

ОБЩЕСТВЕННЫЙ КОМПЬЮТИНГ

Вовлечение людей в науку

* * *

Дэвид П. Андерсон

Лаборатория космических исследований
(Space Sciences Laboratory)

Калифорнийский университет, Беркли

Если раньше основные вычислительные мощности были сосредоточены в суперкомпьютерных центрах и институтских лабораториях, то теперь они распределены по сотням миллионов персональных компьютеров. Еще через несколько лет значительную долю общемирового вычислительного потенциала составят потребительские устройства: игровые или телевизионно-компьютерные приставки.

Эти изменения могут стать решающими для ученых, которым требуются экстремальные вычислительные возможности. Такие проекты, как SETI@home и Folding@home уже привлекли миллионы участников, отдающих время своих домашних ПК для научных расчетов. Ведется работа по подготовке аналогичных проектов в различных областях, что позволит проводить ранее неосуществимые научные исследования.

Реализация принципа "общественного компьютеринга" имеет не только научные, но и социальные последствия. Общественный компьютеринг обеспечивает основу для образования глобальных сообществ, концентрирующихся вокруг общих интересов и целей. Он заставляет общественность интересоваться текущими научными исследованиями. В конечном итоге он даст общественности возможность непосредственного контроля за направлениями научного прогресса.

Введение

Компьютерные технологии произвели революцию в науке. Ученые разработали точные математические модели физического мира, а компьютерные программы, реализующие эти модели, позволили аппроксимировать реальность на разных уровнях: атомное ядро, молекула белка, биосфера Земли, Вселенная. Можно создавать программы, позволяющие предсказывать будущее, подтверждать или опровергать теории и управлять "виртуальными лабораториями", позволяющими изучать химические реакции без приборов.

Вообще говоря, чем большей вычислительной мощностью мы располагаем, тем точнее аппроксимация реальности, что побуждает создавать все более и более быстрые компьютеры. Одним из способов повышения скорости вычислений является их "распараллеливание" – разделение на части, которые могут одновременно выполняться на разных процессорах. Именно так работает большинство современных суперкомпьютеров.

Широкомасштабному развитию технологии способствуют и экономические факторы. Так, компания может вложить больше средств в усовершенствование процессорных плат, если она планирует продавать их миллионами. Так, были быстро усовершенствованы платы для бытовых компьютеров типа Intel Pentium и Motorola PowerPC, фактически они вдвое увеличивали скорость примерно каждые 18 месяцев – тенденция, известная как "Закон Мура".

В 1990-е годы имели место два важных обстоятельства. Во-первых, в соответствии с законом Мура, персональные компьютеры стали очень быстрыми – такими быстрыми, какими были суперкомпьютеры всего несколько лет назад. Во-вторых, на потребительский рынок вышел Интернет. Внезапно появились миллионы быстрых компьютеров, соединенных в сеть. Идея использования этих компьютеров в качестве

одного параллельного суперкомпьютера родилась у многих людей одновременно. В 1997 году появилось два таких проекта: GIMPS – поиск больших простых чисел и Distributed.net – дешифровка закодированных сообщений. Эти проекты привлекли тысячи участников.

В 1999 году был запущен третий проект, SETI@home, цель которого - обнаружение радиосигналов от разумных внеземных цивилизаций [1]. SETI@home действует как экранная заставка, работающая только тогда, когда ПК простаивает, и дающая графическое представление выполненной работы, когда она закончена. SETI@home оказался привлекательным не только для тех, для кого поиск внеземных цивилизаций – это хобби, но и для миллионов участников во всем мире. Этот проект способствовал появлению других академических проектов, а также созданию нескольких компаний, которые стремятся коммерциализировать принцип общественного компьютеринга.

1. Возможности общественного компьютеринга

Общественный компьютеринг может предоставить больше вычислительных возможностей, чем любой суперкомпьютер, кластер или грид, и это расхождение будет со временем только увеличиваться. Сегодня в SETI@home задействовано приблизительно миллион компьютеров, что обеспечивает производительность в 60 TFLOPS, в то время как самый мощный суперкомпьютер IBM ASCI White имеет производительность около 12 TFLOPS. Один миллион компьютеров SETI@home составляет меньше процента от 150 миллионов ПК во всем мире, связанных через Интернет. Предполагается, что это число увеличится до одного миллиарда к 2015 году. Таким образом, общественный компьютеринг имеет потенциальную возможность дать петафлопную вычислительную мощность.

Согласно закону Мура, скорость процессора удваивается каждые 18 месяцев. Еще быстрее прогрессируют "графические сопроцессоры" – время удвоения их производительности составляет около 8 месяцев, причем современные графические процессоры выполняют операции с плавающей запятой во много раз быстрее, чем обычные. Графические процессоры становятся более программируемыми и гибкими, и исследователи активно вкладывают средства в их использование в научном компьютеринге. Поскольку графические процессоры встроены в современные ПК, это дает общественному компьютерингу преимущество перед другими направлениями.

Наряду с вычислительными средствами большинству вычислительных задач требуется память. Общественные ресурсы могут обеспечить беспрецедентный объем памяти. На сегодняшний день типовой ПК имеет около 80 Гбайт памяти, причем это гораздо больше, чем обычно использует его хозяин. Если 100 миллионов пользователей предоставят по 10 Гбайт памяти, получится Эксабайт (10 в 18 степени), что больше, чем объем любой централизованной системы памяти.

2. Социальные аспекты общественного компьютеринга

Общественный компьютеринг эффективен только при условии, что число участников велико, и в этом смысле проект SETI@home оказался успешным – в него было вовлечено 4,6 миллионов участников, из которых около 600 тыс. остаются активными.

Люди узнали о SETI@home благодаря тому, что материалы о проекте появлялись в средствах массовой информации, на форумах в Интернет типа Slashdot [2]. Кроме того,

картинки экранной заставки в SETI@home также являются мощным механизмом популяризации проекта: в офисах и школах, где на компьютерах работают много людей, компьютер, используемый в SETI@home, служит прекрасной рекламой проекта.

Кто участвует в проекте SETI@home и почему? Чтобы ответить на этот вопрос, мы в интерактивном режиме провели социологическое исследование, в котором приняло участие около 130 тыс. респондентов. Наш Web-сайт позволяет интерактивно создавать "профили", где пользователи сообщают сведения о себе, и около 50 000 сделали это. Мы собрали сообщения от многих тысяч участников, а затем провели случайную выборку для тысячи пользователей.

Наш опрос показал, что 92% пользователей SETI@home – мужчины, и большинство из них мотивирует свое участие в проекте, главным образом, научным интересом – они хотят знать, существуют ли внеземные цивилизации. Другая побудительная причина – это общественное признание. SETI@home отслеживает вклад каждого пользователя (т.е. объем выполненных вычислений) и располагает на сайте "списки лидеров", где пользователи упорядочиваются по их вкладу. Пользователи могут образовывать "команды", которые имеют собственные списки лидеров. Механизм команд оказался весьма эффективным для привлечения новых участников.

Некоторые участники SETI@home пытаются жульничать – получить очки за невыполненные вычисления. Еще большую проблему создают пользователи – вандалы, которые намеренно возвращают неверные результаты, что приводит к существенному обесцениванию всей остальной вычислительной работы. Эти проблемы можно обойти, если проводить избыточные вычисления и сравнивать результаты.

Вклад участников SETI@home – это не только процессорное время. Добровольцы перевели сайт SETI@home на 30 языков, разработали различные расширяемые программные средства и создали вспомогательные сайты. Мы придаем большое значение созданию условий для такого рода вкладов в проект.

Вокруг SETI@home образовались различные "сообщества": мировое сообщество, члены которого взаимодействуют через сайт; национальные или языковые сообщества со своими собственными сайтами; группа пользователей SETI@home в Германии заключила коммерческие договоры на несколько лет; по крайней мере, три пары познакомились и поженились благодаря SETI@home.

3. Технические аспекты общественного компьютеринга

Управление проектом общественного компьютеринга требует адаптации прикладных программ к различным платформам, обеспечения работы серверных систем и баз данных, отслеживания оплаты счетов пользователей, учета возможности избыточной и ошибочной информации и решения ряда других проблем. Сейчас мы разрабатываем программное обеспечение, называемое Berkeley Open Infrastructure for Network Computing (BOINC), которое решает или помогает решить большинство из этих проблем. При помощи BOINC можно просто преобразовать уже имеющиеся приложения для их использования в проекте общественного компьютеринга. BOINC-проекты автономны, каждый поддерживает свои собственные серверы и базы данных и не зависит от других проектов. Участники могут регистрироваться в нескольких проектах и могут контролировать использование их

ресурсов (например, пользователь может выделять 60% времени своего процессора на изучение глобального потепления и 40% на SETI).

Сейчас разрабатывается несколько проектов, базирующихся на BOINC, включая SETI@home, биохимический проект Folding@home [4] и проект изучения климата – Climateprediction.net [3]. BOINC является дополнением для грид-систем, которые обеспечивают разделение ресурсов внутри и между институтами, но не поддерживают общественный компьютеринг [5].

4. Применение общественного компьютеринга

Чтобы какая-либо задача могла решаться при помощи общественного компьютеринга, она должна делиться на независимые части, причем доля вычислений по отношению к передаче данных в этих частях должна быть достаточно высока (иначе стоимость пересылки данных через Интернет может превысить стоимость вычислений). Такими свойствами обладают многие задачи:

- Сложные физические системы имеют компоненты случайной и хаотической природы. Их внешнее поведение является вероятностным, не точным. Изучение статистики для таких систем требует многократных операций моделирования с разными случайными начальными и конечными условиями. Эти операции могут выполняться параллельно.
- Развивается область "случайных алгоритмов", которые позволяют получать приближенное решение для точных задач. При этом делаются случайные попытки, которые могут выполняться параллельно.
- Во многих областях применяются "генетические алгоритмы". Этот подход предусматривает создание популяции (совокупности) приближенных решений задачи и использование механизма естественного отбора для получения оптимального решения.
- Модели физических систем часто имеют большое число основных параметров, чьи оптимальные значения неизвестны, а зависимость между ними нелинейна. Исследование области значений таких параметров требует большого числа независимых операций моделирования. Например, алгоритмы Монте-Карло предусматривают большое число независимых вычислений, соответствующих выборке в многомерном пространстве.
- Распараллеливание присуще приложениям, связанным с анализом большого количества данных, таких как телеметрия от радиотелескопа (например, SETI@home) или от ускорителя частиц. Ограничивающим фактором является отношение вычисления/данные.
- Некоторые медицинские проекты связаны с исследованием миллионов или миллиардов молекул (например, исследование потенциальных наркотиков). Эти задачи легко распараллеливаются. Аналогично, распараллеливаются некоторые проекты в области генетики, касающиеся сопоставления набора белков с последовательностью ДНК

5. Заключение

Карл Саган заметил, что общее отношение общественности к науке – это растущее отчуждение и даже враждебность [7]. Общественный компьютеринг может помочь изменить эту тенденцию. Если владельцы компьютеров смогут предоставить свои ресурсы любому из научных проектов, они получают возможность изучить и оценить эти проекты, ознакомиться с их целями, методами и шансами на успех. Дальнейшему развитию этого процесса может способствовать создание "рынков решений", где люди смогут делать виртуальные ставки или инвестиции, основываясь на результатах научных проектов, аналогично тому, как это делается на рынках политических решений [8].

Поскольку владельцы компьютеров могут присоединяться к проектам по своему выбору, контроль за выделением ресурсов для науки может переместиться от учреждений, финансируемых правительством (с несметным числом факторов, определяющих их политику), к обществу. Однако, здесь есть определенный риск: общество легче ввести в заблуждение, чем комиссию экспертов. Но такой подход предлагает очень ясный и демократичный механизм для выбора стратегии исследований.

Если у ученого есть идея, реализация которой требует значительного объема вычислений, и он обнаруживает, что это займет миллионы часов компьютерного времени, естественная реакция – выбросить эту идею в мусорную корзину. Но общественный компьютеринг делает такие идеи осуществимыми, например, в проекте SETI@home использовалось 1,5 миллиона лет процессорного времени.

Библиография

- [1] D. P. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer. SETI@home: An experiment in public-resource computing. Communications of the ACM, Nov. 2002, Vol. 45 No. 11, pp. 56-61. См. также <http://setiathome.berkeley.edu>
- [2] <http://www.slashdot.org>
- [3] <http://climateprediction.net>
- [4] <http://folding.stanford.edu>
- [5] <http://www.globalgridforum.com/>
- [6] R. Motwani and P. Raghavan. Randomized Algorithms. Cambridge University Press, 1995.
- [7] C. Sagan. The Demon-Haunted World: Science As a Candle in the Dark. Random House, 1996.
- [8] R. Forsythe, T. A. Rietz, and T. W. Ross. Wishes, expectations, and actions: A survey on price formation in election stock markets. Journal of Economic Behavior and Organization, 39:83-110, 1999.