

CROWDSOURCING АБО ЯК ДОЛУЧИТИСЯ ДО VOLUNTEER COMPUTING

Цирульник Сергій¹, Непийвода Максим²

¹Вінницький національний технічний університет

²Вінницький технічний коледж

Анотація

У статті розглядається спосіб розв'язку надскладних задач із залученням територіально віддалених один від одного обчислювальних ресурсів. Основна увага приділяється явищу Crowdsourcing, як способу залученню вільних обчислювальних ресурсів користувача (Volunteer Computing) до участі у добровільних обчисленнях наукових проектів у системі BOINC.

Abstract

In the article the way of solving highly complex problems involving geographically distant from each other computing resources. The focus is on the phenomenon of Crowdsourcing, as a way of attracting free computing power user to participate in volunteer computing projects in the scientific system of BOINC.

Вступ

З розвитком мережі Інтернет з'явився новий підхід до побудови GRID-систем, що дозволяє об'єднати обчислювальні ресурси для вирішення завдань обробки великих та надвеликих обсягів даних. Цей підхід полягає в організації системи розподілених обчислень з обчислювальних вузлів відносно невисокої продуктивності. У більшості випадків такі системи побудовані на використанні вільних ресурсів комп'ютерів приватних осіб та організацій, які добровільно приєднуються до таких GRID, щоб допомогти науковим дослідженням (Volunteer Computing).

Актуальність

Для обробки великих масивів даних використовуються суперкомп'ютери або обчислювальні кластери. При цьому для аналізу надвеликих наборів даних та/або обчислювальних алгоритмів використовуються спеціальним чином організовані системи з географічно віддалених кластерів – GRID-кластерів [1, 2].

Краудсорсинг (Crowdsourcing) широко застосовується компаніями, науковими установами, громадськими та політичними організаціями. Це поширена форма організації on-line проектів. Наука – один із самих природних напрямків для використання праці краудсорсерів. Завдання, що встають перед людством, занадто великі для того, щоб їх можна було вирішити силами однієї або навіть декількох наукових груп. Найчастіше до технології розподілених обчислень (Distributed Computing) дослідники звертаються тоді, коли наукові проекти вимагають величезної кількості обчислень, які недоцільно та дорого виконувати з допомогою суперкомп'ютерів або кластерних обчислювальних мереж. Перспективним підходом до обробки складних завдань, підвищення якості результатів та зменшення часу розрахунків є застосування систем паралельної розподіленої обробки.

Основна частина

GRID – це об'єднання декількох комп'ютерів для вирішення єдиної обчислювальної складної задачі, що розбивається на окремі підзадачі. Кожен комп'ютер вирішує кілька підзадач, а результати окремих обчислень об'єднуються. Основна перевага GRID в тому, що вона може складатися з комп'ютерів, що знаходяться один від одного на великій відстані та мають різні характеристики (як апаратні, так і програмні). Завдання

об'єднання цих різномірних комп'ютерів виконує проміжне програмне забезпечення, яке віртуально пов'язує всі комп'ютери через Інтернет в єдиний суперкомп'ютер.

Volunteer Computing – це форма реалізації GRID-обчислень, специфіка якої полягає у використанні часу простою звичайних користувачів комп'ютерів по всьому світу.

Надати ресурс власного комп'ютера можна за допомогою мережі інтернет та технології розподілених обчислень (GRID-обчислень). Існує кілька систем для добровільних обчислень. Система повинна вміти розбивати одну велику задачу на безліч маленьких підзадач, розподіляти ці підзадачі з обчислювальних вузлів, приймати результати обчислень та об'єднувати їх в єдине ціле. Для цього були створені різні програмні «прошарки» між керуючим сервером та обчислювальними вузлами. Це BOINC, XtremeWeb-NEP, OurGrid, Condor та інші [3, 4]. Система BOINC є найбільш популярною та займає практично весь сегмент добровільних обчислень.

Яскравим прикладом BOINC-проекту в мережі Інтернет, що використовує вільні обчислювальні ресурси комп'ютерів добровольців, є проект SETI@home [5, 6]. Середня продуктивність обчислень цього проекту складає 1,6 PFLOPS [5, 6]. При цьому в проекті SETI@home бере участь понад 1,4 млн. користувачів, а обчислювальна система складається з 3 млн. вузлів [5, 6].

Розподілений проект BOINC складається із складеного сервера, який роздає завдання клієнтам та збирає результати, та програм-клієнтів, які підключаються до сервера та виконують основну роботу з одержання результатів. Через деякий час, коли завдання виконано, програма-клієнт знову підключається до сервера та передає йому результати, а також отримує нове завдання. За виконання завдання зареєстрований учасник проекту, якому належить програма-клієнт отримує кредити – бали, що характеризують процесорний час, витрачений на проведення обчислень.

Для участі в проектах BOINC підходить будь-який персональний комп'ютер, який має постійний або періодичний доступ до мережі Internet. З технічної точки зору участь у проекті виглядає так. Користувач сам завантажує та встановлює обчислювальний модуль - програму-клієнт. На більшості комп'ютерів процесор завантажений на 5-10%. На вільні 90-95% процесорної потужності претендує під час роботи програма розподілених обчислень. При цьому користувач не відчуває ніяких незручностей – програма бере для себе нижчий пріоритет в системі та не заважає роботі інших програм. Для кожного завдання кількість дублюючих один одного виконавців індивідуально, але значення за замовчуванням платформи BOINC дорівнює п'яти. Коли програма-клієнт звітує про проведені обчислення, сервер порівнює результати, що прислали різні учасники. Результат, прийнятий «кворумом» програм-клієнтів, вважається правильним, а невірні відповіді відкидаються. Учасники, які повернули до сервера невірні результати, не одержують кредитів. Кількість програм-клієнтів, що становить «кворум», також може бути вказано для кожного завдання окремо, а значення за замовчуванням дорівнює трьом.

Активних BOINC-проектів зараз дуже багато, тому визначити їх популярність досить складно [5]. Але можна виділити групу з SETI@home (проект з пошуку позаземних цивілізацій), Einstein@Home (пошук зірок, що випромінюють гравітаційні хвилі), World Community Grid (дослідження, що пов'язані з генетикою людини і пошуку ліків від хвороб), Rosetta@Home (вивчення тривимірної структури протеїнів), ClimatePrediction.net (прогнозування змін клімату Землі в найближчі 50 років), MilkyWay@ Home (вивчення процесу злиття галактик), LHC@home (обробка результатів з Великого адронного колайдера).

Користувач, який бажає допомогти розвитку науки, просто приєднується до цікавого йому науковому проекту, список яких є на сайті системи BOINC [5, 6]. З сайту проекту на персональний комп'ютер користувача завантажується виконуваний файл програми, потім викачуються вихідні файли з даними, які потрібно обробити. Користувач може вибрати не один проект, а кілька. Зазвичай такий вибір здійснюється на основі аналізу опису проектів. У таблиці 1 наводяться результати участі в деяких проектах системи BOINC української команди [6].

Таблиця 1 – Результати участі команди розподілених обчислень з України

Назва	Тематика	Місце команди	Кількість учасників у команді
DIMES	Інтернет	1 з 1689	223
Magnetism@Home	Фізика	2 з 269	160
ORG-27	Математика	11 з 4845	53
RC5-72	Криптографія	19 з 4845	293
FreeHAL@Home	Штучний інтелект	24 з 798	98
Burp	Рендеринг	28 з 840	24
Fightaids@Home	Медико-біологічне	30 з 30154	1066
Cosmology@home	Космос	33 з 2055	210
Climate Prediction	Планета земля	43 з 7759	240
QMC@Home	Хімія	44 з 2188	132
Chess960@Home	Ігрові проекти	95 з 688	10

Висновки

Використання розподілених обчислень – новий крок у розвитку наукових досліджень. Вчені зі всього світу розробляють проекти, які потребують проведення великої кількості розрахунків, а програмна платформа BOINC дає можливість приєднатися до цих проектів всім охочим, не вимагаючи ні спеціальної освіти, ні особливих апаратних ресурсів.

Розробка нового покоління ліків, пошук позаземних цивілізацій та моделювання молекул – все це вимагає обчислювальних ресурсів, якими можна поділитися без шкоди для себе.

Список використаних джерел:

1. Цирульник С. М. Віртуальні складові інтелектуальної око-процесорної розподіленої геоінформаційно – енергетичної системи/ С.М. Цирульник, В.П. Кожемяко, К.В. Кожемяко// Оптико-електронні інформаційно-енергетичні технології. – 2005. – №1(9). – С. 89–116
2. Барский А. Б. Параллельные информационные технологии: Учебное пособие / А. Б. Барский. – М.: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2007. – 503 с. – ISBN 978-5-94774-546-7.
3. IBM developerWorks [Електронний ресурс] / Ивашко Е. Часть 1. Высокопроизводительные вычисления в каждый дом. – Режим доступа: <http://www.ibm.com/developerworks/ru/library/l-grid/index.html>. – Назва з екрана.
4. IBM developerWorks [Електронний ресурс] / Ивашко Е. Распределенные вычисления: Часть 2. Архитектура высокопроизводительных вычислений на базе BOINC. – Режим доступа: <http://www.ibm.com/developerworks/ru/library/l-boinc/index.html>. – Назва з екрана.
5. BOINC [Електронний ресурс] / Open-source software for volunteer computing. – Режим доступа: <http://boinc.berkeley.edu/index.php>. – Назва з екрана.
6. Ukraine - Distributed Computing Team [Електронний ресурс]. – Режим доступа: <https://distributed.org.ua>. – Назва з екрана.