

К.т.н., доцент Замятін Д.С., магістрант Буткевич Т.В.

Національний технічний університет України
«Київський політехнічний інститут»

МЕТОД БАЛАНСУВАННЯ НАВАНТАЖЕННЯ В GRID-СИСТЕМАХ НА БАЗІ МОБІЛЬНИХ ПРИСТРОЇВ

Abstract

*Denys S. Zamyatin, assoc. prof., PhD; Taras Butkevych, student
The method of load balancing in grid-systems based on mobile devices*

Mobile grid systems which are based on voluntary participation are heterogeneous. Every node has different computation capabilities limited battery life poor performance and unreliable internet connection. Thus the load balancing is more difficult in mobile grid environment and has many issues. There are many techniques for static and dynamic load balancing. In this paper the most appropriate load distribution ways are analyzed and the dynamic load balancing method is proposed. The method bases on connection history, which is forming during period while node is registered in mobile grid system, devices availability type and job classification. The experiment shows that load balancing based on proposed method reduce an amount of tasks that are assigned to the unavailable nodes.

Вступ

Найбільш актуальні завдання при розробці мобільної grid-системи це планування задач, та балансування навантаження [1]. Наприклад, під час виконання завдання, що потребує ресурсів мережі, з'єднання з мережею може бути порушене. У цьому випадку обчислення повинні бути зупинені до моменту відновлення з'єднання, що призводить до зниження продуктивності роботи. Без правильного балансування навантаження всі завдання будуть використовувати стабільні ресурси. Це призводить до “дискримінації” мобільних пристроїв, що мають меншу продуктивність, але мають перевагу у кількості. В результаті не тільки знижується кількість ресурсів, що використовується, але й знижується продуктивність системи за рахунок неправильного розподілу навантаження.

Мобільні пристрої в grid-системах можуть грати роль як споживача, так і провайдера. Як споживач вони використовують обчислювальні можливості grid-системи, а як провайдер — безпосередньо беруть участь в обробці запитів. У порівнянні з фізично фіксованими grid-ресурсами, такими як персональні комп'ютери, мобільні пристрої, як правило, забезпечуються порівняно низькими показниками, з точки зору

можливостей процесора, обсягу оперативної пам'яті. Також, мобільні пристрої використовують бездротовий спосіб підключення до мережі, що зменшує надійність з'єднання в порівнянні з дротовими мережами. Однією з найголовніших проблем тут є обмежений термін служби акумулятора. Тим не менш, крім недоліків існують і переваги: можливість визначення точного місця розташування вузла та надзвичайно стрімкий ріст кількості цих пристроїв.

Постановка задачі

На відміну від традиційних grid-систем, мобільні grid-системи мають особливість. Розташування їх ресурсів залежить від місця розташування їх власника і постійно змінюється. Під час руху, вузол може втрачати з'єднання з мережею. Задача полягає в розробці методу балансування навантаження в grid-системах на базі мобільних пристроїв, який враховує постійну зміну розташування ресурсів системи та їх нестабільне з'єднання з мережею.

Термінологія

Grid — це тип паралельно розподіленої системи, яка дозволяє розділення, вибір та накопичення географічно розподілених «автономних» ресурсів в реальному часі в залежності від їх придатності, можливостей, ціни та вимог до якості обслуговування користувачів.

Споживач — користувач або агент, який використовує доступний ресурс, агента або служби.

Хост — обчислювальний елемент як фізична одиниця. Цей елемент характеризує фізичну конфігурацію обчислювального вузла, включаючи процесори, програмне забезпечення, елементи пам'яті тощо.

Огляд існуючих рішень

Існує три основні параметри, які визначають стратегію балансування навантаження, що буде використовуватися:

- хто приймає рішення про балансування навантаження;
- яка інформація використовується для прийняття рішення про балансування навантаження;
- де приймається рішення про проведення балансування.

У багатьох дослідженнях пропонуються різні підходи до вирішення цієї задачі [2 - 5]. На жаль вони не адаптовані для використання на мобільних пристроях. Концептуально методи балансування навантаження можна

розділити на два типи – статичні та динамічні. В статичних розподіл навантаження здійснюється на етапі проектування розподіленої програми, в динамічних під час аналізу завантаженості вузла. Далі не будемо розглядати статичні алгоритми оскільки мобільне середовище передбачає швидку зміну станів та статусу доступності вузлів. Що стосується динамічних, в них розрізняють різні політики балансування. Розглянемо основні.

Централізована політика використовує одного агента, який є відповідальним за прийняття рішення про включення певного вузла в обчислення тієї чи іншої задачі. Ця політика не придатна для випадку, що розглядається, оскільки завантаженість вузла залежить також і від користувача пристрою, а агент не володіє цією інформацією. Крім того, суттєво збільшується обсяг трафіку в мережі через використання ширококомовних запитів. Інша політика — децентралізований перерозподіл за ініціативою отримувача. В цьому випадку перерозподіл навантаження повністю залежить від обчислювальних вузлів і має суттєвий недолік, а саме: вузли мобільної grid-системи можуть бути вимкненими одразу після отримання задачі і отримувач не буде здатний передати необхідну інформацію.

Також існує політика децентралізованого перерозподілу за ініціативою відправника, коли відправник задачі вже відомий і необхідно визначити необхідні обчислювальні вузли. При використанні такої політики, для прийняття рішення про необхідність розпочати процедуру балансування використовується локальна інформація про рівень завантаження лише «найближчих» вузлів. Це штучно обмежує множину вузлів-кандидатів для передавання навантаження з вузла.

Серед розглянутих динамічних алгоритмів найбільш придатним для використання в мобільному grid-середовищі є алгоритм балансування навантаження на основі агентів. Головний його недолік — неможливість передбачення вимкнення деякого вузла обчислювальної мережі в певний період часу. Тому приходимо до висновку, що традиційні алгоритми балансування не ефективні для їх використання в мобільній grid-системі.

Опис методу

Розподілимо задачі, що надходять до менеджера задач на дві групи. До першої групи віднесемо задачі, що потребують потужних обчислювальних ресурсів GroupP, до другої — стабільного з'єднання з мережею для передачі великої кількості даних GroupC.

Кожен ресурс зареєстрований у мобільній grid-системі отримує ідентифікатор групи. Цей ідентифікатор обирається з таблиці 1.

Таблиця 1.

Групи ресурсів мобільної grid-системи

<i>P</i> \ <i>C</i>	Високий	Середній	Низький
Високий	ВВ	ВС	ВН
Середній	СВ	СС	СН
Низький	НВ	НС	НН
Тимчасово низький	ТВ	ТС	ТН

P — рівень обчислювальної потужності пристрою, *C* — рівень стабільності з'єднання ресурса з мережею. Маючи класифікацію ресурсів по рівню їх обчислювальної потужності та стабільності з'єднання з іншими ресурсами мережі, формуємо для груп Group*P* та Group*C*, список ресурсів, які доцільно використовувати для кожної з них. Найбільш універсальні ресурси відносяться до груп ВВ, СС, НН та можуть використовуватися для будь-якого типу задач. Групи ресурсів СВ, НВ краще використовувати для задач які відносяться до Group*P*, а ВН, СН відповідно до задач з групи Group*C*.

Рівень обчислювальної потужності залежить від вагових коефіцієнтів продуктивності процесору W_i^{mem} , оперативної пам'яті W_i^{proc} та заряду батареї. Параметри W_i^{mem} , W_i^{proc} визначаються як співвідношення характеристики продуктивності *i*-го пристрою і продуктивності пристрою що обраний як стандарт.

Рівень стабільності з'єднання ресурсу з мережею C_i змінюється протягом всього часу існування вузлу у мобільній grid-системі, та визначається наступним чином:

$$T_{wi} = \frac{T_i}{T_{all}}$$

$$C_i = \frac{T_{wi}}{D_i}$$

де T_i — час, протягом якого ресурс використовувався для обчислення задач мобільної grid-системи, T_{all} — час, протягом якого ресурс доступний для задач grid-системи, D_i — кількість втрат з'єднання з мережею, протягом якого ресурс використовувався для обчислення, grid-системою.

До груп ТВ, ТС, ТН відносяться ресурси, що активно використовуються користувачем і в яких, рівень завантаженості процесору та оперативної пам'яті перевищує 80%. Розподілення задач до таких ресурсів не відбувається.

Історія переміщення ресурсів та статус їх з'єднання з мережею постійно фіксується. Протягом певного часу формуються зони найбільшої ймовірності відмови з'єднання. Ресурси, що перебувають у таких зонах

помічаються міткою — “небажаний для використання”. Такі ресурси використовуються мобільною grid-системою, якщо в системі більше не залишилося вільних ресурсів.

Можна виділити основні етапи балансування навантаження.

1. Розподілення задач менеджером задач на дві основні групи (GroupP та GroupC).
2. Визначення вагового коефіцієнту заряду батареї кожного пристрою, та прийняття рішення про додавання ресурсу до списку доступних ресурсів.
3. Визначення для кожного ресурсу зі списку доступних ресурсів його групи.
4. Додавання мітки “небажаний для використання” до ресурсів, що знаходяться у зоні великої ймовірності втрати з’єднання.
5. Розподілення задач відповідно до груп ресурсів.

Запропонований метод динамічного балансування навантаження дозволяє розподіляти запити, що надходять до системи з урахуванням продуктивності обчислювальних елементів, що динамічно змінюються. Метод враховує велику кількість і високу ймовірність відмови вузлів системи.

Висновки

Таким чином, запропонований метод динамічного балансування навантаження дозволяє розподіляти запити, що надходять до системи з урахуванням продуктивності обчислювальних елементів, що динамічно змінюються. Розроблений метод враховує більшість обмежень, які притаманні мобільним пристроям та дозволяє збільшити ефективність обчислень. У подальшому планується розробка та тестування алгоритму, що базується на цьому методі.

Література

1. Волк М.А., Филимончук Т.В., Гридель Р.Н. Методы распределения ресурсов для Grid- систем // Збірник наукових праць Харківського університету Повітряних Сил. – 2009. Вип. 1(19). – С. 100 – 104.
2. K. Abini. Decision Making for Load Balancing in a Distributed System, Proceedings of the 36th Midwest Symposium Circuits and Systems, 1993, pp. 500–502.
3. C.W. Cheong, V. Ramachandran. Genetic Based Web Cluster Dynamic Load Balancing in Fuzzy Environment, Proceedings of the Fourth

International Conference on High Performance Computing in the Asia-Pacific Region, Beijing, China, Vol. 2, 2000, pp. 714–719.

4. *P. Chulhye, J.G. Kuhl* A fuzzy-based distributed load balancing algorithm for large distributed systems, Proceedings of the Second International Symposium on Autonomous Decentralized Systems, April 1995, pp. 266–273.

5. *G.Cybenko*. Dynamic load balancing for distributed memory multiprocessors, J. Parallel Distributed Computng, 7 (1989) 279–301.