

Адаптація схем розпаралелювання реальних процесів до архітектури кластерних систем

Криковлюк О.О., Пепеляєв В.А.

*Інститут кібернетики ім. В.М. Глушкова НАН України,
03680, ГСП, Київ-187, проспект Академіка Глушкова, 40*

alenska1331@gmail.com, emk160ik@gmail.com

Анотація. *Пропонується один із можливих підходів до реалізації процесів направлено пошуку оптимальних рішень на кластерних архітектурах. Визначено основні принципи формування такого підходу.*

Ключові слова

Оптимізаційно-імітаційна інтеграція, розподілені обчислення, High Performance Computing, схема розпаралелювання, методологія Data Farming, кластерні системи.

1 Вступ

Однією з основних тенденцій в галузі сучасного імітаційного моделювання є перехід від технологій розподілених обчислень до технологій високопродуктивних обчислень, що, як правило, реалізуються на суперпотужних платформах обчислювальної техніки [1,2].

В Інституті кібернетики ім. В.М. Глушкова НАН України накопичено певний досвід реалізації процесів пошуку оптимальних рішень на мережних архітектурах із використанням можливостей технологій розподілених обчислень. Для підтримки таких процесів розроблено спеціальні сценарії та відповідне програмне забезпечення в рамках системи NEDISOPT_D [3,4].

Оскільки розподілені обчислення базуються на паралельних процесах, то початковим етапом створення даної системи була розробка відповідних схем розпаралелювання (на концептуально-алгоритмічному рівні). При цьому в основу вказаної схеми покладена дворівнева декомпозиція: розподіл функцій між основними компонентами системи NEDISOPT_D (оптимізатором та імітатором); декомпозиція на основі збалансованих даних (на рівні окремих хромосом).

Зауважимо, що всі модулі оптимізатора підтримують прийнятну оптимізаційну стратегію, зокрема стратегію на базі однопопуляційного паралельного генетичного алгоритму, а модулі імітатора підтримують процеси прогону імітаційних моделей. Така схема розпаралелювання може бути адекватним чином адаптована до структури відповідної мережі. При цьому на головному комп'ютері мережі розміщується оптимізатор, а на відповідних периферійних комп'ютерах мережі – імітатор.

В такому випадку може бути запущено $m-1$ паралельних асинхронних процесів, де m – кількість комп'ютерів мережі. В цілому за один крок пошуку оптимальних рішень запускається число паралельних процесів $PopulSize$, де $PopulSize$ – розмір популяції, на основі еволюції якої визначається відповідне оптимальне рішення (як правило, $PopulSize > m$).

Застосування технологій високопродуктивних обчислень в задачах пошуку оптимальних рішень потребує використання нових типів обчислювальних платформ, зокрема кластерних систем [5,6]. В свою чергу, останнє вимагає розробки нових схем розпаралелювання (в першу чергу, на концептуально-алгоритмічному рівні). Саме такі дослідження в даний момент виконуються в Інституті кібернетики при розробці програмних засобів підтримки процесів пошуку оптимальних рішень на кластерних архітектурах.

2 Розробка нової схеми процесів пошуку оптимальних рішень

В основу вказаної схеми покладена комбінація трьох типів декомпозиції: тривіальна, функціональна та декомпозиція за даними.

Розробка схеми базуються на наступних принципах:

- інтеграція методів імітаційного моделювання, оптимізації та технологій високопродуктивних обчислень;
- використання концепцій популяції та підпопуляції хромосом-рішень;
- використання оптимізаційної стратегії, що базується на багатопопуляційному паралельному генетичному алгоритмі (острівна модель генетичного алгоритму). Для вказаної моделі характерним є обмін хромосомами в процесі еволюції під популяції;

- трьохрівневе розпаралелювання процесів пошуку оптимальних рішень (на рівні різних версій сценарію відповідної оптимізаційної стратегії (функціональне розпаралелювання), на рівні підпопуляцій (збалансована декомпозиція даних) та на рівні окремих хромосом (тривіальна декомпозиція)). Для підтримки першого рівня розпаралелювання буде використана така концепція методології Data Farming як *мультисценарні обчислення*. На основі управляючої інформації до імітаційних експериментів, що реалізуються в форматах сесій моделювання, формуються версії острівного сценарію генетичного алгоритму у відповідності з рекомендаціями дослідника (користувача). Такий підхід забезпечує динамічну реконфігурацію сценарію оптимізатора та реалізацію на кожному ядрі кластера процесу пошуку оптимальних рішень з власними параметрами стосовно операцій відбору кроссовера та мутації. Весь простір пошуку оптимальних рішень представлений відповідною популяцією хромосом-рішень. Перед запуском паралельних процесів на виділених ядрах кластера відбувається поділ популяції на однакові за об'ємом підпопуляції, які виступають у якості початкових підпопуляцій для відповідних ядер;

- управляюча інформація до експериментів, що реалізуються в форматі сесій моделювання, має фрагментарну структуру. Кількість таких фрагментів відповідає виділенню на дану сесію ресурсам кластера (кількості ядер);

- завантаження на кожне ядро виділених ресурсів кластера сценаріїв підтримки оптимізатора та імітатора. Задана для даного ядра версія оптимізаційної стратегії буде визначатись відповідним фрагментом управляючої інформації до сеансу моделювання;

- оцінювання альтернатив на кожному ядрі кластера відбувається в режимі однопроцесорного (послідовного) моделювання;

- базування системи зберігання даних на прийнятій в рамках кластера СКІТЗ глобальній файлової системі Lustre [7].

3 Висновки

Оскільки множина даних, що використовується в процесах пошуку оптимальних рішень на кластерних архітектурах, має досить складну структурну організацію (управляюча інформація для сесій моделювання, управляюча та числова інформація для оптимізатора та імітатора), то першочерговим питанням подальших досліджень є визначення та формування інформаційних моделей для вказаних даних, їх розміщення на відповідних ресурсах кластера та використання.

Література

- [1] G. E. Home, T.E. Meyer: Data Farming: Discovering Surprise. *Proc. of the Winter Simulation Conf. 2005*: 1082 – 1087, 2005.
- [2] SEED Center for Data Farming: <http://harvest.nps.edu/>
- [3] Т.Н. Галаган, В.А. Пепеляев, М.А. Сахнюк: Особенности реализации многослойного сценария распределенного поиска оптимальных решений. *Проблеми програмування*, 2-3: 636 – 640, 2008.
- [4] В.А. Пепеляев, М.А. Сахнюк, Ю.М. Чёрный: Параллельная реализация процессов направленного поиска оптимальных решений. *Проблеми програмування*, 2-3: 572 – 576, 2010.
- [5] А.Е. Дорошенко Математические модели и методы организации высокопроизводительных параллельных вычислений. - Киев: «Наук. думка», 2000.-177С.
- [6] Системі паралельної обробки /под редакцією Д. Ивенса/ Москва: Мир, – 1985 – 412 С.
- [7] Суперкомп'ютер ІК НАН України <http://icybcluster.org.ua/>