

ОПТИМАЛЬНАЯ РЕПЛИКАЦИЯ ЗАДАНИЙ В DESKTOP GRID

Чернов И.А., Никитина Н.Н.

ИПМИ КарНЦ РАН, г. Петрозаводск; {chernov,nikitina}@krc.karelia.ru

Доступность и широкое распространение персональных компьютеров и других неспециализированных вычислительных устройств повышают интерес к высокопроизводительным вычислительным системам типа Desktop grid и расширяют круг задач, которые могут быть решены с их помощью. При этом возникают новые задачи, связанные с моделированием самого вычислительного процесса в такой системе. В частности, это минимизация рисков, связанных с принятием неверного ответа.

Задача, решаемая в вычислительной сети типа Desktop grid, обычно состоит из большого количества относительно простых (с вычислительной точки зрения) заданий. Задания можно трактовать как «черные ящики», генерирующие элемент некоторого множества. Типичными примерами являются задачи распознавания с ответами «да» и «нет», вычисление расстояния между двумя объектами в некотором метрическом пространстве, расчет целочисленной или вещественной характеристики изучаемого объекта и т. п. Не исключены ошибки, то есть получение неверных ответов от вычислительных узлов. Ошибки могут быть разной природы: от злонамеренных действий до генерации неверного ответа правильным алгоритмом. Последний случай вполне возможен, например, в итерационных алгоритмах при неудачно выбранных параметрах сходимости и точности или в алгоритмах численной оптимизации при обнаружении «ложного» локального минимума. При этом недетерминированный алгоритм (использующий случайные числа) может выявить ошибку при повторном расчете.

Репликация — рассылка одинаковых заданий двум и более вычислительным узлам — способна снизить вероятность неверного ответа и тем самым — численное выражение затрат. Цена ошибки может быть очень высока (например, стоимость лабораторного исследования молекулы, претендующей на прочное связывание с белком и представляющей, потенциально, интерес для медицины, измеряется тысячами евро), особенно если сравнивать ее с ценой однократного расчета одного задания (минуты работы простаивающего персонального компьютера). Вероятность ошибки, даже «малую» с интуитивной точки зрения, приходится принимать во внимание в больших проектах из-за огромного количества заданий.

Вероятность выявления ошибки при репликации зависит от распределения вероятностей возможных ответов. Так, при непрерывном распределении вероятностей неверных ответов репликация гарантированно выявляет ошибку при любой вероятности правильного ответа. Ошибки могут также иметь разную цену; например, если множество возможных ответов содержит три элемента («хороший», «пригодный», «плохой»), то шесть видов ошибок, очевидно, различаются своими последствиями.

В ряде случаев достаточно факта обнаружения ошибки, после чего задание направляется доверенному узлу для нахождения «правильного» ответа. Однако зачастую целесообразнее решать вопрос «Quid est veritas» через кворум: принимается ответ, полученный заданное число раз, даже если другие варианты ответа также были получены. При этом сохраняется вероятность, что будет принят неверный ответ и он, позже обнаруженный на практике, приведет к тем или иным убыткам.

Противоречие между затратами вычислительных ресурсов на репликацию и снижением средних затрат приводит к оптимизационным задачам. В работах [1-3] мы рассматриваем ряд вопросов выбора оптимального кворума, подробно изучив задачи распознавания и коснувшись задач с тремя и бесконечным множеством ответов. В докладе будут обсуждаться эти результаты, в том числе новые численные эксперименты с параметрами из статистических данных реальных проектов, а также некоторые новые результаты для различных распределений вероятностей и цен ошибок разного рода.

Литература

1. *Chernov I., Nikitina N.* Virtual Screening in a Desktop Grid: Replication and the Optimal Quorum // Lecture Notes in Computer Science. Parallel Computing Technologies 13th International Conference, PaCT 2015. P. 258.
2. *Chernov I.* Optimal Quorum for the Model of Computational Grid with Redundancy // Proceedings of The Autumn Session of the VIII International Workshop "Applied Problems in Theory of Probabilities and Mathematical Statistics Related to Modeling of Information Systems", pp. 648-651.
3. *Чернов И.А.* Оптимальное дублирование заданий в вычислительной системе // Труды Карельского научного центра РАН, №4, 2014 г, стр. 130–136.