

Методы секторного моделирования в лабораторном практикуме по параллельному программированию для студентов магистратуры

И.Г. Захарова, А.А. Захаров

ФГБОУ ВО «Тюменский государственный университет»

В работе представлен опыт использования заданий, опирающихся на секторное моделирование или декомпозицию области решения, в лабораторном практикуме по параллельному программированию и распределенным вычислениям для магистерской программы «Высокопроизводительные вычислительные системы». Задания имеют различный уровень сложности: от склейки формально обрабатываемых массивов или изображений до сопряжения секторных гидродинамических моделей. В дополнение к классическим заданиям по параллельному программированию они включают исследование алгоритмов и правил сопряжения, решение вопросов балансировки нагрузки и обеспечения отказоустойчивости.

Ключевые слова: параллельное программирование, программа магистратуры, секторное моделирование, балансировка нагрузки, отказоустойчивость

1. Введение

В Тюменском государственном университете с 2013 г. реализуется магистерская программа «Высокопроизводительные вычислительные системы» по направлению «Математическое обеспечение и администрирование информационных систем». Она ориентирована на комплексную подготовку профессионалов в области разработки и сопровождения программного обеспечения, комплексирования, администрирования и защиты высокопроизводительных вычислительных систем. Среди обучающихся в магистратуре выпускники не только одноименного направления бакалавриата, но и других родственных образовательных программ (Информационная безопасность, Информационные системы и технологии, Математика, Математика и компьютерные науки, Механика и математическое моделирование и др.), имеющие различный уровень подготовки в области алгоритмов и структур данных, технологий и языков программирования. Кроме того, многие магистранты уже работают, как правило, в проектных и ИТ-подразделениях крупных компаний (Роснефть, Газпром, Schlumberger и др.), имея при этом вполне определенные профессиональные интересы.

Поэтому требуется предусмотреть возможность индивидуализации обучения в зависимости от уровня начальной подготовки и предпочтительности углубленного изучения системного или прикладного программирования, администрирования или защиты информации. При этом в некоторых дисциплинах общего характера (например, «Методика преподавания компьютерных наук») или, наоборот, узкоспециальных («Вычислительный эксперимент в MATLAB», «Разработка мобильных приложений» и др.) допускается смещение акцентов и даже исключение некоторых тем с учетом предпочтений студентов и тематики их научно-исследовательской работы или проектных разработок. Однако, в базовых курсах «Алгоритмы и технологии разработки параллельных программ» и «Архитектура многопроцессорных вычислительных систем» необходимо добиться освоения на уровне «готовность к использованию» для всех предусмотренных разделов. Основным инструментом здесь безусловно является выполнение проектных заданий лабораторного практикума. При разработке таких заданий необходимо обеспечить, с одной стороны, вариативность по степени трудности, уровню формализации или, напротив, открытости постановки исходной задачи. С другой же стороны, у всех заданий должно быть некое общее основание, гарантирующее достижение целей обучения. В качестве такого основания представляется возможным использовать в постановках заданий лабораторного практикума магистратуры обобщенные секторные модели.

2. Логика построения заданий лабораторного практикума

Вопросам разработки лабораторных практикумов по параллельному программированию уделяется достаточно большое внимание [2, 4, 5, 7]. При этом на уровне бакалавриата задания ориентированы, как правило, на программную реализацию достаточно стандартного набора алгоритмов решения задач дискретной и вычислительной математики с использованием технологий OpenMP, MPI, CUDA, multithreading (C/C++, Java, Python) и проведение типовых вычислительных экспериментов. Однако, такой практической подготовки для проведения исследований уровня магистратуры и реализации прикладных проектов, использующих технологии распределенных вычислений, явно недостаточно. Рано или поздно программист, разрабатывающий программное обеспечение для распределенных систем, сталкивается с необходимостью управления балансировкой нагрузки и обеспечения отказоустойчивости. Безусловно, что понимать суть соответствующих проблем и иметь представление о методах их решения также должны специалисты в области администрирования и защиты распределенных вычислительных систем. Именно поэтому в контексте изучения методов повышения эффективности параллельных программ в соответствии с содержанием свода знаний и умений в области суперкомпьютерных технологий [2-4, 11] в традиционные задания (близкие по содержанию изучаемых параллельных алгоритмов, в частности, к курсу [5, 7]) практикума по дисциплинам «Алгоритмы и технологии разработки параллельных программ» и «Архитектура многопроцессорных вычислительных систем» необходимо включить соответствующие вопросы. При этом необходимо подчеркнуть, что цель практикума – это освоение технологий программирования, которое должно интегрироваться с научно-исследовательской работой магистранта.

Основная проблема, которую потребовалось решать при разработке практикума – это создание заданий в единой логике с учетом отличий в уровне начальной подготовки студентов магистратуры в области параллельных вычислений. Суть в том, что только выпускники укрупненных групп направлений «Информационная безопасность» и «Компьютерные и информационные науки» ранее слушали соответствующие курсы [8] и, в целом, имеют наиболее высокий уровень знаний по программированию, операционным системам и сетевым технологиям. В то же время, математики и механики, отставая в плане вышеперечисленных дисциплин, свободнее в вопросах математического моделирования. Наконец, выпускники бакалавриата укрупненной группы «Информатика и вычислительная техника» больше ориентированы не на разработку математического и программного обеспечения вычислительных систем, а на проектирование и сопровождение программных продуктов. Соответственно, и задания должны учитывать эти особенности.

Учитывая вышесказанное, полагаем, что достаточно эффективным может оказаться использование обобщенных секторных моделей в качестве шаблона для постановки задач. Речь идет не только об изучении подходов к решению задач, где применение таких моделей для организации вычислений является не только органичным, но и, зачастую, единственно возможным способом получения решения (например, секторные гидродинамические модели пласта). Более полезным в учебных целях оказывается использование обобщенной постановки, которую можно наполнять разным содержанием. Использование термина «обобщенные» объясняется тем, что здесь можно говорить и о декомпозиции области счета, и об изначально существующих сопрягающихся областях, и о декомпозиции данных.

При разработке подходов к решению соответствующих задач естественным образом возникает необходимость исследования следующих проблем: 1) правила и алгоритмы сопряжения моделей (решений в подобластях); 2) балансировка нагрузки на вычислительных узлах; 3) обеспечение отказоустойчивости при проведении вычислений. Решать вопросы балансировки и отказоустойчивости в учебных приложениях, с одной стороны, можно достаточно формально, не связывая их с природой исходной задачи (разве что выделяя динамическую или статическую постановку). С другой же стороны, первая проблема требует особого внимания, поскольку именно особенности сопряжения моделей могут оказаться определяющими для выбора способов балансировки нагрузки и решения вопросов отказоустойчивости.

Представляется возможным именно в учебных целях, учитывая также и различные профессиональные интересы студентов, использовать задачи трех типов: 1) сопряжение на основе условий предметной области (физических законов [1, 12], связей социальных сетей [6, 10] и др.);

2) сопряжение на основе регламентирующих критериев (задачи полного или частично-полного перебора, связанные с компрометацией систем защиты информации, в частности, взлом криптоалгоритмов и криптопротоколов; задачи разработки надежных контроллеров для программно-конфигурируемых сетей; задачи поиска уязвимостей в распределенных компьютерных системах [9]); 3) сопряжение по заранее заданным формальным правилам. В соответствии с этим варьируются и требования по балансировке и отказоустойчивости. Этот подход позволяет выстроить единую систему заданий проектного типа, имеющих различный уровень сложности. Задания выполняются в течение первого семестра и, как показал опыт, полученные решения достаточно часто получают развитие в курсовых работах второго семестра.

3. Примеры заданий

Ниже в качестве иллюстрации представлены задания лабораторного практикума. Студенты могут выбирать любые из предложенных заданий (в том числе, и для групповых проектов), кроме задания 1, которое носит формальный характер и выполняется индивидуально. Возможно использование любых технологий (OpenMP, OpenACC, CUDA, Microsoft TPL, MPI, Microsoft WCF и др.) и языков параллельного (распределенного) программирования (C/C++, Java, Python, C#, Erlang и др.). Допускается эмуляция распределенных вычислений на одном узле, а также симуляция отказа запросом на отмену задачи. Вариации заданий представлены в скобках.

Задание 1: реализовать обработку N массивов (списков, графов, текстовых файлов) различной длины одним и тем же алгоритмом известной вычислительной сложности (поиск, сортировка) на K вычислительных узлах заданной производительности (производительность узлов определяется экспериментально).

Правило сопряжения 1: итоговый результат (найденные объекты, объединенный отсортированный список, файл и т.д.) вычисляется на главном узле после завершения полной обработки (поэтапно после завершения каждого процесса).

Балансировка 1: нагрузка на узлы определяется программно на основе характеристик объема исходных данных, вычислительной сложности алгоритма и производительности узлов (с учетом истории отказов).

Задание 2: сгенерировать граф, моделирующий социальную сеть (граф сетевых атак на основе моделирования действий нарушителя и атакуемых объектов). Для расчета выполнить декомпозицию графа на основе выделения групп «близких» объектов и определить K наиболее активных участников сообществ (уязвимых объектов).

Правило сопряжения 2: итоговый результат вычисляется на главном узле после завершения полной обработки и корректировки по весам ребер сопряжения (попарно после завершения каждого процесса для смежных подграфов специальным процессом для сопряжения).

Балансировка 2: нагрузка на узлы определяется программно на основе размеров подграфов и производительности узлов с учетом истории отказов (с учетом трудоемкости сопряжения).

Задание 3: реализовать численное решение задачи Дирихле для уравнения Пуассона (расчет 2D фильтрации по закону Дарси), выполнив декомпозицию области вычислений на N смежных (попарно пересекающихся) подобластей, в каждой из которых используется одинаковая (различная) однородная (неоднородная) сетка.

Правило сопряжения 3: итоговый результат вычисляется в ходе итерационного процесса на главном узле после завершения расчетов по смежным подобластям и пересчета после сопряжения решений (попарно после завершения каждого процесса для смежных подобластей специальным процессом для сопряжения).

Обеспечение отказоустойчивости (общее требование): 1) сохранение контрольных точек на локальных дисках узлов (в центральном хранилище); 2) периодическая (в зависимости от предполагаемой частоты отказов) пересылка локальных промежуточных результатов на другие узлы (только на центральный); 3) повторная полная обработка (данные заново пересылаются с центрального узла). Предполагается отсутствие отказов центрального узла и возможность (невозможность) перезапуска процесса на том же узле.

4. Заключение

При создании заданий практикума была учтена практика выполнения курсовых и выпускных квалификационных работ студентами направлений бакалавриата «Математическое обеспечение и администрирование информационных систем» и «Информационная безопасность», научных исследований магистрантов и аспирантов. Этот опыт показал, что представленный подход к разработке приложений на основе обобщенных секторных моделей позволяет поэтапно пройти путь от решений учебных задач до реализации полноценных исследовательских проектов.

Лабораторный практикум вместе с методическими рекомендациями по выполнению заданий будет доступен в электронном виде в 2016-2017 учебном году в составе учебно-методического комплекса на образовательном сервере Тюменского государственного университета (<http://www.umk3plus.utmn.ru/?section=speciality&id=1014>).

Литература

1. Dolea V., Jolivet P., Nataf F. An Introduction to Domain Decomposition Methods: algorithms, theory and parallel implementation. URL: <https://hal.archives-ouvertes.fr/cel-01100932v4> (дата обращения: 10.06.2016).
2. Gergel V., Linirov A., Meyerov I. NSF/IEEE-TCPP curriculum implementation at the State University of Nizhni Novgorod //Parallel & Distributed Processing Symposium Workshops (IPDPSW), 2014 IEEE International. IEEE, 2014. P. 1079-1084.
3. Voevodin V., Gergel V., Popova N. Challenges of a Systematic Approach to Parallel Computing and Supercomputing Education //Euro-Par 2015: Parallel Processing Workshops. Springer International Publishing, 2015. P. 90-101.
4. Антонов А.С., Воеводин В.В., Гергель В.П., Соколинский Л.Б. Системный подход к суперкомпьютерному образованию //Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. 2013. Т. 2. №. 2. С. 5-17.
5. Баркалов К.А., Мееров И.Б., Бастратов С.И. Об опыте разработки и чтения курса лекций «Параллельные численные методы» // Суперкомпьютерные дни в России: Труды международной конференции (28-29 сентября 2015 г., г. Москва). М.: Изд-во МГУ, 2015. С. 772-775.
6. Батура Т. В. Методы анализа компьютерных социальных сетей //Вестник НГУ, серия «Информационные технологии». 2012. Т. 10. №. 4. С. 13-28.
7. Гергель В.П., Баркалов К.А., Мееров И.Б., Сысоев А.В., и др. Параллельные вычисления. Технологии и численные методы: Учебное пособие в 4 томах. Н. Новгород: Изд-во Нижегородского госуниверситета, 2013. 1394 с.
8. Захарова И.Г., Захаров А.А. Интегративные возможности курса параллельного программирования // Суперкомпьютерные дни в России: Труды международной конференции (28-29 сентября 2015 г., г. Москва). М.: Изд-во МГУ, 2015. С. 759-762.
9. Котенко И.В., Степашкин М.В. Анализ защищенности компьютерных сетей на основе моделирования действий злоумышленников и построения графа атак //Труды ИСА РАН. 2007. Т. 31. С. 126-207.
10. Райгородский А.М. Модели случайных графов и их применения //Труды МФТИ. 2010. Т. 2. №. 4. С. 130 - 140.
11. Свод знаний и умений суперкомпьютерного образования. URL: <http://hpc-education.unn.ru/обучение/свод-знаний/> (дата обращения: 10.06.2016).
12. Снытников Н.В. Параллельный алгоритм для решения 2D-уравнения Пуассона в контексте нестационарных задач //Вычислительные методы и программирование. 2015. Т. 16. С. 39-51.

Methods of Sector Modeling in the Practical Work on Parallel Programming for Graduate Students

I.G. Zakharova, A.A. Zakharov

Tyumen State University

The paper presents the experience of using project tasks, based on the sectoral modeling or decomposition solutions area, in the practical work on parallel programming and distributed computing for the master's program "High Performance Computing". The tasks are provided different levels of difficulty: from integration formally processed arrays or images to conjugation sectoral hydrodynamic models. In addition to the classic tasks for parallel programming they include the study of algorithms and conjugating rules, issues of load balancing and fault tolerance.

Keywords: parallel programming, master program, sector modeling, load balancing, fault tolerance.

References

1. Dolea V., Jolivet P., Nataf F. An Introduction to Domain Decomposition Methods: algorithms, theory and parallel implementation. URL: <https://hal.archives-ouvertes.fr/cel-01100932v4> (accessed: 10.06.2016).
2. Gergel V., Linirov A., Meyerov I. NSF/IEEE-TCPP curriculum implementation at the State University of Nizhni Novgorod //Parallel & Distributed Processing Symposium Workshops (IPDPSW), 2014 IEEE International. IEEE, 2014. P. 1079-1084.
3. Voevodin V., Gergel V., Popova N. Challenges of a Systematic Approach to Parallel Computing and Supercomputing Education //Euro-Par 2015: Parallel Processing Workshops. Springer International Publishing, 2015. P. 90-101.
4. Antonov A. S., Voevodin V. V., Gergel, V. P., Sokolinskii L. B. Sistemnyj podhod k superkomp'yuternomu obrazovaniju [A Systematic Approach to Supercomputing Education]. Vestnik Yuzhno-Ural'skogo Gosudarstvennogo Universiteta. Seriya "Vychislitel'naya Matematika i Informatika" [Bulletin of South Ural State University. Series: Computational Mathematics and Informatics]. 2012. Vol. 2. No.2. P. 5-17.
5. Barkalov K.A., Meyerov I.B., Bastrakov S.I. Ob opyte razrabotki i chtenija kursa lekcij «Parallelnye chislennye metody» [On the Experience of Development and Teaching the Lecture Course "Parallel Numerical Methods"]. Superkomp'yuternye dni v Rossii: Trudy mezhdunarodnoj konferencii (28-29 sentjabrja 2015 g., g. Moskva) [Russian Supercomputing Days: Proceedings of the International Conference (September 28-29, 2015, Moscow, Russia)]. Moscow, Publishing of the Moscow State University, 2015. P. 772-775.
6. Batura T. V. Metody analiza komp'yuternyh social'nyh setej [Methods of Computer Analysis of Social Networks] Vestnik NGU, serija «Informacionnye tehnologii» [Bulletin of Novosibirsk State University. Series: Information Technology]. Vol. 10. No. 4. P. 13-28.
7. Gergel V.P., Barkalov K.A., Meyerov I.B., Sysoev A.V. et al. Parallelnye vychislenija. Tehnologii i chislennye metody: Uchebnoe posobie v 4 tomah [Parallel Computing. Technology and Numerical Analysis: A Tutorial, volumes 1-4]. Nizhni Novgorod, Publishing of the State University of Nizhni Novgorod, 2013. 1394 P.
8. Zakharova I.G., Zakharov A.A. Integrativnye vozmozhnosti kursa parallel'nogo programmirovaniya [Integrative Possibilities of the Course "Parallel Programming"]

Superkomp'juternye dni v Rossii: Trudy mezhdunarodnoj konferencii (28-29 sentjabrja 2015 g., g. Moskva) [Russian Supercomputing Days: Proceedings of the International Conference (September 28-29, 2015, Moscow, Russia)]. Moscow, Publishing of the Moscow State University, 2015. P. 759-762.

9. Kotenko I.V., Stepashkin M.V. Analiz zashhishhennosti komp'juternyh setej na osnove modelirovaniya dejstvij zloumyshlennikov i postroeniya grafa atak [Security Analysis of Computer Networks Based on the Actions of Intruders Modeling and Attack Graph]. Trudy ISA RAN [Proceedings of Institute of Systems Analysis of the Russian Academy of Sciences]. 2007. Vol. 31. P. 126-207.
10. Raigorodskii A.M. Modeli sluchajnyh grafov i ih primenenija [Models of random graphs and their applications]. Trudy MFTI [Proceedings of Moscow Institute of Physics and Technology]. 2010. Vol. 2. No. 4. P. 130-140.
11. Svod znaniy i umenij superkomp'juternogo obrazovaniya [Russian Supercomputing Education Project Curriculum]. URL: <http://hpc-education.unn> (accessed: 10.06.2016).
12. Snytnikov N.V. Parallel'nyj algoritm dlja reshenija 2D-uravnenija Puassona v kontekste nestacionarnyh zadach [A Parallel Algorithm for Solving 2D Poisson's Equation in the Context of Nonstationary Problems]. Vychislitel'nye metody i programmirovanie [Numerical Methods and Programming]. 2015. Vol. 16. P. 39-51.