

Запропоновано один із можливих підходів до розробки моделей та сценаріїв процесів пошуку оптимальних рішень з використанням кластерних обчислювальних систем. Розглянуто особливості реалізації зазначеного підходу на прикладі острівної моделі паралельного генетичного алгоритму.

© О.О. Криковлюк, 2013

УДК 519.711,519.81

О.О. КРИКОВЛЮК

ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ СХЕМ РОЗПАРАЛЕЛЮВАННЯ ПРОЦЕСІВ ПОШУКУ ОПТИМАЛЬНИХ РІШЕНЬ НА КЛАСТЕРНИХ АРХІТЕКТУРАХ

Вступ. Однією з основних тенденцій, що намітились останнім часом у світовій та вітчизняній практиці комп'ютерного моделювання, є використання НРС-технологій (High Performance Computing). Такі технології, як правило, реалізуються на платформах суперпотужної обчислювальної техніки, представлених мережними, кластерними або грід архітектурами. Широке впровадження даного класу обчислювальної техніки сприяло розвитку нових методологій імітаційного моделювання, зокрема методології Data Farming. Ця методологія була розроблена в США при реалізації проекту ALBERT і за останні роки набула статусу одного з найбільш ефективних інструментаріїв у галузі імітаційного моделювання складних стохастичних систем. Відмінною рисою даної методології є інтеграція методів імітаційного моделювання, оптимізації та технологій високопродуктивних обчислень [1 – 4].

В Інституті кібернетики імені В.М. Глушкова НАН України накопичено певний досвід дослідження складних стохастичних систем за умов використання мережних архітектур та спеціально розробленої системи NEDISOPT_D. Зазначена система інтегрує можливості методів імітаційного моделювання, оптимізації та технологій розподілених обчислень і забезпечує направлений пошук відповідних оптимальних рішень [5 – 7].

Оскільки оптимізаційна стратегія системи NEDISOPT_D базується на концепціях паралельного генетичного алгоритму, то один із

основних аспектів розробки даної системи був пов'язаний із вибором та реалізацією на мережних архітектурах моделі розпаралелювання процесів пошуку оптимальних рішень. Такою моделлю була обрана модель розпаралелювання на основі декомпозиції даних (на рівні окремих альтернатив, представлених відповідними хромосомами-рішеннями). Для підтримки зазначеної моделі розроблені спеціальні сценарії, які забезпечують проведення відповідних розподілених оптимізаційно-імітаційних експериментів із можливістю вибору заданих типів операторів генетичного алгоритму щодо селекції, схрещування, мутації та форматів представлення пошуку оптимальних рішень.

На рис. 1 показана узагальнена схема процесів пошуку оптимальних рішень із використанням функціональних можливостей основних компонент системи NEDISOPT_D – оптимізатора та імітатора. При цьому всі модулі підтримки оптимізатора розміщуються на головному комп'ютері мережі, а на периферійних комп'ютерах – система розподіленого імітаційного моделювання NEDIS_D [8].

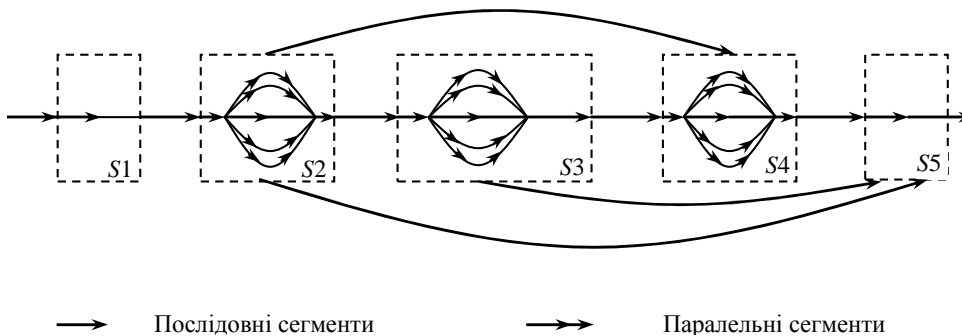


РИС. 1. Узагальнена схема функціонування системи NEDISOPT_D

Допускається розміщення всіх компонент системи NEDISOPT_D на одному процесорі. За таких умов система функціонує у режимі послідовного (однопроцесорного) моделювання.

Функціональні модулі підтримки групи сегментів S1 забезпечують ввід управляючої інформації для сесії моделювання. Зауважимо, що оптимізаційно-імітаційні експерименти на базі NEDISOPT_D реалізуються у форматі сесій моделювання. Модулі групи S2 підтримують формування початкової популяції та оцінювання хромосом останньої. У загальному випадку множина зазначених модулів може бути використана для підтримки оптимізаційної стратегії на основі методу послідовного перебору варіантів. Модулі групи сегментів S3 відображають реалізацію однопопуляційного, паралельного генетичного алгоритму. Модулі сегменту S4 виконують оцінювання статистичної достовірності результатів пошуку оптимальних рішень. Сегменти групи S5 призначені для підтримки процесів видачі окремих сесій моделювання в заданих форматах [9].

1. Постановка задачі. За умов широкого розповсюдження НРС-технологій актуальною стає проблема використання в практиці вітчизняного комп'ютерного моделювання високопродуктивних обчислювальних систем, зокрема таких, як кластерні архітектури. При цьому виникає потреба у виборі або розробці нових моделей для підтримки оптимізаційних стратегій. Так, однопопуляційні моделі паралельного генетичного алгоритму, які були реалізовані в рамках системи NEDISOPT_D і ефективно використовувались при проведенні експериментів на мережних архітектурах, мають бути замінені новими моделями паралельного генетичного алгоритму, більш адаптованими до особливостей організації кластерних архітектур.

Мета роботи полягає у розробці нового підходу до реалізації процесів пошуку оптимальних рішень на кластерних архітектурах за умов використання концепцій методології Data Farming, моделей паралельного багатопопуляційного генетичного алгоритму та досвіду розробки системи оптимізаційно-імітаційного моделювання NEDISOPT_D.

2. Особливості реалізації сценарію пошуку оптимальних рішень на кластерних архітектурах. Першочерговою тут є проблема відображення паралелізму, властивого алгоритмам пошуку оптимальних рішень на основі паралельного генетичного алгоритму на конфігурацію виділених для конкретних сесій моделювання ресурсів кластерних систем.

Зменшення часових витрат на взаємодію паралельних процесів, що базується на обміні повідомленнями, робить доцільним вибір такої оптимізаційної стратегії, яка мінімізує зазначені витрати. Типовим прикладом такого роду оптимізаційної стратегії є острівна модель паралельного генетичного алгоритму. Основна ідея тут полягає у розділенні великих за об'ємом множин хромосом-рішень на однакові за розміром підпопуляції. Процес еволюції кожної підпопуляції реалізується на окремому ядрі кластера (острові) за допомогою одного з можливих сценаріїв послідовного (непаралельного) генетичного алгоритму. Через деяку кількість поколінь підпопуляцій відбувається обмін домінуючими хромосомами рішеннями (міграція хромосом), що в цілому підвищує ефективність оптимізаційних стратегій на основі паралельного генетичного алгоритму.

Особливості реалізації сценарію підтримки обраної стратегії на основі острівної моделі можуть бути сформульовані наступним чином:

- використання концепцій популяція та підпопуляція хромосом-рішень;
- використання концепцій методології Data Farming щодо декомпозиції даних, представлених багатовимірним простором параметрів – характеристик оцінюваних альтернатив, та реалізації імітаційних експериментів на основі мульти-сценарних обчислень;
- адаптація структури та об'єму підпопуляцій до конфігурації реальної апаратури (виділених на дану сесію моделювання ресурсів кластера);
- стратегія ініціалізації початкових підпопуляцій та розпаралелювання процесів пошуку оптимальних рішень базуються на так званій моделі збалансованої декомпозиції даних;

- використання концепції міграції для реалізації обміну доміантними хромосомами між островами;
- реалізація прогонів імітаційних моделей у рамках оцінювання кожного покоління підпопуляції виконується в режимі послідовного (однопроцесорного) моделювання на відповідних ядрах кластера;
- використання бібліотеки MPI для підтримки схем розпаралелювання процесів пошуку оптимальних рішень;
- сценарії пошуку оптимальних рішень мають забезпечити рівномірне і, за можливістю, максимальне завантаження процесорів кластера;
- використання ядра з номером 0 для розміщення головного процесора P_0 сценарію пошуку оптимальних рішень на кластерних архітектурах;
- використання СКІТЗ як експериментального полігону для оцінювання розробленого підходу [10].

Узагальнена схема підтримки запропонованої оптимізаційної стратегії показана на рис. 2. Тут прийняті наступні позначення: Sij – модулі, які підтримують операції, аналогічні операціям однопопуляційного генетичного алгоритму, реалізованого в системі NEDISOPT_D. Блоки з позначенням $Smij$ визначають модулі, призначені для модифікації сценарію однопопуляційного генетичного алгоритму з метою перетворення його на багатопопуляційний паралельний генетичний алгоритм для підтримки нової оптимізаційної стратегії; N – кількість ядер кластера, виділених на дану сесію моделювання.

Керуюча інформація для сесій моделювання має включати множину фрагментів, що визначається кількістю виділених на дану сесію моделювання ядер кластера. До складу вказаних фрагментів обов'язково включається інформація, яка визначає відповідний сценарій генетичного паралельного алгоритму, що реалізується на кожному окремому ядрі кластерного вузла. Такі сценарії будуть відрізнятися схемами селекції, операторами схрещування та мутації. До зазначених фрагментів також має включатися інформація про параметри вибраної моделі міграції.

Після формування та оцінювання поточного покоління підпопуляції хромосом аналізуються умови виконання міграції. За виконання цих умов реалізується однонаправлений обмін доміантними хромосомами за кільцевим маршрутом між ядрами кластера. При цьому хромосоми-імігранти витісняють з даної підпопуляції хромосоми з більш низькими показниками якості. В протилежному випадку хромосоми іммігранти не включаються до складу підпопуляції на даному ядрі. Після виконання міграції виконується перевірка умов завершення еволюції. У разі невиконання умов процес еволюції продовжується.

По завершенню еволюції на всіх ядрах кластера головному процесу P_0 пересилається група доміантних хромосом кінцевого покоління кожної підпопуляції. Спеціально розроблений модуль $Sm70$ (характеристичний порівнювач) аналізує множину таких хромосом з метою вибору найбільш оптимальної.

Для оцінки показників статистичної достовірності оптимальної альтернативи (модуль $S40$) відбувається багатократний прогін імітаційної моделі з використанням характеристик такої альтернативи як факторів останньої.

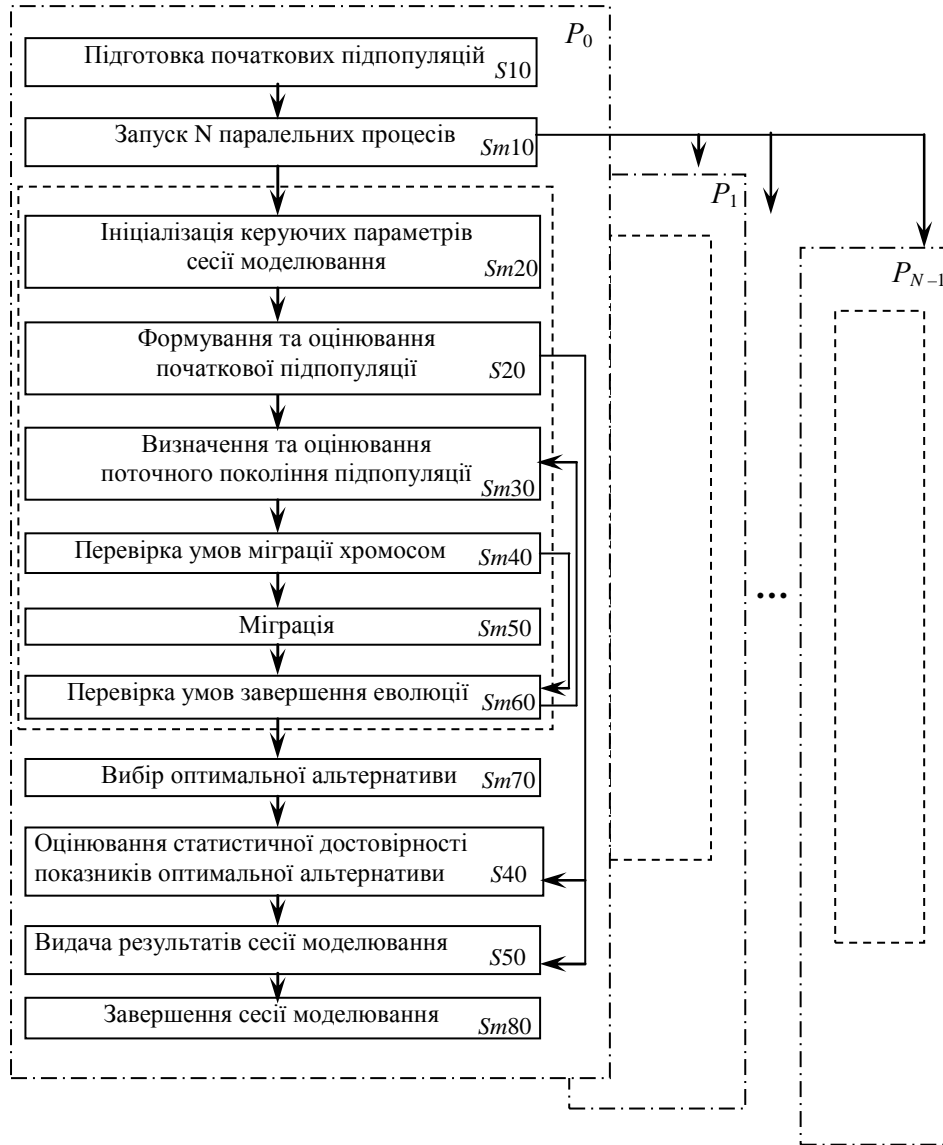


РИС. 2. Узагальнена схема розпаралелювання процесів пошуку оптимальних рішень на кластерних архітектурах

Модуль S_{50} забезпечує видачу кінцевих результатів сесії моделювання у форматах, визначених управляючою інформацією до сесії моделювання.

Модуль Sm_{80} виконує операції по завершенню паралельних процесів та сесії моделювання в цілому.

При використанні засобів бібліотеки MPI щодо розпаралелювання процесів пошуку оптимальних рішень на всі, виділені для даної сесії моделювання, ядра кластера симетрично завантажуються всі модулі зазначеного сценарію.

Зауважимо, що всі модулі сценарію виконуються тільки на ядрі з номером 0. На інших ядрах виконуються множина модулів, обведена дефіс-контуром. Така схема виконання регулюється управляючою інформацією до сесії моделювання.

Висновки. Розроблено один із можливих підходів до підвищення ефективності оптимальних рішень за умов використання паралельного генетичного алгоритму як оптимізаційної стратегії та можливостей кластерних обчислювальних систем. Зазначені засоби можуть бути використані при дослідженні різного роду складних стохастичних систем на відповідних кластерних архітектурах.

Одним із перспективних напрямків розвитку виконаних досліджень є розширювання класу моделей оптимізаційних стратегій та використання можливостей графічних прискорювачів (GPU) у процесах направлено пошуку оптимальних рішень використанні кластерних архітектур, зокрема SKIT-4.

Е.А. Криковлюк

ОСОБЕННОСТИ РЕАЛИЗАЦИИ СХЕМ РАСПАРАЛЛЕЛИВАНИЯ ПРОЦЕССОВ ПОИСКА ОПТИМАЛЬНЫХ РЕШЕНИЙ НА КЛАСТЕРНЫХ АРХИТЕКТУРАХ

Предложен один из возможных подходов к разработке моделей и сценариев процессов поиска оптимальных решений с использованием кластерных вычислительных систем. Рассмотрены особенности реализации указанного подхода на примере островной модели параллельного генетического алгоритма.

О.О. Krikovluik

FEATURES OF PARALLEL PROCESS REALISATION SCHEMES OF OPTIMAL DECISIONS SEARCHING ON CLUSTER ARCHITECTURES

An approach to a design of models and scenarios of optimal decisions searching processes with the use of cluster computing systems is proposed. The applications of this approach on the example of island model of parallel genetic algorithm are considered.

1. *Horne G.E., Meyer T.E.* Data Farming: Discovering Surprise // Proc. of the Winter Simulation Conf. – 2005. – P. 1082–1087.
2. *Horne G.E., Schwierz K.-P.* Data Farming around the world overview // Proc. of the Winter Simulation Conf. – 2008. – P. 1442–1447.
3. *Choo C.S., Ng E.C., Ang Dave, Chua C.L.* Data Farming in Singapore: a brief history // Proceedings of the 2008 Winter Simulation Conf. – 2008. – P. 1448–1455.

4. SEED Center for Data Farming: <http://harvest.nps.edu/>
5. M. Fu: Optimization for Simulation: Theory and Practice. INFORMS // J. on Computing. – 2002. – 14(3). – P. 192 – 215.
6. April J., Glover F., Kelly J.P., Laguna M. Practical introduction to simulation optimization. Proc. of the 2003 Winter Simul. Conf. – 2003. – P. 71 – 78.
7. Галаган Т.Н., Пепеляев В.А., Сахнюк М.А. Особенности реализации многослойного сценария распределенного поиска оптимальных решений // Проблемы програмування. – 2008. – № 2–3. – С. 636 – 640.
8. Пепеляев В.А., Сахнюк М.А., Черный Ю.М. Параллельная реализация процессов направленного поиска оптимальных решений // Проблемы програмування. – 2010. – № 2–3. – С. 572 – 576.
9. Гусев В.В., Галаган Т.Н., Яценко Н.М. Технологическая система распределенного имитационного моделирования NEDIS_D // Тр. Первой научно-практической конф. «Математичне та імітаційне моделювання систем – МОДС 2006». – 2006. – С. 139 – 143.
10. Суперкомп'ютер ІК НАН України <http://icybcluster.org.ua/>

Одержано 04.02.2013

Про автора:

Криковлюк Олена Олександрівна,
аспірантка Інституту кібернетики імені В.М. Глушкова НАН України.