

ОЦЕНКИ И ХАРАКТЕРИСТИКИ РАСПРЕДЕЛЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ СРЕД

Ю.А. Жолудев, С.И. Соболев, К.С. Стефанов

Научно-исследовательский вычислительный центр МГУ им. М.В. Ломоносова

При решении больших задач с помощью параллельных вычислений важным моментом является анализ эффективности использования вычислительных ресурсов. Результатом анализа становятся оценки эффективности и масштабируемости, служащие ориентиром как при разработке и оптимизации параллельных программ, так и при проектировании и модернизации вычислительных комплексов. Цель таких работ очевидна – необходимо решать конкретную задачу или класс задач за минимальное время. Для суперкомпьютерных систем с традиционными архитектурами давно сформирован комплекс характеристик и показателей, описывающих основные свойства вычислительной системы и поведение прикладной программы в ее среде. Прежде всего это пиковая и реальная производительность (на синтетическом тесте или конкретной задаче), эффективность (отношение реальной и пиковой производительности), пропускная способность и латентность коммуникационной среды. Существенно, что и пользователи, и администраторы системы хорошо представляют себе основные архитектурные решения вычислительной системы (состав вычислительных устройств, операционную систему, топологию сети). Это и позволяет с использованием известных методик делать выводы относительно поведения в такой системе прикладных задач заданного класса.

Статья открывает серию публикаций по формированию методов оценки качества распределенных компьютерных систем, полученных в рамках научно-исследовательского проекта, ведущегося в НИВЦ МГУ им. М.В. Ломоносова. Цель этого проекта – создание методов и средств для выявления основных характеристик распределенных неоднородных вычислительных сред и описания поведения прикладных задач в таких средах. Безусловно, первым шагом выполнения проекта стало исследование уже имеющихся наработок по анализу характеристик распределенных систем и сред. В статье описывается методы формирования таких характеристик, приводится анализ их применимости для технологии распределенных вычислений X-Com [9-10], а также предложения по созданию собственной системы характеристик и показателей.

Известные оценки для параллельных вычислительных систем чаще всего основываются на моделях, в которых система рассматривается как совокупность вычислительных элементов, способных решать общую задачу. Рассмотрим несколько моделей производительности для таких систем.

Наиболее широко распространенная модель, описанная в [1], рассматривает совокупность вычислительных элементов без учета структуры связей между ними. В рамках модели вводится несколько характеристик производительности: коэффициент накладных расходов ε , ускорение S и эффективность E , вычисляемые по следующим формулам:

$$T_0 = nT_p - T_s, \varepsilon = \frac{T_0}{T_p}, S = \frac{T_s}{T_n}, E = \frac{S}{n}$$

где T_s – время последовательного решения задачи на одном процессоре,

T_p – время параллельного решения задачи на n процессорах,

T_0 – время накладных расходов на организацию параллельных вычислений.

Это классическая общепризнанная модель, используемая для оценки производительности однородных вычислительных комплексов, таких как компьютеры с разделяемой памятью и одинаковыми процессорами, многопроцессорные вычислительные комплексы, вычислительные кластеры.

Обобщением классической модели производительности вычислений является модель коллектива вычислителей [2], которая описывает структуру связей между элементами вычислительной системы. Структура коллектива вычислителей представляется неориентированным графом $G = \langle C, E \rangle$, где вершинам сопоставлены вычислители из множества C , а ребрам графа E соответствуют связи между вычислителями, по которым возможны обмены данными. Модель коллектива вычислителей $R = \langle H, A \rangle$ при этом представляется в виде совокупности описания структуры связей $H = \langle C, E \rangle$ и алгоритма работы вычислителей A . Среди характеристик вычислительной системы в рамках данной модели выделяют порядок вершин (число инцидентных ребер) и диаметр графа (максимальное число вершин на пути между двумя произвольными вершинами графа). Структуры с наименьшим диаметром обеспечивают минимальные задержки при передаче данных. Алгоритм A для такой модели представляется в виде $A = A(P(D))$, где D – исходные данные для параллельной программы P , и D представимы в виде объединения индивидуальных для каждого вычислителя массивов данных

$$D = \bigcup_{i=1}^N D_i,$$

а параллельная программа представляется в виде совокупности ветвей P_i , выполняемых разными вычислителями:

$$P = \bigcup_{i=1}^N P_i, \bigcap_{i=1}^N P_i \neq \emptyset.$$

Для вычислительных систем в рамках такой модели характерны принципы однородности (одинаковая функциональность и производительность всех вычислителей) и локальности (вычислители взаимодействуют с ограниченной частью других вычислителей).

Обобщением двух описанных моделей на неоднородные системы являются модели систем функциональных устройств [3], вводящая пиковые производительности отдельных устройств π_i , при этом пиковая производительность всей системы вычисляется в виде суммы $\pi = \sum \pi_i$. Реальная производительность r вычислительной системы вычисляется как отношение количества элементарных операций в задаче A ко времени ее выполнения T : $r = \frac{N}{T}$.

На данном этапе возникают такие понятия как загруженность устройств на интервале времени

$p_i(t_1, t_2) = \frac{\pi(t_2 - t_1)}{V(t_2) - V(t_1)}$, где V – функция стоимости использования выполненной работы, и загруженности системы:

$$P = \sum_{i=1}^n \alpha_i p_i, \alpha_i = \frac{\pi_i}{\pi}.$$

Тогда коэффициент ускорения таких систем будет вычисляться по формуле

$$S = \frac{\sum_{i=1}^n p_i \pi_i}{\max_i \pi_i}.$$

Как видно из описанного выше, для применения указанных моделей необходимо иметь представление не только о вычислительной сложности задачи A , но и об индивидуальных характеристиках каждого из узлов вычислительной системы. Представленная модель находит свое дальнейшее развитие в моделях систем с расписанием [4], моделях с предположением о сложности индивидуальных задач для каждого из вычислителей, моделях с использованием алгоритмов балансировки [5] и так далее.

Описанный подход для определения характеристик вычислительных систем и решаемых с их помощью задач, заключающийся в скрупулезном моделировании каждого из аспектов системы, встречается ряд проблем, когда речь заходит о системах на основе распределенных вычислительных сред, построенных с помощью системы метакомпьютинга X-Com. Для таких систем характерна динамичность практически всех компонентов вычислительной среды и отсутствие информации о вычислительной сложности исходной задачи, заключающейся в обработке порций данных на стороне вычислительного клиента и обработке порций результатов на стороне сервера. Фактическая производительность узлов и характеристики каналов связи неодинаковы и могут меняться в ходе решения задачи.

Исходя из этих предпосылок, предлагается инструментальный подход, заключающийся в сборе сведений обо всех этапах функционирования каждого из компонентов вычислительной среды в период решения задачи (передачи запросов к серверу, обработка порций, и т.д.) и оценке временных интервалов и количественных характеристик каждого из этих этапов. С одной стороны, это позволяет построить более точную модель вычислительной системы, не являющейся обобщением одной из общепринятых моделей, с другой – избежать использования в модели значений, не поддающихся точному измерению или вычислению.

Такой подход к оценке характеристик систем часто используется в тестах производительности (benchmark) [6]. Например, для задач в распределенных вычислительных системах исследуется интенсивность вычислений (computation intensity) и интенсивность обменов данными (data intensity) [7]. Интенсивность вычислений с различными вариациями определяется как количество вычислительной работы, приходящейся на элемент массива данных, передаваемых за один обмен между элементами системы. Интенсивность обмена данными – обратная интенсивности вычислений характеристика, которая отражает объем передаваемых данных, приходящийся на выполнение единицы вычислительной работы.

Для самих распределенных вычислительных систем вводится так называемое время вычислительного цикла (turnaround time) и пропускная способность (throughput) вычислительной системы [8]. Время вычислительного цикла определяется как время между началом решения части задачи и получением результатов ее решения. Пропускная способность вычислительной системы – максимальный объем передаваемых между узлами системы данных, при котором время вычислительного цикла не возрастает. Обе эти оценки измеряются в ходе решения синтетических или реальных задач. Для каждой из них дается более точное определение и способ вычисления указанных оценок с учетом специфики как самой задачи, так и вычислительной среды, на которой они решаются.

В связи с ограниченностью класса задач системы метакомпьютинга X-Com схемой bag-of-tasks [6] вычисление подобных характеристик возможно в ходе решения не только синтетических задач в условиях статичности состава вычислительной среды, но и при решении реальных задач. Этого можно достичь путем включения измерений характеристик рабочего цикла вычислительных узлов непосредственно в программный инструментарий системы: для каждой задачи известны конкретные этапы вычислительных циклов, а основная информация об объемах передаваемых данных доступна на серверной части вычислительной среды и отражается в журналах сервера. Тем не менее, для точного определения конкретных показателей производительности, накладных расходов и интенсивности вычислений могут представлять трудности такие особенности хода решения задач, как динамичность производительности вычислительных узлов (которые могут принимать участие в решении задачи не в монопольном режиме и временно выбывать из расчета) и характеристик каналов связи.

Для создания комплекса характеристик, описывающих решение задачи в распределенной среде X-Com, построена базовая модель системы X-Com, в которой были выделены основные этапы вычислительного цикла. Базовая модель предполагает простейшую архитектуру распределенной среды, в которой существует один сервер и множество общающихся с ним клиентов. За рамками этой модели остаются более сложные случаи, в частности, использование промежуточных серверов, взаимодействующих со своими подмножествами клиентов либо с системами очередей вычислительных комплексов. Эти случаи будут описаны в последующих публикациях по тематике проекта.

Для данной модели в настоящее время возможно вычисление следующих показателей.

«Серверная эффективность» – отношение суммарного процессорного времени, затраченного на обработку вычислительной порции на клиенте (т.е. суммарное время работы вычислительных модулей прикладной задачи) к условному суммарному серверному времени, которое для каждой порции считается с начала обработки запроса на создание порции до сохранения результатов ее обработки. Данный показатель используется как оценка накладных расходов, т.е. всех затрат, которые возникают помимо собственно вычислений. Показатель легко интерпретируем в случае простейшей архитектуры распределенной среды. При использовании промежуточных серверов последние становятся источником дополнительных накладных расходов, которые на данный момент не учитываются моделью.

«Клиентская эффективность» – соотношения числа отправленных и числа принятых порций. Данный показатель отражает фактор «потерянных порций» в результате отключения или перезапуска клиентов, а также неизбежно возникающих в конце решения задачи вследствие алгоритма распределения порций в системе X-Com. Показатель хорошо иллюстрирует динамичность вычислительной среды.

«Эффективность проведения расчета» – отношение суммарного времени всех клиентов, участвующих в расчете, затраченного на проведение элементарных операций, к суммарному времени участия всех клиентов в расчете. Такой показатель отражает степень вовлеченности всех узлов в ход распределенного расчета, на основе которого можно делать выводы о наличии внешних факторов, влияющих на общую производительность вычислительной среды. Данная оценка также может учитывать и избыточность операций, т.е. перерасчет вычислительных порций.

Имеется еще ряд показателей, представляющих определенный интерес с точки зрения характеристик распределенных сред и проведенных расчетов, которые, однако, нуждаются в дополнительных исследованиях.

«Пиковая производительности распределенной среды». Понятно, что простейший способ получения этой оценки суммированием пиковых производительностей всех участвовавших в расчете узлов слишком груб, т.к. полностью игнорирует динамичность состава среды. В этом варианте он может рассматриваться как оценка максимального потенциала среды, своего рода оценка сверху. Более адекватным вариантом представляется оценка суммарной производительности реально работающих над задачей узлов в заданный интервал времени. Однако она оказывается зависимой от времени и, кроме того, может быть заниженной, поскольку число работающих узлов в X-Com можно определить только по приходящим от них запросам, при этом не обязательно в заданный интервал времени будут получены запросы от всех реально работающих узлов. Тем не менее, данный показатель можно представлять графически, интерпретируя его как нижнюю оценку пиковой производительности среды.

«Реальное клиентское процессорное время». С помощью системной информации можем оценить эффективность выполнения исполнимого модуля прикладной задачи на клиенте. Это может представлять интерес при анализе поведения вычислительного мо-

дуля на различных программно-аппаратных платформах, а также при работе клиента в фоновом режиме, т.е. одновременно с другими задачами. Взвешенная по показателям производительности вычислительных узлов, такая оценка может использоваться для оценивания вычислительной сложности прикладной задачи и сложности обработки каждой порции, что позволит применить оценки производительности системы с предположениями трудоемкости алгоритма решения задачи [11].

Адекватность и применимость этих и других показателей будет рассмотрена в ходе дальнейших работ по проекту. Кроме того, будет расширена модель вычислительного цикла системы X-Com, уточнены алгоритмы получения описанных выше показателей для этой модели, созданы инструментальные средства для получения и анализа показателей, а также для формирования рекомендаций пользователям по оптимизации их задач с учетом заданных требований (минимизация времени решения задачи, минимизация используемых ресурсов и т.д.).

Работа выполняется при поддержке гранта Президента Российской Федерации для молодых российских ученых-кандидатов наук МК-5104.2011.9.

Литература

1. Grama A., Gupta A., Karypis G., Kumar V. Introduction to Parallel Computing. – 2nd ed. – USA: Addison Wesley, 2003.
2. Хорошевский В.Е. Архитектура вычислительных систем: Учеб. пособие для вузов. – М.: Изд-во МГТУ им. Н. Э. Баумана, 2005.
3. Воеводин В.В., Воеводин Вл. В. Параллельные вычисления. – СПб.: БХВ-Петербург, 2002.
4. Посыпкин М.А., Хританков А.С. О понятии ускорения и эффективности в распределенных системах // Труды Всероссийской научной конференции «Научный сервис в сети Интернет: решение больших задач». – 2008. – С.149-155.
5. Хританков А.С. Модели и алгоритмы балансировки нагрузки. Модели коллектива вычислителей. Модели с соперником // Информационные технологии и вычислительные системы. – 2009. – № 2. – С. 65-80.
6. Allan Snaveley, Greg Chun, Henri Casanova, Rob F. Van der Wijngaart, Michael A. Frumkin «Benchmarks for Grid Computing: A Review of Ongoing Efforts and Future Directions» // A CM SIGMETRICS Performance Evaluation Review Homepage. Vol. 30. Issue 4, March 2003.
7. Vudutala China Venkata Rao, B. Ranga Swami. Low-Level Grid Computing Benchmarks // In proc. International Conference on High Performance Computing (HiPC-2005), Dec 18-21, 2005 Goa, India.
8. Michael Frumkin, Rober Hood. Using Grid Benchmarks for Dynamic Scheduling of Grid Applications. // NAS Technical Report NAS-03-015 October 2003, URL: <http://www.nas.nasa.gov/News/Techreports/2003/PDF/nas-03-015.pdf> [PDF].
9. Воеводин Вл.В., Жолудев Ю.А., Соболев С.И., Стефанов К.С. Эволюция системы метакомпьютинга X-Com // Вестник Нижегородского государственного университета им. Н.И. Лобачевского. 2009. №4. С 157-164.
10. Система метакомпьютинга X-Com: официальный сайт, <http://x-com.parallel.ru>.
11. Хританков А.С. Анализ производительности распределенных вычислительных комплексов на примере системы X-Com. // Труды Всероссийской суперкомпьютерной конференции «Научный сервис в сети Интернет: масштабируемость, параллельность, эффективность» (Новороссийск, 21-26 сентября 2009). 2009. С. 46-52.