(\sum_{i=1}^{k-1} \vec{s}_i < j \leq \sum_{i=1}^k \vec{s}_i \right). \quad (5)$$

Реалізація адаптивного підходу до розв'язування такої СЛАР полягає у пристосуванні алгоритму поділу завдання на підзавдання до обчислювальної потужності підлеглих виконавців з метою, по-перше, мінімізації загального часу виконання завдання:

$$\vec{t}_{\max} \rightarrow \min, \quad (6)$$

та, по-друге, мінімізації сумарного часу простою системи розподілення обчислень:

$$\sum_{i=1}^k (\vec{t}_{\max} - \vec{t}_i) \rightarrow \min, \quad (7)$$

де k — кількість підлеглих виконавців у вузла, який реалізує адаптивний підхід, $\vec{t} \in R^k$ — дійсний вектор розмірності k , що містить значення часу виконання частин завдання підлеглими виконавцями, а \vec{t}_{\max} — найбільше значення у векторі \vec{t} .

Проте у цьому випадку значення загального часу виконання завдання (6) та сумарного часу простою системи розподілення обчислень (7) є лінійно залежними; а відтак можуть бути мінімізованими водночас завдяки реалізації адаптивного підходу до розподілення обчислень при розв'язуванні СЛАР.

Реалізація адаптивності

Очевидно, що для реалізації адаптивного підходу до ієрархічно розподіленого розв'язування СЛАР великої розмірності необхідним є ефективний метод оцінювання реальної обчислювальної потужності вузлів системи розподілення обчислень.

Сьогодні найпоширенішою одиницею вимірювання обчислювальної потужності апаратних засобів є FLOPS (англ. FLoating point Operations Per Second – «операцій з плаваючою комою на секунду»). Обчислювальною потужністю в 1 FLOPS володіє система, здатна виконувати одну операцію з плаваючою комою протягом однієї секунди. Очевидно, що FLOPS є надто малою одиницею для вимірювання обчислювальної потужності сучасних комп'ютерів, тому найпоширенішою є кратна одиниця – MFLOPS. Одним із популярних способів визначення обчислювальної потужності комп'ютера є розв'язування на ньому СЛАР з насиченою матрицею методом Гаусса без жодних оптимізацій. Зазвичай використовуються СЛАР зі ста рівнянь, проте, якщо є потреба в підвищенні точності результату, то можуть бути використані СЛАР з тисячі рівнянь.

Відомо, що при розв'язуванні СЛАР з насиченою матрицею розміром $n \times n$ елементів методом Гаусса виконується загалом $\frac{1}{2}n(n+1)$ операцій ділення, $\frac{1}{6}(2n^3 + 3n^2 - 5n)$ операцій множення та $\frac{1}{6}(2n^3 + 3n^2 - 5n)$ операцій віднімання [10]. Відтак, загальна кількість операцій з плаваючою комою становить $\frac{1}{2}n(n+1) + \frac{2}{6}(2n^3 + 3n^2 - 5n) = \frac{2}{3}n^3 + \frac{3}{2}n^2 - \frac{7}{6}n$. Тому при

розв'язуванні тестової СЛАР зі ста рівнянь буде всього виконано $\frac{2}{3}100^3 + \frac{3}{2}100^2 - \frac{7}{6}100 = 681550$ операцій з плаваючою комою. Обчислювальна потужність комп'ютера визначається елементарно:

$$p = \frac{681550}{t} \text{ FLOPS}, \quad (8)$$

де t – час розв'язування тестової СЛАР зі ста рівнянь в секундах.

Оскільки обчислювальна складність розв'язування СЛАР з насиченою матрицею становить в загальному випадку $O(n^3)$, то для мінімізації сумарного часу простою системи при поділі кожного завдання на підзавдання, вони повинні бути такими, щоб сума кубів розмірів секцій, які входять до кожного підзавдання була пропорційною обчислювальній потужності підлеглого виконавця, для якого це підзавдання призначене:

$$\forall i \in \overline{1, k} : \sum_{j=1}^{m_i} \vec{s}_{ij}^3 \propto \vec{p}_i, \quad (9)$$

де k – кількість підлеглих виконавців у вузла, який реалізує адаптивний підхід, натуральні вектори $\vec{s}_1, \vec{s}_2, \dots, \vec{s}_k$ з розмірностями m_1, m_2, \dots, m_k відповідно – розміри секцій призначених для кожного підлеглого виконавця, $\vec{p} \in R^k$ – дійсний вектор розмірності k , що містить значення обчислювальних потужностей підлеглих виконавців.

Оскільки вузли системи розподілення обчислень можуть бути також задіяні у виконанні інших задач чи в обслуговуванні користувачів, то значення їх обчислювальної потужності, доступної системі розподілення обчислень, може значно коливатися протягом доби, а, відтак, постає необхідність здійснення повторних тестувань обчислювальної мережі час від часу, а найкраще — перед початком виконання кожного нового завдання.

Саме такий адаптивний підхід було реалізовано в розроблюваній системі розподіленого розв'язування СЛАР з ієрархічною структурою. Систему було розроблено засобами середовища Microsoft Visual Studio 2005 на мові C# для платформи Microsoft .NET Framework 2.0.

Результати експериментів

Тестування реалізованого адаптивного підходу до ієрархічно розподіленого розв'язування СЛАР великої розмірності проводилося в обчислювальній мережі з семи комп'ютерів на базі процесорів AMD Athlon з тактовою частотою 1,81 GHz та 512 MB оперативної пам'яті. Загальна обчислювальна потужність такого кластера, визначена описаним вище способом, становить близько 150 MFLOPS.

Всі тести проводились для кожної з чотирьох можливих комбінацій структури мережі та обчислювальних ресурсів: нормалізованої мережі з гомогенними ресурсами, нормалізованої мережі з гетерогенними ресурсами, денормалізованої мережі з гомогенними ресурсами та денормалізованої мережі з гетерогенними ресурсами. На рис. 1 графічно представлено структури мережі, утворювані під час тестування: нормалізовану (рис. 1а) та денормалізовану (рис. 1б).

Тестування проводилося на трьох різних тестових СЛАР з блоково-стрічковими матрицями з розміром блоку, що дорівнював 32 елементам, подібних до СЛАР, що розв'язуються в реальних задачах теплового проектування електронних пристроїв. Тестові СЛАР склалися з 512, 1024 та 2048 рівнянь.

Під час тестування було виконано ієрархічно розподілене розв'язування кожної з трьох тестових СЛАР: спочатку без застосування адаптивного підходу, а потім з його використанням. З метою підвищення точності вимірів часу в кожному з випадків розв'язування СЛАР виконувалося тричі, а значення вимірів часу усереднювалися. Отримані результати наведено в таблиці.

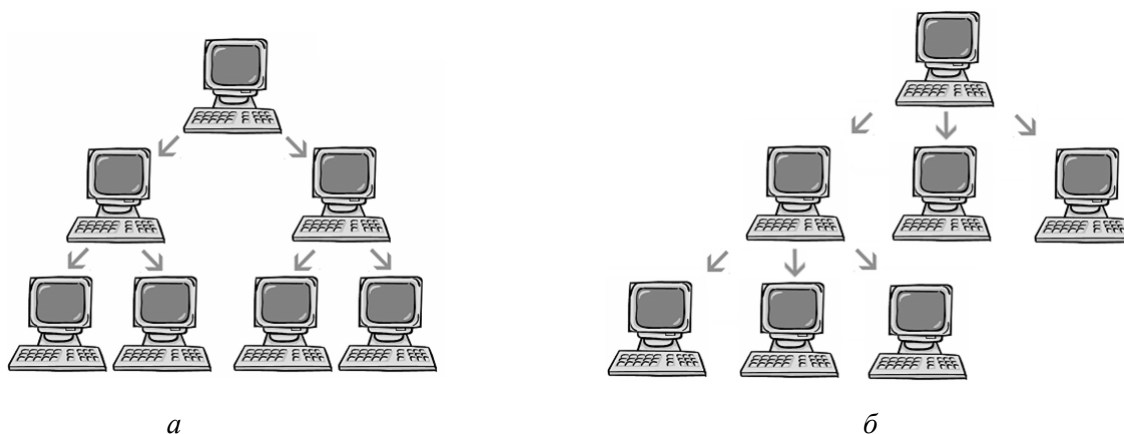


Рис. 1. Структура мережі:
а – нормалізована; б – денормалізована

Результати експериментів

Мережа	Ресурси	Розмір матриці	Без адаптивного підходу				З адаптивним підходом				Приріст продуктивності
			Тест 1	Тест 2	Тест 3	Середній	Тест 1	Тест 2	Тест 3	Середній	
Нормалізована	Гомогенні	512x512	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	0.00%
		1024x1024	24.00	24.00	23.00	23.67	24.00	22.00	22.00	22.67	4.23%
		2048x2048	181.00	169.00	176.00	175.33	158.00	159.00	159.00	158.67	9.51%
	Середній										4.58%
	Гетерогенні	512x512	8.00	7.00	7.00	7.33	4.00	4.00	4.00	4.00	45.45%
		1024x1024	35.00	35.00	35.00	35.00	23.00	26.00	27.00	25.33	27.62%
2048x2048		260.00	263.00	261.00	261.33	180.00	186.00	214.00	193.33	26.02%	
Середній										33.03%	
Денормалізована	Гомогенні	512x512	8.00	7.00	8.00	7.67	6.00	6.00	6.00	6.00	21.74%
		1024x1024	30.00	30.00	30.00	30.00	29.00	29.00	29.00	29.00	3.33%
		2048x2048	230.00	230.00	230.00	230.00	216.00	216.00	216.00	216.00	6.09%
	Середній										10.39%
	Гетерогенні	512x512	18.00	18.00	18.00	18.00	8.00	8.00	8.00	8.00	55.56%
		1024x1024	60.00	60.00	60.00	60.00	57.00	61.00	57.00	58.33	2.78%
2048x2048		449.00	449.00	449.00	449.00	239.00	239.00	239.00	239.00	46.77%	
Середній										35.03%	
Середній										20.76%	

Співвідношення середнього часу ієрархічно розподіленого розв'язування трьох тестових СЛАР звичайним чином та з адаптивним підходом, а також значення приросту продуктивності системи розподілення обчислень внаслідок застосування адаптивного підходу подано в графічному вигляді на рис. 2. На окремих діаграмах зображено дані для нормалізованої мережі з гомогенними ресурсами (рис. 2а), нормалізованої мережі з гетерогенними ресурсами (рис. 2б), денормалізованої мережі з гомогенними ресурсами (рис. 2в) та денормалізованої мережі з гетерогенними ресурсами (рис. 2г).

У жодному з випадків застосування адаптивного підходу до ієрархічно розподіленого розв'язування СЛАР не погіршило існуючої продуктивності системи розподілення обчислень. Натомість використання адаптивного підходу дало змогу скоротити час, необхідний для розв'язування СЛАР, в окремих випадках на 55,56 %.

Використання адаптивного підходу є доцільним, оскільки дає середній приріст продуктивності на рівні 20,76 %. Особливо вигідним він може бути у випадку мереж з денормалізованою структурою (середній приріст продуктивності 22,71 %) та мереж з гетерогенними обчислювальними ресурсами (середній приріст продуктивності 34,03 %).

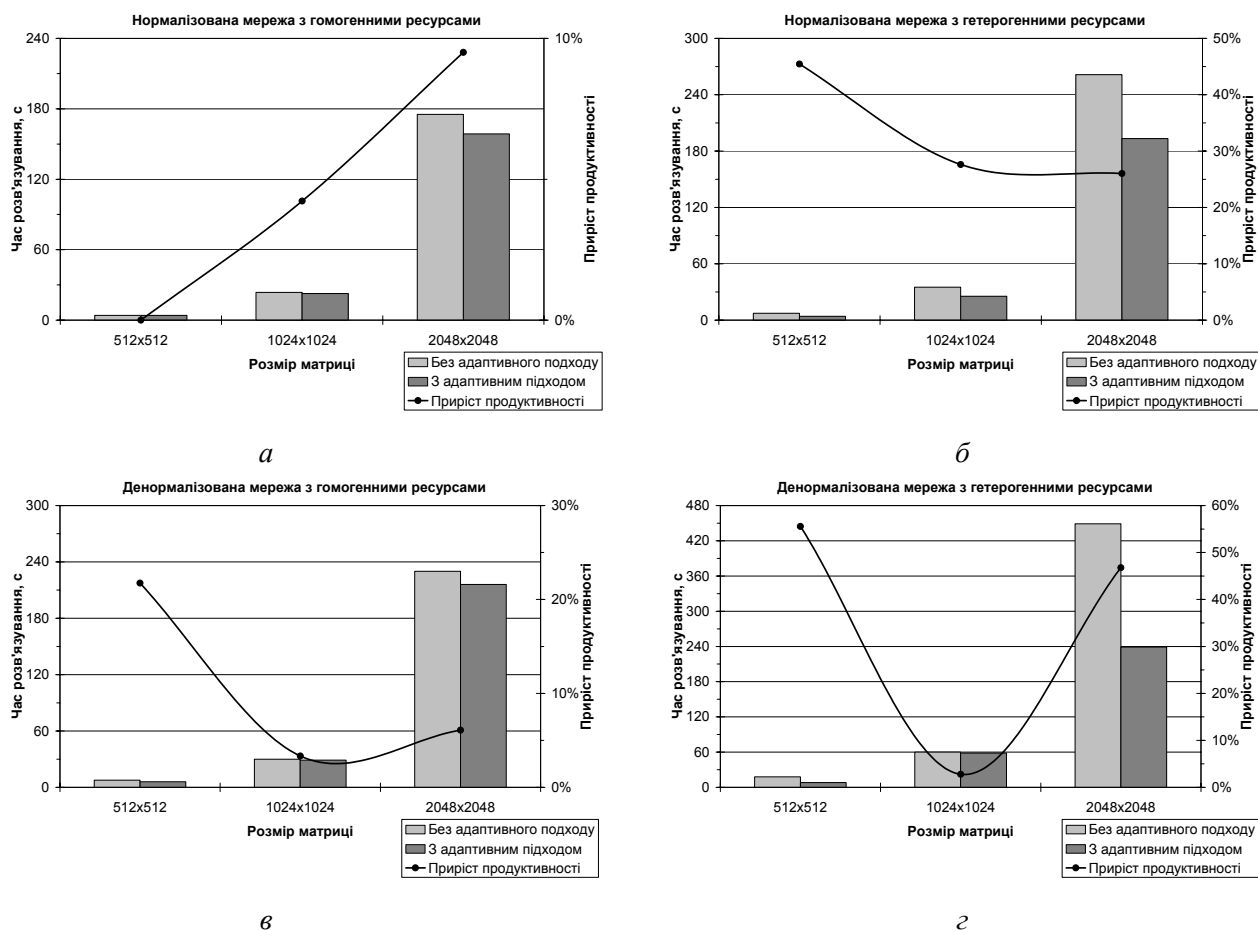


Рис. 2. Результати експериментів: а – нормалізована мережа з гомогенними ресурсами; б – нормалізована мережа з гетерогенними ресурсами; в – денормалізована мережа з гомогенними ресурсами; г – денормалізована мережа з гетерогенними ресурсами

Висновки

У роботі сформульовано задачу реалізації адаптивного підходу до ієрархічно розподіленого розв'язування СЛАР великої розмірності та обґрунтовано її актуальність.

Розроблена та досліджена реалізація адаптивного підходу до ієрархічно розподіленого розв'язування СЛАР великої розмірності дозволяє мінімізувати сумарний час простою системи розподілення обчислень.

Результати серії експериментів, проведеної на тестових СЛАР, підтвердили ефективність реалізації адаптивного підходу до ієрархічно розподіленого розв'язування СЛАР великої розмірності, особливо у випадках мереж з денормалізованою структурою або гетерогенними обчислювальними ресурсами. Так, середній приріст продуктивності після реалізації адаптивного підходу становить 20,76 %, зокрема у випадку мереж з денормалізованою структурою — 22,71 %, мереж з гетерогенними обчислювальними ресурсами — 34,03 %.

1. Баландин М.Ю., Шурина Э.П. Методы решения СЛАУ большой размерности. – Новосибирск: Изд-во НГТУ, 2000. – 70 с. 2. Федасюк Д.В., Сердюк П.В., Семчишин Ю.Б. Математичне та програмне забезпечення для розподіленого розв'язування параметричних задач математичної фізики // Вісник НУ «Львівська політехніка» Комп'ютерні системи проектування: теорія і практика. – Львів: Вид-во НУ «Львівська політехніка». – 2008. – № 626. – С. 94 – 102. 3. Федасюк Д.В., Сердюк П.В., Семчишин Ю.Б. Ієрархічне розподілення розв'язування блоково-стрічкових систем лінійних рівнянь великої розмірності // Вісник НУ «Львівська політехніка» Комп'ютерні науки та інформаційні технології. – Львів: Вид-во НУ «Львівська політехніка». –

2009. – № 638. – С. 261 – 266. 4. Симоненко В.П., Демиденко А.А. Измерение неоднородности распределенных вычислительных систем // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка. – 1999. – № 32. – С. 53 – 58. 5. Симоненко В.П., Осадчий О.Є. Оцінка рівня неоднорідності розподілених обчислювальних систем // Вісник НТУУ «КПІ». – 2000. – С. 35 – 41. 6. Князькова З.В. Аналіз та врахування неоднорідності в системах розподільної обробки даних // Математичні машини і системи. – 2003. – № 2. – С. 136 – 139. 7. F. Giandomenico Di, Bondavalli A., Xu J., Chiaradonna S. Hardware and software fault tolerance: definition and evaluation of adaptive architectures in a distributed computing environment // Proceedings of the International Conference on Safety and Reliability ESREL 1997. – Portugal: Lisbon. – 17 – 20 June 1997. – Volume 1. – PP. 341 – 348. 8. A. Bondavalli, S. Chiaradonna, F. Di Giandomenico, J. Xu. An adaptive approach to achieving hardware and software fault tolerance in a distributed computing environment // Journal of Systems Architecture. – March 2002. – Volume 47, Issue 9. – PP. 763 – 781. 9. B. Hong, V. K. Prasanna. Distributed adaptive task allocation in heterogeneous computing environments to maximize throughput // Proceedings of the 18th International Parallel and Distributed Processing Symposium. – 2004. – PP. 52. 10. R. W. Farebrother. Linear Least Squares Computations. – STATISTICS: Textbooks and Monographs. – Marcel Dekker Inc. – 1988. – PP. 12.