РАСПРЕДЕЛЕННЫЕ ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ И ПРОБЛЕМЫ ПАРАЛЛЕЛЬНОГО МУЛЬТИПРОГРАММИРОВАНИЯ

В.Г. Хорошевский

Институт физики полупроводников им. А.В. Ржанова СО РАН

DISTRIBUTED COMPUTER SYSTEMS AND PROBLEMS OF PARALLEL MULTIPROGRAMMING

V.G. Khoroshevsky

A.V. Rzhanov Institute of Semiconductor Physics SB RAS

*Considered are hardware and software of large-scale distributed computer systems (DCS) and parallel multiprogramming modes of their functioning. N.N. Yanenko mathematical technique of complex problem solving by DCS is described. Game theory and stochastic programming approaches are studied to obtain stochastic-optimal functioning organization of these systems.*

**Введение**

Инструментарий решения и математического моделирования сложных проблем составляют высокопроизводительные средства вычислительной техники. С середины 1960-х годов концептуальную основу таких средств стала составлять модель коллектива вычислителей [1–3], за ними закрепляется термин «вычислительные системы (ВС)”. Современные суперкомпьютеры базируются на парадигме мультиархитектуры, они имеют иерархическую функциональную структуру. Любой суперкомпьютер в целом и его иерархические функционально-структурные уровни представляются как коллективы вычислителей. Следовательно, суперкомпьютеры терминологически более правильно называть суперВС.

1. Максимумом потенциальных архитектурных возможностей обладают вычислительные системы с программируемой структурой [2,3]. Концепция, теория функционирования и первые в мире проекты ВС с программируемой структурой были разработаны в СО АН CCCР. Рассматриваемые ВС диалектически сочетают в себе архитектурные свойства универсальных и специализированных средств обработки информации. В них допускается генерация (при выполнении соответствующих команд и программ) виртуальных специализированных конфигураций, адекватных структурам и характеристикам решаемых задач. Отмеченные архитектурные свойства ВС с программируемой структурой оценены математиком и механиком Н.Н. Яненко: ”Чем шире класс задач, охватываемый специализированной машиной, тем сложнее ее структура, и как наиболее совершенную форму ЭВМ следует рассматривать ЭВМ с перестраиваемой структурой ” [4].

Обобщая опыт развития индустрии обработки информации, можно заключить, что независимо от изначальной архитектурной парадигмы фирмы-создатели суперкомпьютеров к началу ХХI столетия перешли на платформу ВС с программируемой структурой. Следует отметить, что все суперВС, включенные в список Тор500, обладают программируемой структурой. Каждая очередная редакция Top500 свидетельствует о все более полном воплощении принципа программируемости структуры в суперВС.

2. Начало ХХI в. характеризуется переходом от “виртуальных метакомпьютеров” (использующих ресурсы нескольких суперВС) к большемасштабным пространственно-распределенным вычислительным системам. Пространственно-распределенная ВС – это объединение географически удаленных друг от друга сосредоточенных ВС, основанное на принципах:

– параллельности функционирования ресурсов (т.е. способности любых композиций сосредоточенных ВС работать в режимах мультипрограммирования);

– превалирующего использования массовых аппаратурно-программных средств и существующих компьютерных сетей, включая Internet.

К данному классу систем относится пространственно-распределенная мультикластерная ВС, созданная коллективом ведущей научной школы НШ-5176.2010.9 (объединенным коллективом Лаборатории вычислительных систем Института физики полупроводников им. А.В. Ржанова СО РАН и Центром параллельных вычислительных технологий ГОУ ВПО “Сибирский государственный университет телекоммуникаций и информатики – СибГУТИ”). Данная мультикластерная ВС – масштабируемая GRID-модель, ее действующая конфигурация объединяет десятки кластеров, расположенных в институтах СО РАН и в СибГУТИ. Для формирования кластеров использовались, в частности, процессоры AMD Opteron и Intel Xeon (включая двух- и четырехядерные). Мультикластерная ВС оснащена операционной системой GNU/Linux и стандартными средствами разработки параллельных программ, а также специально разработанными подсистемами: параллельного мультипрограммирования, оптимизации вложения параллельных программ в ВС, анализа MPI-программ, обеспечения отказоустойчивости выполнения параллельных программ, организации распределенной очереди задач, удаленного доступа и мониторинга ВС и др.

Пространственно-распределенная мультикластерная ВС используется не только в научных исследованиях, но и при подготовке специалистов и научных кадров высшей квалификации в области GRID и параллельных вычислительных технологий.

3. Под параллельным мультипрограммированием понимается теоретическая и практическая деятельность по созданию моделей [5], методов, алгоритмов и программ для оптимизации использования ресурсов ВС при решении задач. Параллельное мультипрограммирование следует воспринимать и как раздел математического программирования, ориентированный на распределенные вычислительные системы (включая пространственно-распределенные и массово-параллельные ВС). Мультипрограммирование реализуется при работе ВС в режимах: решения одной сложной задачи, обработки наборов задач и обслуживания потоков задач. В первом режиме задача представляется в виде параллельной программы – композиции взаимосвязанных ветвей (последовательных программ); все ресурсы ВС используются для решения задачи. При работе ВС во втором и третьем режимах одновременно решается множество задач (представленных в общем случае параллельными программами), ресурсы системы делятся между задачами.

Математические модели и методы, алгоритмы и программы, организующие работу ВС в указанных режимах составляют инструментарий параллельного мульпрограммирования. Этот инструментарий должен быть практически эффективным. Он не должен приводить к временным затратам ВС, существенно уменьшающим долю системного времени, приходящуюся на реализацию пользовательских параллельных программ. Эффективность [2] инструментария должна оцениваться такими же количественными характеристиками, как и параллельные алгоритмы (например, коэффициентами ускорения и накладных расходов).

В докладе будет рассмотрен инструментарий параллельного мульти-программирования, практически эффективный для ВС любой сложности.

4. Математическая технология Н.Н. Яненко решения сложных задач на ВС (постоянно развиваемая) ориентирована на оптимизацию вычислительной цепочки [6]: объект исследования → физическая модель → математическая модель → численный алгоритм → программа → структура вычислительной системы → расчет. Особенностью этой цепочки является то, что в ней допускаются цикличности как для цепочки в целом, так и между отдельными звеньями. Так, при разработке физических и математических моделей следует учитывать как структуру численных алгоритмов и программ, так и структуру ВС, предназначенную для реализации этих моделей. При выборе или разработке численного алгоритма должны учитываться особенности его программной реализации, структура и аппаратурно-программная конфигурация ВС. При этом структура алгоритма может определять (виртуальную) конфигурацию аппаратурно-программных средств ВС, если она имеет программируемую структуру. Таким образом, необходимо оптимизировать структуру всех звеньев вычислительной технологической цепочки с учетом их взаимного влияния, оптимизировать реализацию всей цепочки на ВС с учетом ее структуры и возможностей аппаратурно-программных средств. Такую технологию решения сложных задач на ВС следует называть глобальным программированием [6]. Важнейшей современной проблемой глобального программирования является “вложение” параллельных программ в вычислительную систему. Сложность данной проблемы обусловливается тем, что суперВС (см. Тор500) основываются на парадигме мультиархитектуры, следовательно они имеют иерархическую функциональную структуру и многоуровневую неоднородную память, характеризуются различными показателями эффективности своих подсистем и каналов связи между ними и т. п.

5. Средства вложения параллельных программ в мультиархитектурные ВС [7] основаны на многоуровневых (multilevel) алгоритмах разбиения информационных графов  параллельных программ;  – множества вершин и ребер, соответственно. Суть этих алгоритмов:

– граф  разбивается на *k* подмножеств; , ;  
– отношение числа  ветвей программы к числу  ядер, составляющих узел ВС; в каждое из подмножеств включаются ветви, обменивающиеся большими объемами данных;

– параллельные ветви из *i*-го подмножества распределяются по ядрам  
-го вычислительно узла ВС, {1, 2, …, }.

Пример вложения в 16-ядерную ВС параллельной MPI-программы Conjugate Gradient из пакета NAS Parallel Benchmarks, реализующей решение системы линейных алгебраических уравнений методом сопряженных градиентов (= 16; = 4; *k*= 4) представлен на рис.



Созданные для мультиархитектурных иерархических ВС средства вложения параллельных MPI-программ [7] сокращают время их выполнения в 1,5 – 2 раза (по сравнению с возможностями стандартных утилит).

6. Мультизадачные режимы функционирования ВС – это режимы обработки наборов задач и обслуживания потоков задач. В этих режимах ресурсы ВС распределяются между одновременно решаемыми задачами, представленными параллельными программами с различными количествами ветвей.

Режим обработки на ВС набора задач организуется тогда, когда они известны априори, и для каждой из них заданы такие параметры как ранг (в общем случае масштабируемый, число параллельных ветвей в программе), время решения и/или штраф за задержку решения на единицу времени. Режим обслуживания потока задач является наиболее общим. В нем ВС используется для обслуживания вероятностных потоков заданий, представленных параллельными программами со случайными параметрами.

Современная суперВС (в силу своей природы) относится к сложным вероятностным системам. Следовательно, эффективный инструментарий параллельного мультипрограммирования должен, безусловно, базироваться на эвристике [1, 3, 8] и стохастике [9 – 11]. Применяются два подхода к стохастической оптимизации функционирования большемасштабных и масштабируемых распределенных ВС. Оба подхода ориентированы на организацию работы ВС при обслуживании потоков параллельных задач (заданий) со случайными параметрами. Первый подход – теоретико-игровой [3, 9, 10] – применим к потокам с очередями заданий, а второй основан на методах стохастического программирования [3, 11].

Для решения “теоретико-игровых“ проблем организации функционирования большемасштабных распределенных ВС и суперВС (с массовым параллелизмом) эффективен параллельный алгоритм, основанный на композиции симплекс-метода и модифицированного метода Брауна-Робинсон. Проблемы оптимизации мульти-программирования для ВС, поставленные как задачи стохастического программирования, эффективно решаются с использованием техники динамического программирования.

Созданный стохастический инструментарий параллельного мультипрограммирования нетрудоемок, он оптимизирует загрузку ресурсов ВС множеством задач, представленных параллельными программами с произвольными количествами ветвей, и может быть положен в основу распределенных операционных систем.

**Заключение**

Современные суперВС (пространственно сосредоточенные и распределенные) полностью основываются на модели коллектива вычислителей. Для них характерно, в частности, следующее: мультиархитектурность и иерархичность функциональных структур, большемасштабность (массовый параллелизм), масштабируемость и стохастический характер ресурсов. Анализ и оптимизация функционирования суперВС относятся к числу сложных проблем, которые не могут быть решены при помощи традиционного аппарата исследования операций и методов математического программирования.

Итак, исследования в области параллельного мультипрограммирования позволяют сделать следующие выводы:

– распределенная ВС – это большемасштабный вероятностный объект, обслуживаю-щий стохастические потоки параллельных задач (в общем случае);

– техника теории игр и стохастическое программирование составляют основу для организации стохастически оптимального использования ресурсов ВС;

– стохастическая оптимизация функционирования распределенных ВС осуществляется однократно для достаточно большого интервала времени;

– параллельные алгоритмы и теоретико-игровые, и стохастического программирования реализуются эффективно на распределенных ВС;

– нет сложных вычислительных проблем при создании распределенной операционной системы, поддерживающей параллельное мультипрограммирование;

– разработанный алгоритмический и программный инструментарий вложения параллельных программ в иерархические ВС эффективнее стандартных MPI-утилит.

**Список литературы**

1. Евреинов Э.В., Хорошевский В.Г. Однородные вычислительные системы. Новосибирск: Наука, 1978.
2. Хорошевский В.Г. Архитектура вычислительных систем. М.: МГТУ им. Н.Э. Баумана, 2008.
3. Хорошевский В.Г. Распределенные вычислительные системы с программируемой структурой // Вестник СибГУТИ. 2010. № 2. С. 3 – 41.
4. Яненко Н.Н. Перспективы развития вычислительной математики на основе вычислительных систем // Препринт «ЭВМ. Перспективы и гипотезы”. Новосибирск: ИТПМ СО АН СССР, 1981. № 46. С. 3 – 6.
5. Хорошевский В.Г. Модели функционирования большемасштабных распределенных вычислительных систем // Электросвязь. 2004. № 10. С. 30 – 34.
6. Яненко Н.Н., Хорошевский В.Г., Рычков А.Д. Параллельные вычисления в задачах математической физики на вычислительных системах с программируемой структурой // Электронное моделирование. 1984. Т. 6. № 1. С. 3 – 8.
7. Хорошевский В.Г., Курносов М.Г. Алгоритмы распределения ветвей параллельных программ по процессорным ядрам вычислительных систем // Автометрия. 2008. Т. 44, № 2. С. 56 – 67.
8. Хорошевский В.Г., Седельников М.С. Эвристические алгоритмы распределения задач по машинам вычислительной системы // Автометрия. 2004. Т. 40. № 4. С. 76 – 87.
9. Хорошевский В.Г., Власюк В.В. Теоретико-игровой подход к организации стохастически оптимального функционирования распределенных вычислительных систем // Автометрия. 2000. № 3. С. 17 – 25.
10. Хорошевский В.Г., Мамойленко С.Н. Стратегии стохастически оптимального функционирования распределенных вычислительных систем // Автометрия. 2003. Т. 39. № 2. С. 81 – 91.
11. Хорошевский В.Г., Подаков М.Н. Поиск стохастически оптимального разбиения большемасштабных вычислительных систем // Автометрия. 2000. № 2. С. 52 – 59.