На правах рукописи

Пазников Алексей Александрович

АЛГОРИТМЫ ОРГАНИЗАЦИИ ФУНКЦИОНИРОВАНИЯ МУЛЬТИКЛАСТЕРНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ С ИЕРАРХИЧЕСКОЙ СТРУКТУРОЙ

Специальность 05.13.15 – Вычислительные машины, комплексы и компьютерные сети

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Новосибирск – 2013

Работа выполнена на Кафедре вычислительных систем Федерального государственного образовательного бюджетного учреждения высшего профессионального образования “Сибирский государственный университет телекоммуникаций и информатики” Федерального агентства связи.

Научный руководитель –	доктор технических наук, профессор, член-корреспондент РАН, заслуженный деятель науки РФ **Хорошевский Виктор Гаврилович**
Научный консультант –	кандидат технических наук **Курносов Михаил Георгиевич**
Официальные оппоненты:	доктор технических наук, профессор, лауреат Государственной премии СССР, старший научный сотрудник Лаборатории вычислительных технологий Федерального государственного бюджетного учреждения науки Института вычислительных технологий Сибирского отделения Российской академии наук **Рычков Александр Дмитриевич**
	доктор технических наук, заведующий Лабораторией реконфигурируемых высокопроизводительных систем Федерального государственного бюджетного образовательного учреждения высшего профессионального образования “Национальный исследовательский Томский государственный университет” **Шидловский Станислав Викторович**
Ведущая организация –	Федеральное государственное бюджетное учреждение науки Институт систем информатики им. А.П. Ершова Сибирского отделения Российской академии наук

Защита состоится “23” мая 2013 г. в 12 часов на заседании диссертационного совета Д 219.005.02 при ФГОБУ ВПО “Сибирский государственный университет телекоммуникаций и информатики”, по адресу: 630102, г. Новосибирск, ул. Кирова, 86, ком. 625.

С диссертацией можно ознакомиться в библиотеке ФГОБУ ВПО “СибГУТИ”.

Автореферат разослан “17” апреля 2013 г.

Ученый секретарь
диссертационного совета Д 219.005.02
кандидат технических наук, доцент Иван Иванович Резван

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. В настоящее время при решении сложных задач науки и техники широкое распространение получили пространственно-распределённые вычислительные системы (ВС). В архитектурном плане они представляют собой макроколлективы рассредоточенных вычислительных средств (подсистем), взаимодействующих через локальные и глобальные сети связи (включая сеть Internet). Подсистема такой ВС может быть представлена ЭВМ, вычислительным кластером или отдельной проприетарной ВС с массовым параллелизмом. К пространственно-распределённым относятся мультикластерные вычислительные и GRID-системы.

Одним из основных режимов функционирования пространственно-распределённых ВС является мультипрограммный режим обслуживания потоков параллельных задач. В этом режиме в систему (в распределённую очередь) поступает поток задач. Для решения каждой задачи требуется выделять элементарные машины (ЭМ) с одной или нескольких подсистем с целью оптимизации заданных показателей эффективности функционирования ВС.

Одним из таких показателей является время обслуживания задачи, которое включает время доставки входных и выходных данных задачи до подсистем, время ожидания в локальных очередях и время выполнения программы на ЭМ.

Актуальной является разработка моделей, методов и программного обеспечения организации функционирования пространственно-распределённых ВС. В моделях и алгоритмах должны учитываться архитектурные свойства современных ВС: большемасштабность, мультиархитектурная организация (наличие SMP, NUMA-узлов и специализированных ускорителей) и иерархическая структура коммуникационной среды.

После того, как сформирована подсистема ЭМ, необходимо оптимально вложить задачу в неё: распределить ветви по ЭМ так, чтобы минимизировать накладные расходы на межмашинные обмены информацией. Проблема вложения (Task mapping, task allocation, task assignment) в недостаточной степени проработана для пространственно-распределённых ВС, поэтому востребованы алгоритмы оптимизации вложения параллельных программ в мультикластерные и GRID-системы.

Пространственно-распределённые ВС комплектуются из неабсолютно надёжных вычислительных ресурсов (вычислительных узлов, сетевых коммутаторов, процессорных ядер и др.), поэтому немаловажной задачей является разработка средств (математических моделей, методов и программного обеспечения) организации их живучего функционирования.

Отечественные и зарубежные исследования в области распределённых ВС активно ведутся со второй половины XX столетия. Ряд фундаментальных

работ посвящен проблемам создания и эксплуатации высокопроизводительных вычислительных средств: проведены исследования по теории функционирования и построению оптимальных (макро)структур ВС, проработаны многие аспекты создания программного обеспечения, исследован широкий круг задач, допускающих эффективную реализацию на распределённых ВС. Построены отечественные вычислительные системы: “Минск-222”, СУММА, МИНИМАКС, МИКРОС, МВС, Эльбрус и др. Создана первая в мире пространственно-распределённая ВС – система АСТРА.

Фундаментальный вклад в теорию и практику вычислительных систем и параллельных вычислительных технологий внесли выдающиеся учёные, среди которых Е. П. Балашов, В. Б. Бетелин, В. С. Бурцев, В. В. Васильев, В. В. Воеводин, В. М. Глушков, В. Ф. Евдокимов, Э. В. Евреинов, А. В. Забродин, В. П. Иванников, М. Б. Игнатьев, А. В. Каляев, И. А. Каляев, Л. Н. Королев, В. Г. Лазарев, С. А. Лебедев, В. К. Левин, Г. И. Марчук, В. А. Мельников, Ю. И. Митропольский, Д. А. Поспелов, И. В. Прангишвили, Д. В. Пузанков, Г. Е. Пухов, А. Д. Рычков, Г. Г. Рябов, А. А. Самарский, В. Б. Смолов, А. Н. Томилин, Я. А. Хетагуров, В. Г. Хорошевский, Б. Н. Четверушкин, Ю. И. Шокин, Н. Н. Яненко, P. Balaji, R. Buyya, S. Cray, J. Dongarra, M. Flynn, I. Foster, A. Gara, D. Grice, W. Gropp, D. Hillis, C. Kesselman, D. L. Slotnick, R. Thakur и др.

При решении проблем оптимизации функционирования ВС в мультипрограммных режимах большую роль сыграли фундаментальные работы по исследованию операций и оптимальному управлению выдающихся ученых: В. Л. Береснева, Э. Х. Гимади, В. Т. Дементьева, С. В. Емельянова, Ю. И. Журавлева, А. А. Корбут, С. К. Коровина, Ю. С. Попкова, К. В. Рудакова, D. P. Agrawal, R. Baraglia, S. H. Bokhari, P. Bouvry, A. Gara, G. Karypis, B. W. Kernighan, V. Kumar, S. Lin, R. Perego, K. Steiglitz и др.

В диссертации предложены децентрализованные алгоритмы диспетчеризации параллельных программ в мультикластерных ВС с иерархической структурой и алгоритмы оптимизации вложения в них параллельных программ. Полученные результаты легли в основу инструментария организации функционирования мультикластерных ВС.

Цель работы и задачи исследования. Цель диссертации заключается в разработке и исследовании алгоритмов и программных средств организации функционирования мультикластерных ВС с иерархической структурой.

В соответствии с целью определены следующие задачи исследования.

1. Анализ архитектурных свойств современных пространственно-распределённых мультикластерных вычислительных и GRID-систем, методов диспетчеризации и вложения в них параллельных программ.

2. Разработка алгоритмов децентрализованной диспетчеризации в мультикластерных ВС параллельных программ с целью минимизации времени их обслуживания.

3. Создание программного инструментария децентрализованной диспетчеризации параллельных программ в мультикластерных ВС.

4. Построение алгоритмов оптимизации вложения в иерархические пространственно-распределённые ВС параллельных программ с целью минимизации времени их выполнения.

5. Реализация программного инструментария субоптимального вложения параллельных MPI-программ в мультикластерные ВС.

6. Разработка средств мониторинга производительности каналов связи и загрузки подсистем мультикластерных ВС.

Методы исследования. Для достижения цели и решения поставленных задач применялись методы теории функционирования распределённых вычислительных систем, теории множеств, теории графов, теории алгоритмов и математический аппарат исследования операций. Экспериментальные исследования проводились путём моделирования на пространственно-распределённой мультикластерной вычислительной системе.

Научная новизна работы. В диссертационной работе разработаны и исследованы алгоритмы организации функционирования мультикластерных ВС с иерархической структурой.

1. Создано семейство алгоритмов децентрализованной диспетчеризации параллельных программ. Алгоритмы учитывают переменный характер загрузки ресурсов и каналов связи пространственно-распределённых ВС и позволяют обеспечить живучее обслуживание потоков параллельных программ.

2. На основе методов разбиения графов на непересекающиеся подмножества предложены эвристические алгоритмы вложения параллельных программ в мультикластерные ВС. Алгоритмы учитывают все уровни иерархической структуры ВС, что позволяет сократить время выполнения информационных обменов в параллельных программах.

3. Выработаны рекомендации по формированию структур логических связей децентрализованных диспетчеров мультикластерных ВС. Создан эвристический алгоритм поиска субоптимальных структур локальных окрестностей диспетчеров, минимизирующий функцию штрафа при обслуживании потоков параллельных задач.

Практическая ценность работы. Разработанные в диссертации модели и алгоритмы реализованы в компонентах системного программного обеспечения мультикластерных и GRID-систем.

Предложенные алгоритмы диспетчеризации легли в основу пакета GBroker децентрализованной диспетчеризации параллельных задач в мультикластерных ВС. Применение пакета GBroker позволяет организовать жи-

вучее обслуживание потоков параллельных программ. Алгоритмы диспетчеризации характеризуются незначительной вычислительной трудоёмкостью, что обеспечивает их применимость в большемасштабных ВС.

Разработаны программные средства мониторинга производительности каналов связи и состояния вычислительных ресурсов мультикластерных ВС.

На основе эвристических алгоритмов вложения создан пакет MPIGridMap оптимизации вложения MPI-программ, позволяющий сократить время их выполнения в мультикластерных ВС. Пакет включает средства формирования информационных графов программ и оптимизации их вложения в мультикластерные ВС.

Компоненты программного обеспечения внедрены в действующую пространственно-распределённую мультикластерную ВС Центра параллельных вычислительных технологий ФГОБУ ВПО “СибГУТИ” (ЦПВТ ФГОБУ ВПО “СибГУТИ”) и Лаборатории вычислительных систем Института физики полупроводников им. А.В. Ржанова СО РАН (ИФП СО РАН).

Реализация и внедрение результатов работы. Результаты диссертационного исследования нашли применение в работах по созданию и развитию пространственно-распределённой мультикластерной ВС ЦПВТ ФГОБУ ВПО “СибГУТИ” и Лаборатории ВС ИФП СО РАН.

Исследования выполнялись в рамках федеральной целевой программы “Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 годы” (госконтракт № 07.514.11.4015 “Сверхмасштабируемые средства вложения и отказоустойчивого выполнения параллельных программ для вычислительных систем экзафлопсного уровня производительности”) и при выполнении работ по междисциплинарному интеграционному проекту СО РАН № 113 “Методы параллельной обработки данных и моделирование на распределенных вычислительных системах”. Работа поддержана грантами Российского фонда фундаментальных исследований № 12-07-31016 (научный руководитель – Пазников А.А.), 12-07-00145, 11-07-00105, 09-07-00095, 08-07-00018, грантами Президента РФ по поддержке ведущих научных школ № НШ-2175.2012.9, НШ-5176.2010.9, НШ-2121.2008.9 и грантом по Программе “У.М.Н.И.К.” Фонда содействия развитию малых форм предприятий в научно-технической сфере.

Результаты диссертации внедрены в учебный процесс. Они используются при чтении курсов лекций на Кафедре вычислительных систем ФГОБУ ВПО “СибГУТИ” по дисциплинам “Теория функционирования распределённых вычислительных систем” и “Высокопроизводительные вычислительные системы”.

Внедрение результатов диссертационных исследований подтверждено соответствующими актами.

Достоверность полученных результатов подтверждается проведёнными экспериментами и моделированием, согласованностью с данными, имеющимися в отечественной и зарубежной литературе, а также экспертизами работы, прошедшими при получении грантов.

Апробация работы. Основные результаты работы докладывались и обсуждались на международных, всероссийских и региональных научных конференциях, в том числе:

– Международной конференции “International Conference on Ubiquitous Information Management and Communication (ACM ICUIMC)” (г. Кота-Кинабалу, Малайзия, 2013);

– Международной конференции “Математические и информационные технологии (MIT)” (г. Врнячка Баня, г. Будва, Сербия, Черногория, 2011);

– Международных научных студенческих конференциях “Студент и научно-технический прогресс (МНСК)” (г. Новосибирск, 2008, 2009, 2011, 2012);

– Всероссийской научно-технической конференции “Суперкомпьютерные технологии” (с. Дивноморское Геленджикского района, 2012);

– Российской конференции с международным участием “Распределенные информационные и вычислительные ресурсы (DICR)” (г. Новосибирск, 2010);

– Российской научной конференции с участием зарубежных учёных “Моделирование систем информатики” (г. Новосибирск, 2011);

– Всероссийской конференции молодых ученых по математическому моделированию и информационным технологиям (г. Новосибирск, 2011);

– Российских конференциях “Новые информационные технологии в исследовании сложных структур (ICAM)”, (г. Томск, 2010, Алтайский Край, 2012);

– Российских научно-технических конференциях “Информатика и проблемы телекоммуникаций” (г. Новосибирск, 2008, 2009, 2010, 2011);

– Российской научно-технической конференции “Обработка информационных сигналов и математическое моделирование” (г. Новосибирск, 2012);

– Всероссийских научно-технических конференциях “Научное и технические обеспечение исследований и освоения шельфа Северного Ледовитого океана” (г. Новосибирск, 2010, 2012);

– Всероссийской научной конференции молодых учёных “Наука. Технологии. Инновации” (г. Новосибирск, 2011);

– Сибирской конференции по параллельным и высокопроизводительным вычислениям (г. Томск, 2009).

Публикации. По теме диссертации опубликовано 30 работ: 5 – в изданиях из списка ВАК, 2 свидетельства о государственной регистрации програм-

мы для ЭВМ, 23 – в материалах всероссийских и международных конференций. Результаты исследований отражены в отчётах по грантам и НИР.

Основные результаты диссертации, выносимые на защиту.

1. Алгоритмы децентрализованной диспетчеризации в пространственно-распределённых ВС параллельных программ, обеспечивающие минимизацию среднего времени их обслуживания и увеличение пропускной способности системы.

2. Программный пакет децентрализованной диспетчеризации параллельных программ в мультикластерных ВС, реализующий живучее обслуживание потоков параллельных программ.

3. Эвристические алгоритмы вложения в пространственно-распределённые ВС параллельных программ, минимизирующие время информационных обменов между параллельными ветвями.

4. Программный инструментарий оптимизации вложения параллельных MPI-программ в иерархические мультикластерные ВС.

Структура и объем диссертации. Диссертационная работа состоит из введения, четырёх глав, заключения и списка литературных источников, изложенных на 145 страницах, а также приложения на 1 странице.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении раскрыта актуальность исследования, обозначены цель и задачи диссертации, сформулированы основные положения диссертационной работы, выносимые на защиту.

В первой главе даётся понятие о распределённых ВС, описываются особенности систем с программируемой структурой, пространственно-распределённых мультикластерных и GRID-систем.

В архитектурном плане распределённая ВС представляется композицией множества элементарных машин (ЭМ) и коммуникационной сети. Конфигурация ЭМ варьируется в широких пределах – от процессорного ядра до многопроцессорного вычислительного узла (в т.ч. оснащённого GPU и/или FPGA). При выполнении параллельных программ на пространственно-распределённых ВС необходимо учитывать, что их ресурсы могут быть географически рассредоточены, а состав и загрузка подсистем, вследствие отказов или различных политик предоставления доступа, может динамически изменяться. Время доставки файлов на целевую подсистему зависит от текущей загрузки каналов связи.

Важной особенностью современных мультикластерных и GRID-систем является их большемасштабность. Например, такие современные GRID-системы, как European Grid Infrastructure, Enabling Grids for E-sciencE, Open Science Grid, NorduGrid и др., включают десятки и сотни подсистем, каждая

из которых является автономной ВС со своей системой управления ресурсами (СУР: TORQUE, SLURM, Altair PBS Pro и др.).

К наиболее значимым задачам организации функционирования пространственно-распределённых ВС в мультипрограммном режиме обслуживания потоков задач относится диспетчеризация параллельных программ (Scheduling, metascheduling, brokering): для решения поступающих в систему параллельных задач необходимо выделять элементарные машины с одной или нескольких подсистем.

Существующие средства диспетчеризации параллельных программ являются централизованными, что предполагает наличие в системе единого диспетчера, который поддерживает глобальную очередь задач. Отказ такого диспетчера может привести к выходу из строя всей ВС. Кроме того, применение централизованной диспетчеризации ограничено в большемасштабных ВС вследствие возрастания временных затрат на поиск требуемых ресурсов.

В настоящее время востребованным является создание средств децентрализованной диспетчеризации задач в пространственно-распределённых ВС. При децентрализованной диспетчеризации в системе присутствует коллектив диспетчеров, совместно принимающих решение о выборе ресурсов для задач. Это позволяет достичь живучести функционирования ВС – способности системы продолжать работу при отказах отдельных компонентов и подсистем. Каждый диспетчер взаимодействует с ограниченным числом других диспетчеров, образующих его локальную окрестность. Тем самым снижается сложность поиска ресурсов, что актуально в большемасштабных ВС.

Для современных пространственно-распределённых ВС характерна иерархическая структура коммуникационных сред. Обычно в них можно выделить минимум три уровня: первый уровень – сеть связи между подсистемами (как правило, Internet), второй уровень – сеть связи между вычислительными узлами отдельной подсистемы (технологии InfiniBand, Gigabit Ethernet), третий уровень – общая память вычислительных узлов. Уровни характеризуются различными значениями пропускной способности и латентности каналов связи. В зависимости от размещения процессорных ядер в системе, накладные расходы на передачу информации между ними существенно различаются.

Время выполнения параллельных программ на распределённых ВС в значительной степени зависит от того, насколько эффективно они вложены в систему. Под эффективным вложением (Task mapping, task allocation, task assignment) понимается такое распределение ветвей параллельной программы по процессорным ядрам ВС, при котором достигается минимум накладных расходов на межмашинные обмены.

Существующие методы вложения параллельных программ применимы и в пространственно-распределённых ВС, но они не учитывают все иерархические уровни коммуникационной среды и могут не обеспечивать предельной

эффективности использования ресурсов ВС. Кроме того, некоторые алгоритмы ориентированы на оптимизацию вложения в системы, имеющие определённые структуры (например, тороидальную или гиперкубическую), другие направлены только на вложение задач определённого класса. При оптимизации вложений в целевых функциях не учитывается возможность интенсивных обменов сообщениями небольшого размера, что характерно для программ на языках семейства Partitioned Global Address Space (PGAS).

Существующие СУР и библиотеки MPI реализуют алгоритмы вложения параллельных программ с учётом только двух уровней коммуникационной сети – сети межузловых связей и общей памяти вычислительных узлов. В пространственно-распределённых ВС особенно важно учитывать наличие медленных каналов связи между подсистемами, которые существенно влияют на время выполнения параллельных программ.

Во второй главе описана модель пространственно-распределённой ВС, рассматривается задача диспетчеризации параллельных программ с целью минимизации времени их обслуживания, предложены алгоритмы децентрализованной диспетчеризации параллельных программ.

Рассмотрим пространственно-распределённую ВС, состоящую из H подсистем, объединённых каналами связи и укомплектованную N ЭМ. Под ЭМ понимается единица вычислительного ресурса, предназначенного для выполнения ветви параллельной программы (как правило, процессорное ядро). Приняты следующие обозначения: n_i – количество ЭМ в составе подсистемы $i \in S = \{1, 2, \ldots, H\}$; c_i – число ЭМ подсистемы i, не задействованных для решения задач; q_i – длина локальной очереди задач подсистемы i; s_i – число задач, запущенных на подсистеме i; $t_{ij} = t(i, j, m)$ – время передачи сообщения размером m байт между подсистемами $i, j \in S$ ($[t(i, j, m)] = $ с). Время доставки может быть получено на основе различных аналитических моделей (LogP, LogGP, Hockney) оценки времени выполнения информационных обменов в параллельных программах.

На подсистемах функционируют локальная СУР и децентрализованный диспетчер. Коллектив диспетчеров представлен ориентированным графом $G(S, E)$, в котором вершинам соответствуют диспетчеры, а рёбрам – логические связи между ними (рис. 1). Наличие дуги $(i, j) \in E$ означает, что диспетчер i может отправлять задачи из своей очереди диспетчеру j. Локальная окрестность вершины i образована вершинами j, ей смежными: $L(i) = \{j \in S \mid (i, j) \in E\}$.

При работе с системой пользователь отправляет ресурсный запрос любому диспетчеру i. Ресурсный запрос включает в себя параллельную программу ранга r (количество параллельных ветвей), входные файлы, размеры $z_1, z_2, \ldots, z_k$ файла программы и входных данных ($[z_l]$ = байт), а также номера

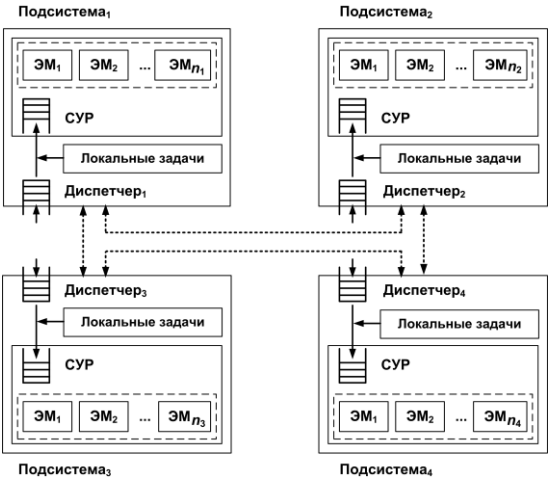


Рис. 1. Пример локальных окрестностей диспетчеров
$H = 4$, $L(1) = \{2, 3\}$, $L(2) = \{1, 4\}$, $L(3) = \{1, 4\}$, $L(4) = \{2, 3\}$

$h_1, h_2, \ldots, h_k$ подсистем, на которых размещены соответствующие файлы ($h_l \in S$). Согласно реализованным алгоритмам, диспетчер i назначает (суб)оптимальную подсистему $j^* \in L(i) \cup \{i\}$ (или подсистемы $j_1^*, j_2^*, \ldots, j_m^*$) из его локальной окрестности для выполнения программы.

Разработано семейство алгоритмов децентрализованной диспетчеризации параллельных задач в мультикластерных ВС. Каждый алгоритм задаёт поведение диспетчера i при поступлении задачи в его очередь.

На первом шаге все алгоритмы предполагают обращение диспетчера i к системе мониторинга и получение текущих значений параметров t_{ij}, c_j, s_j, q_j и n_j ($j \in L(i) \cup \{i\}$). Затем строится множество допустимых подсистем $S(i) = \{j \mid n_j \geq r, j \in L(i) \cup \{i\}\}$, имеющих число ЭМ не ниже требуемого.

Алгоритм локально-оптимальной диспетчеризации (ДЛО) осуществляет выбор локально-оптимальной подсистемы.

Шаг 1. В окрестности $S(i)$ диспетчера i выбирается подсистема j^* с минимальным значением функции $F(j)$, $j \in S(i)$:

$$j^* = \arg\min_{j \in S(i)} F(j), \qquad F(j) = \begin{cases} \dfrac{t_j}{t_{\max}} + \dfrac{c_{\max}}{c_j} + \dfrac{w_j}{w_{\max}}, & \text{если } c_j < r \text{ или } q_j > 0, \\ \dfrac{t_j}{t_{\max}}, & \text{иначе,} \end{cases}$$

где $t_j = \sum_{l=1}^{k} t(h_l, j, z_l)$ – время доставки файлов задачи до подсистемы j; $t_{\max} = \max_{j \in S(i)}\{t_j\}$; $c_{\max} = \max_{j \in S(i)}\{c_j\}$; $w_j = q_j / n_j$ – количество задач в очереди, приходящееся на одну ЭМ подсистемы j; $w_{\max} = \max_{j \in S(i)}\{w_j\}$. Целевая функция $F(j)$ учитывает время доставки данных задачи до подсистем, а также их относительную загруженность.

Шаг 2. Задача направляется в очередь локальной СУР подсистемы j^*, после чего осуществляется доставка файлов программы на эту подсистему.

Алгоритм на основе репликации задач (ДР) реализует назначение задачи одновременно на несколько подсистем.

Шаг 1. Из окрестности $S(i)$ выбирается m подсистем $j_1^*, j_2^*, \ldots, j_m^*$ в порядке неубывания значений функции $F(j)$.

Шаг 2. Задача ставится в очередь подсистем $j_1^*, j_2^*, \ldots, j_m^*$, выполняется доставка данных до этих подсистем.

Шаг 3. С интервалом времени Δ_1 диспетчер i проверяет состояние задачи на подсистемах $j_1^*, j_2^*, \ldots, j_m^*$ и определяет подсистему j', на которой задача запущена на выполнение раньше других подсистем.

Шаг 4. Задача удаляется из очередей СУР подсистем, отличных от j'.

Алгоритм диспетчеризации на основе миграции задач (ДМ) реализует периодический поиск новых подсистем для задач из очереди диспетчера.

Шаг 1. Выбирается локально-оптимальная подсистема $j^* \in S(i)$.

Шаг 2. Задача направляется в очередь СУР подсистемы j^*.

Шаг 3. Диспетчер i с интервалом времени Δ_2 запускает процедуру поиска новой подсистемы j'.

Шаг 4. Если для найденной подсистемы выполняется $F(j^*) - F(j') > \varepsilon$, то задача удаляется из очереди диспетчера j^* и мигрирует на подсистему j'.

Алгоритм на основе репликации и миграции задач (ДРМ) реализует комбинацию двух подходов – назначения задачи на несколько подсистем и миграции из очереди.

Вычислительная сложность алгоритмов ДЛО, ДР, ДМ, ДРМ не зависит от числа H подсистем. Поиск подсистем выполняется только в пределах локальных окрестностей диспетчеров и имеет вычислительную сложность $T = O(|S(i)|\Delta t)$, где Δt – время получения информации о производительности одной подсистемы и времени доставки данных.

Разработанные алгоритмы ДЛО, ДР, ДМ, ДРМ реализованы в программном пакете GBroker децентрализованной диспетчеризации параллельных программ. Натурное моделирование алгоритмов проводилось на пространственно-распределённой мультикластерной ВС, созданной ЦПВТ ФГОБУ

ВПО “СибГУТИ” совместно с Лабораторией ВС ИФП СО РАН (рис. 2). На сегментах системы установлена операционная система GNU/Linux, локальная СУР TORQUE 2.3.7, пакет Globus Toolkit 5.0 и компоненты пакета GBroker. В качестве тестовых задач использовались MPI-программы из пакета SPEC MPI 2007: Weather Research and Forecasting (WRF) – пакет моделирования климатических процессов; The Parallel Ocean Program (POP) – пакет моделирования процессов в океане; LAMMPS – пакет решения задач молекулярной динамики; RAxML – пакет моделирования задач биоинформатики; Tachyon – пакет расчета графических сцен. Входные данные для тестовых задач размещались на сегменте Xeon80; на эту же подсистему доставлялись результаты выполнения программ. Для оценки эффективности алгоритмов диспетчеризации использовались следующие показатели: пропускная способность B системы, среднее время T обслуживания задачи и среднее время W пребывания задачи в очереди диспетчера и СУР.

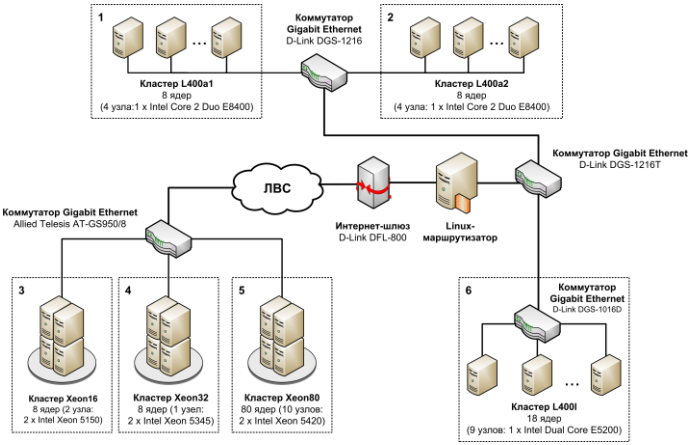


Рис. 2. Тестовая конфигурация мультикластерной ВС ($H = 6$, $N = 130$)

Пропускная способность системы при использовании алгоритма ДР для $m \in \{2, 3\}$ выше пропускной способности, полученной при использовании алгоритма ДЛО (рис. 3). Деградация показателей с ростом m объясняется увеличением загрузки каналов связи при передаче входных файлов одной задачи на несколько сегментов. Наименьшие среднее время обслуживания задач и среднее время ожидания в очереди достигнуты при использовании алгоритмов ДЛО и ДМ, причём большая пропускная способность системы получена для ДМ. Алгоритмы ДР и ДРМ рекомендуется применять при малой интенсивности потоков задач или небольших размерах входных данных.

Выполнено сравнение эффективности обслуживания потоков задач централизованным диспетчером GridWay и разработанным децентрализованным

пакетом GBroker. Пропускная способность диспетчера GBroker при обслуживании потока задач высокой интенсивности превосходит пропускную способность пакета GridWay на 10-15%. Среднее время обслуживания и среднее время пребывания задач в очереди близки с GridWay и незначительно возрастают в случае централизованного обслуживания.

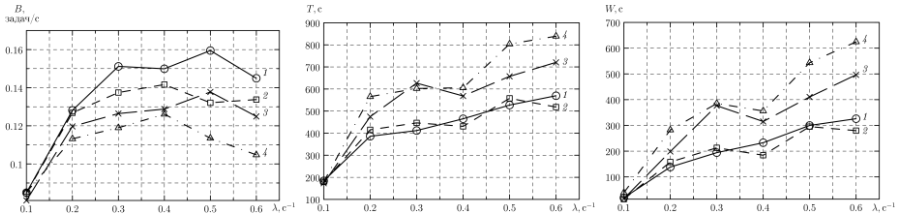


Рис. 3. Сравнение эффективности алгоритмов ДЛО, ДМ и ДРМ:
1 – алгоритм ДМ; 2 – алгоритм ДЛО;
3 – алгоритм ДРМ, $m = 2$; 4 – алгоритм ДРМ, $m = 3$

Проведено исследование влияния выбора логических структур локальных окрестностей диспетчеров на эффективность диспетчеризации. Высокие значения пропускной способности (при сопоставимых значениях T и W), помимо полносвязного графа, были получены для конфигураций на основе решётки, $2D$-тора и D_2-графа, при этом для двух последних достигнуты наибольшие значения. Использование неполносвязных логических структур при формировании локальных окрестностей диспетчеров не приводит к снижению показателей эффективности диспетчеризации.

На основе метода цепей Монте-Карло предложен эвристический алгоритм поиска субоптимальных логических структур локальных окрестностей децентрализованных диспетчеров. Критерий оптимизации (штраф при обслуживании потока параллельных задач) учитывает стоимость использования ресурсов, интенсивности потоков поступления задач от пользователей и миграции задач между подсистемами. Имитационное моделирование показало, что логические структуры, полученные с помощью разработанного алгоритма, обеспечивает уменьшение в 2,5 раза значения целевой функции по сравнению с известными перспективными структурами.

При диспетчеризации мультикластерных ВС возникает острая необходимость в мониторинге производительности каналов связи. Предложен алгоритм оценки времени доставки файлов между подсистемами. Алгоритм предполагает периодическое измерение времени передачи файлов различных размеров между подсистемами и заполнение таблицы измерений. Прогноз времени доставки файла размера z байт рассчитывается по значениям времени передачи пакетом GridFTP путём интерполяции. Такой подход позволяет учитывать особенности реализации службы GridFTP. Относительное откло-

нение прогнозируемого времени передачи файлов от значений GridFTP является приемлемым при диспетчеризации в мультикластерных и GRID-системах.

В третьей главе описана математическая модель коммуникационных сред мультикластерных ВС с иерархической структурой, рассматриваются созданные алгоритмы вложения параллельных программ в такие системы.

Пусть ВС состоит из H подсистем. Коммуникационная среда системы имеет иерархическую организацию и может быть представлена в виде дерева, содержащего L уровней.

Для параллельной программы ранга N выделяется подсистема из N ЭМ (на рис. 4 обозначена серым цветом). Подсистема также имеет иерархическую структуру. Обозначим n_l – число элементов на уровне $l \in \{1, 2, \ldots, L\}$; n_{lk} – количество прямых дочерних узлов элемента $k \in \{1, 2, \ldots, n_l\}$, находящегося на уровне l; c_{lk} – количество ЭМ, принадлежащих потомкам этого элемента.

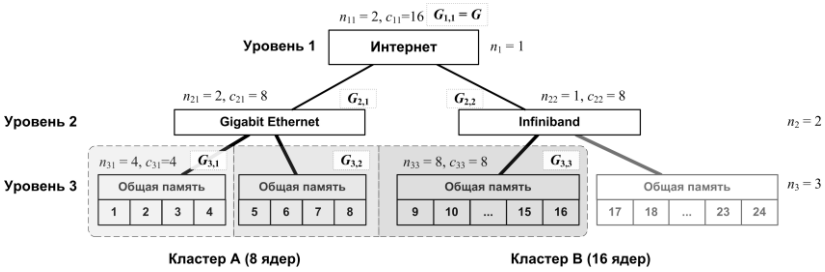


Рис. 4. Пример подсистемы $N = 16$ ЭМ для решения параллельной задачи

Параллельная программа, созданная в модели передачи сообщений, представлена информационным графом $G = (V, E)$, где $V = \{1, 2, \ldots, N\}$ – множество ветвей параллельной программы, а $E \subseteq V \times V$ – множество информационно-логических связей между ветвями. Обозначим через d_{ij} вес ребра $(i, j) \in E$, характеризующий интенсивность обменов данными между ветвями i и j при выполнении программы.

Вложение в ВС задаётся значениями переменных $x_{ij} \in \{0, 1\}$: $x_{ij} = 1$, если $i \in V$ назначена на процессорное ядро $j \in \{1, 2, \ldots, N\}$, иначе $x_{ij} = 0$.

В качестве критерия оценки эффективности вложения использовалось время T выполнения информационных обменов. Оно определяется максимальным из времён выполнения обменов ветвями программы. Требуется построить вложение X, доставляющее минимум T:

$$T(X) = \max_{i \in V} \left\{ \sum_{j=1}^{N} \sum_{p=1}^{N} \sum_{q=1}^{N} x_{ip} x_{jq} t(i, j, p, q) \right\} \to \min_{(x_{ij})} \qquad (1)$$

при ограничениях

$$\sum_{j=1}^{N} x_{ij} = 1, \quad i = 1,2,\ldots,N, \tag{2}$$

$$\sum_{i=1}^{N} x_{ij} = 1, \quad j = 1,2,\ldots,N, \tag{3}$$

$$x_{ij} \in \{0,1\}, \quad i \in V, j \in \{1,2,\ldots,N\}. \tag{4}$$

Ограничения (2), (4) гарантируют назначение каждой ветви параллельной программы на единственную ЭМ. Ограничение (3) обеспечивает назначение на машину не более одной ветви.

Задача (1)-(4) относится к дискретной оптимизации и является трудноразрешимой. Предложен метод *HierarchicMap* её приближённого решения. Метод основан на разбиении графа задачи на подмножества интенсивно обменивающихся параллельных ветвей и вложения их в ЭМ, связанные быстрыми каналами связи. Разбиение выполняется многократно: для каждого уровня иерархии коммуникационной среды.

Суть метода в следующем. На вход алгоритма подаётся граф G_{lk}. В начале алгоритма полагаем $l = 1$, $k = 1$.

Шаг 1. Если текущий уровень равен L, выполнение алгоритма завершается. В противном случае, граф G_{lk} разбивается на n_{lk} частей $G_{l+1,1}, G_{l+1,2}, \ldots, G_{l+1,n_{lk}}$ по $c_{l+1,1}, c_{l+1,1}, \ldots, c_{l+1,n_{lk}}$ вершин.

Шаг 2. Для каждого из подграфов $G_{l+1,1}, G_{l+1,2}, ..., G_{l+1,n_{lk}}$ выполняется процедура разбиения в соответствии с шагом 1.

На основе метода *HierarchicMap* построены алгоритмы, различающиеся между собой учётом уровней коммуникационной среды при формировании разбиения. *L1Map* – алгоритм, учитывающий только уровень $l = 1$ связи между подсистемами и не учитывающий уровень связи между узлами, *L2Map* – алгоритм, учитывающий только уровень $l = 2$ связи между узлами и не учитывающий уровень связи между подсистемами, *L12Map* – алгоритм, учитывающий как уровень связи между подсистемами, так и уровень связи между узлами, и т.д.

Трудоёмкость алгоритмов вложения определяется количеством выполнений процедуры разбиения графа. При использовании многоуровневых алгоритмов разбиения графов вычислительная сложность данной процедуры составляет $T = O(|E|\log_2 z)$, где $|E|$ – количество рёбер в графе, z – число подмножеств разбиения графа. Разбиение выполняется для всех элементов $k = 1, \ldots, n_l$ каждого уровня $l = 1, \ldots, L - 1$.

Моделирование алгоритмов вложения проводилось на мультикластерной ВС ЦПВТ ФГОБУ ВПО “СибГУТИ” и Лаборатории ВС ИФП СО РАН. Для

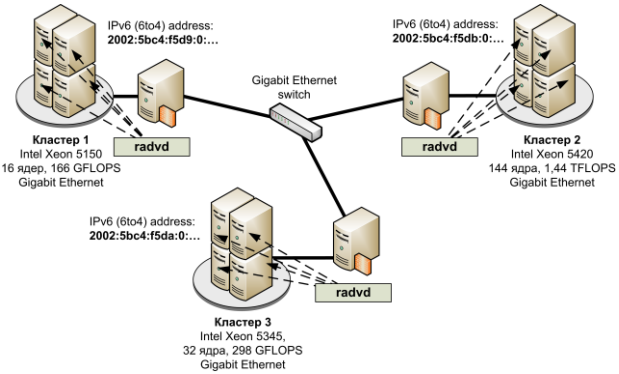


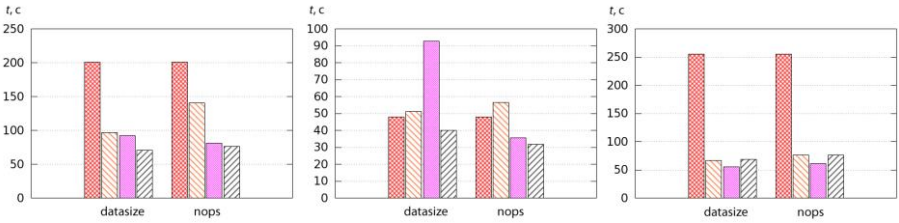
Рис. 5. Конфигурация тестовой подсистемы мультикластерной ВС

запуска MPI-программ на ресурсах нескольких подсистем применялся подход на основе межсетевого протокола IPv6 (рис. 5).

Использовались тестовые MPI-программы POP, SWEEP3d, GRAPH500 и тесты LU, SP, MG, BT из пакета NAS Parallel Benchmarks. Формирование вложений выполнялось при помощи библиотек Scotch, METIS и gpart разбиения графов. Заметим, что библиотека Scotch используется в пакете hwloc при вложении параллельных MPI-программ.

Рассмотрено два способа задания весов рёбер: 1) d_{ij} – суммарный объём данных, передаваемых между ветвями i и j за время выполнения программы ([d_{ij}] = байт), 2) d_{ij} – количество переданных сообщений между ветвями i и j.

Время выполнения MPI-программ при вложении их алгоритмом, учитывающим все уровни коммуникационной среды ВС, на некоторых задачах от 1,1 до 5 раз ниже по сравнению с линейным вложением (реализуется по умолчанию библиотеками MPI) (рис. 6). Алгоритмы вложения эффективны для параллельных задач, имеющих разреженные информационные графы с преобладанием дифференцированных MPI-обменов. В таких программах можно выделить группы интенсивно взаимодействующих параллельных ветвей и распределить их по ЭМ, соединённым быстрыми каналами связи. Время работы алгоритмов на одном процессоре не превышает 1 с. Выбор способа формирования информационного графа зависит от типа коммуникационных взаимодействий в параллельной программе: если преобладают частые обмены сообщениями небольшого размера (характерно для PGAS-программ), рекомендуется учитывать количество переданных сообщений, в остальных случаях – объём сообщений.



а *б* *в*

Рис. 6. Сравнение алгоритмов вложения параллельных программ
а – POP, *N* = 120, *б* – Sweep3D, *N* = 120, *в* – NPB MG, *N* = 64
– линейное вложение, – *L1Map*, – *L2Map*, – *L12Map*

Сравнение пакетов METIS, Scotch, gpart разбиения графов показало, что на большинстве тестовых программ все библиотеки дают сопоставимые результаты. Применение библиотеки Scotch в некоторых случаях (POP) обеспечивает незначительное сокращение времени выполнения программы.

Полученные результаты по вложению параллельных программ могут быть использованы в таких пакетах, как MPICH2, Open MPI, GridMPI, PACX, NumGRID, а также в runtime системах языков семейства PGAS (Cray Chapel, IBM X10).

В четвёртой главе описана архитектура и программное обеспечение мультикластерной ВС (рис. 7), разработанной при непосредственном участии диссертанта.

Система включает 10 пространственно-распределённых сегментов, представленных кластерными ВС. Кластеры A-H расположены в ЦПВТ ФГОБУ ВПО “СибГУТИ”, а кластеры I, J – в Лаборатории ВС ИФП СО РАН.

Коммуникационная среда мультикластерной ВС построена на основе локальных сетей (InfiniBand QDR, Gigabit Ethernet), и глобальной сети Internet (технология VPN). Связь осуществляется через выделенные серверы сегментов ЦПВТ ФГОБУ ВПО “СибГУТИ” и Лаборатории ВС ИФП СО РАН. Система включает более 300 процессорных ядер и имеет пиковую производительность несколько TeraFLOPS. Мультикластерная ВС допускает масштабирование путём организации взаимодействия с другими системами.

Основу программного обеспечения мультикластерной ВС составляет программный комплекс Globus Toolkit, который реализует интерфейс с локальными СУР подсистем (GRAM) и осуществляет передачу файлов (GridFTP) между подсистемами. В качестве СУР применяются TORQUE и SLURM.

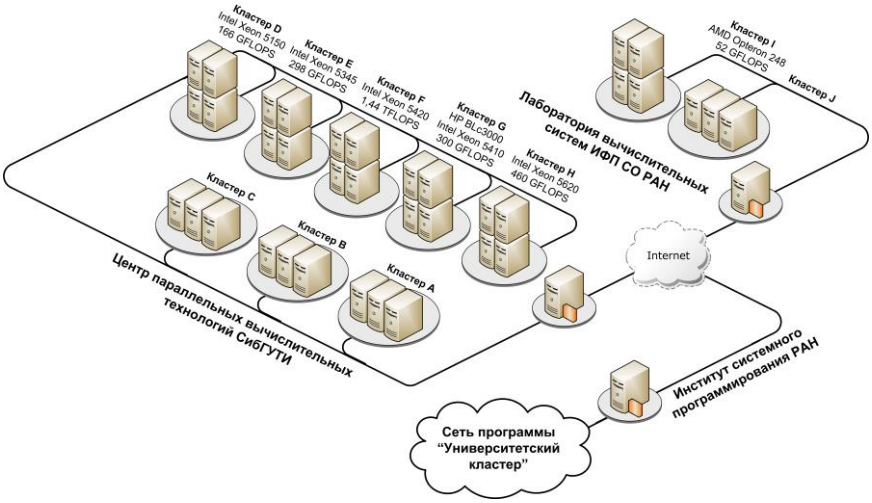


Рис. 7. Конфигурация пространственно-распределённой мультикластерной ВС

Предложенные алгоритмы диспетчеризации реализованы в программном пакете GBroker децентрализованной диспетчеризации параллельных задач в мультикластерных ВС. Пакет разрабатывается ЦПВТ ФГОБУ ВПО "СибГУТИ" совместно с Лабораторией ВС ИФП СО РАН. Он включает в себя диспетчер GBroker, модуль GClient интерфейса и системы мониторинга NetMon и DCSMon производительности каналов связи и загрузки подсистем. Все модули установлены на сегментах мультикластерной ВС. Администратор задаёт локальные окрестности диспетчеров и систем мониторинга.

Предложенные алгоритмы вложения реализованы в программном пакете MPIGridMap оптимизации вложения MPI-программ в мультикластерные ВС. Он включает в себя средство OTFParse анализа протокола выполнения программ и модуль MPIGridMap формирования вложения MPI-программ.

Созданный программный инструментарий позволяет повысить эффективность эксплуатации мультикластерных ВС и минимизировать время выполнения на них параллельных программ.

В заключении изложены основные результаты, полученные в диссертационной работе.

В приложениях приведено описание структурной организации сегментов мультикластерной ВС.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

Разработаны и исследованы алгоритмы и программные средства организации функционирования пространственно-распределённых мультикластерных ВС с иерархической структурой.

1. Предложены алгоритмы децентрализованной диспетчеризации параллельных программ в мультикластерных ВС, позволяющие сократить среднее время обслуживания задач и повысить пропускную способность системы. В алгоритмах учитывается текущая загрузка подсистем и каналов связи. Построенные алгоритмы не менее эффективны, чем централизованные, при этом позволяют обеспечить живучее обслуживание потоков параллельных программ.

2. Созданные алгоритмы составили основу программного инструментария децентрализованной диспетчеризации параллельных программ в мультикластерных ВС. Инструментарий включает в себя клиентскую программу, диспетчер и средства мониторинга состояния каналов связи и подсистем.

3. Разработаны эвристические алгоритмы вложения параллельных программ в мультикластерные ВС с иерархической структурой. Алгоритмы учитывают все иерархические уровни мультикластерных ВС, что позволяет сократить время выполнения параллельных программ от 1,1 до 5 раз по сравнению с вложением стандартными средствами. Алгоритмы применимы для программ, имеющих разреженные информационные графы с преобладанием дифференцированных обменов.

4. Разработан программный пакет вложения параллельных задач в мультикластерные ВС с иерархической структурой. Пакет позволяет получать информационные графы параллельных программ и вкладывать их в мультикластерные ВС в соответствии с предложенными алгоритмами.

5. При участии диссертанта создана мультикластерная ВС, конфигурация которой расширена инструментарием децентрализованной диспетчеризации параллельных программ и средствами их субоптимального вложения.

Основные результаты диссертации опубликованы в работах [1-30].

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в журналах из списка ВАК

1. Курносов, М.Г. Эвристические алгоритмы отображения параллельных MPI-программ на мультикластерные вычислительные и GRID-системы / М.Г. Курносов, **А.А. Пазников** // Вычислительные методы и программирование. – 2013. – № 14. – С. 1-10.

2. Курносов, М.Г. Децентрализованные алгоритмы диспетчеризации пространственно-распределённых вычислительных систем / М.Г. Курносов, **А.А. Пазников** // Вестник ТГУ. Управление, вычислительная техника и информатика. – 2012. – № 1 (18). – С. 133-143.

3. Хорошевский, В.Г. Масштабируемый инструментарий параллельного мультипрограммирования пространственно-распределенных вычислительных систем / В.Г. Хорошевский, М.Г. Курносов, С.Н. Мамойленко, К.В. Павский, А.В. Ефимов, **А.А. Пазников**, Е.Н. Перышкова // Вестник СибГУТИ. – 2011. – № 4. – С. 3-19.

4. Курносов, М.Г. Инструментарий децентрализованного обслуживания потоков параллельных MPI-задач в пространственно-распределенных мультикластерных вычислительных системах / М.Г. Курносов, **А.А. Пазников** // Вестник ТГУ. Управление, вычислительная техника и информатика. – 2011. – № 3 (16). – С. 78-85.

5. Курносов, М.Г. Программный пакет децентрализованного обслуживания потоков параллельных задач в пространственно-распределенных вычислительных системах / М.Г. Курносов, **А.А. Пазников** // Вестник СибГУТИ. – 2010. – № 2 (10). – С. 79-86.

Авторские свидетельства

6. Свид. 2012660249 Российская Федерация. Свидетельство о государственно регистрации программы для ЭВМ. Программа децентрализованной диспетчеризации параллельных задач в пространственно-распределённых вычислительных системах / **Пазников А.А.**, Курносов М.Г. Заявитель и патентообладатель ФГОБУ ВПО “СибГУТИ”. Заявл. 23.07.2012, опубл. 14.11.2012.

7. Свид. 2012613763 Российская Федерация. Свидетельство о государственно регистрации программы для ЭВМ. Средства вложения и отказоустойчивого выполнения параллельных программ для вычислительных систем экзафлопсного уровня производительности / Хорошевский В.Г., Курносов М.Г., Молдованова О.В., **Пазников А.А.**, Поляков А.Ю., Павский К.В., Мамойленко С.Н. Заявитель и патентообладатель ФГОБУ ВПО “СибГУТИ”. Заявл. 05.03.2012, опубл. 20.04.2012.

Публикации в журналах и материалах конференций

8. Kurnosov, M.G. Efficiency analysis of decentralized grid scheduling with job migration and replication / M.G. Kurnosov, **A.A. Paznikov** // Proc. of ACM International Conference on Ubiquitous Information Management and Communication (IMCOM/ICUIMC). – 2013. – 7 p.

9. Курносов, М.Г. Моделирование алгоритмов децентрализованного обслуживания потоков параллельных задач в GRID-системах / М.Г. Курносов, **А.А. Пазников** // Проблемы информатики. – 2012. - № 2. – С. 45-54.

10. Курносов, М.Г. Вложение параллельных программ в пространственно-распределённые вычислительные системы на основе методов разбиения графов / М.Г. Курносов, **А.А. Пазников** // Материалы 2-й Всероссийской научно-технической конференции “Суперкомпьютерные технологии” (СКТ-2012). – Ростов-на-Дону: Издательство Южного федерального университета, 2012. – С. 135-139.

11. Курносов, М.Г. Эвристические алгоритмы вложения параллельных MPI-программ в мультикластерные вычислительные системы / М.Г. Курносов, **А.А. Пазников** // Материалы

Девятой российской конференции с международным участием “Новые информационные технологии в исследовании сложных структур”. – Томск: НТЛ, 2012. – С. 10.

12. **Пазников, А.А.** Эвристические алгоритмы вложения параллельных MPI-программ в мультикластерные вычислительные системы / А.А. Пазников // Материалы 50-й Международной научной студенческой конференции “Студент и научно-технический прогресс”: Программирование и вычислительные системы. – Новосибирск: НГУ, 2012. – С. 37.

13. Курносов, М.Г. Сравнительный анализ методов вложения параллельных MPI-программ в мультикластерные вычислительные системы / М.Г. Курносов, **А.А. Пазников** // Материалы Российской научно-технической конференции “Обработка информационных сигналов и математическое моделирование”. – Новосибирск: СибГУТИ, 2012. – С. 160-162.

14. **Пазников, А.А.** Стохастический алгоритм формирования локальных окрестностей децентрализованных диспетчеров мультикластерных систем / А.А. Пазников // Материалы Российской научно-технической конференции “Обработка информационных сигналов и математическое моделирование”. – Новосибирск: СибГУТИ, 2012. – С. 167-168.

15. Курносов, М.Г. Применение многоуровневых методов разбиения графов параллельных программ для оптимизации их вложения в мультикластерные вычислительные системы / М.Г. Курносов, **А.А. Пазников** // Материалы Второй всероссийской научно-технической конференции “Научное и технические обеспечение исследований и освоения шельфа Северного Ледовитого океана”, 2012. – С. 109-113.

16. Курносов, М.Г. Децентрализованные алгоритмы управления ресурсами распределенных вычислительных и GRID-систем / М.Г. Курносов, **А.А. Пазников** // Материалы Международной конференции “Математические и информационные технологии, MIT-2011”. – Сербия, 2011. – 6с.

17. Курносов, М.Г. Моделирование алгоритмов децентрализованного обслуживания потоков параллельных задач в GRID-системах / М.Г. Курносов, **А.А. Пазников** // Материалы Российской научной конференции с участием зарубежных исследователей “Моделирование систем информатики”. – Новосибирск, 2011. – 10 с.

18. Курносов, М.Г. Исследование алгоритмов диспетчеризации задач в пространственно-распределенных вычислительных системах / М.Г. Курносов, **А.А. Пазников** // Материалы Российской научно-технической конференции “Информатика и проблемы телекоммуникаций”. – Новосибирск: СибГУТИ, 2011. – Т. 1. – С. 199-200.

19. **Пазников, А.А.** Анализ алгоритмов диспетчеризации параллельных программ в пространственно-распределённых вычислительных системах / А.А. Пазников // Материалы XLIX Международной научной студенческой конференции “Студент и научно-технический прогресс”. – Новосибирск: НГУ, 2011. – С. 228.

20. **Пазников, А.А.** Алгоритмы диспетчеризации мультикластерных вычислительных систем на основе миграции и репликации параллельных MPI-программ / А.А. Пазников // Материалы всероссийской научной конференции молодых учёных “Наука. Технологии. Инновации”. – Новосибирск: НГТУ, 2011. – Т. 1. – С. 63-66.

21. Курносов, М.Г. Моделирование алгоритмов децентрализованной диспетчеризации параллельных задач в пространственно-распределенных мультикластерных вычислительных системах / М.Г. Курносов, **А.А. Пазников** // Материалы XIII Российской конференции с участием иностранных ученых “Распределенные информационно-вычислительные ресурсы” (DICR-2010). – Новосибирск: ИВТ СО РАН, 2010. – 7 с.

22. Курносов, М.Г. Инструментарий организации эффективного выполнения параллельных программ на распределенных вычислительных системах / М.Г. Курносов, **А.А. Пазников**, Ю.Е. Макарова, В.В. Апостолов // Материалы Всероссийской научно-технической конференции “Научное и технические обеспечение исследований и освоения шельфа Северного Ледовитого океана”, 2010. – С. 49-53.

23. Курносов, М.Г. Об организации функционирования пространственно-распределенных мультикластерных вычислительных систем / М.Г. Курносов, **А.А. Пазников** // Материалы Российской научно-технической конференции “Информатика и проблемы телекоммуникаций”. – Новосибирск: СибГУТИ, 2010. – Т. 1. – С. 159-161.

24. Курносов, М.Г. Децентрализованная диспетчеризация параллельных программ в распределенных вычислительных системах / М.Г. Курносов, **А.А. Пазников** // Материалы Девятой международной конференции “Высокопроизводительные параллельные вычисления на кластерных системах”. – Владимир: ВлГУ, 2009. – С. 260-265.

25. **Пазников, А.А.** Алгоритмы децентрализованной диспетчеризации параллельных программ в пространственно-распределённых вычислительных системах / А.А. Пазников // Материалы Пятой сибирской конференции по параллельным вычислениям. – Томск, 2009. – С. 161-165.

26. **Пазников, А.А.** Средства децентрализованной диспетчеризации задач в распределенных вычислительных системах / А.А. Пазников // Материалы XLVII Международной научной студенческой конференции “Студент и научно-технический прогресс”. – Новосибирск: НГУ, 2009. – С. 210.

27. Курносов, М.Г. Диспетчеризация параллельных задач в пространственно-распределенных вычислительных системах / М.Г. Курносов, **А.А. Пазников** // Материалы Российской научно-технической конференции “Информатика и проблемы телекоммуникаций”. – Новосибирск: СибГУТИ, 2009. – Т.1 – С. 125-127.

28. **Пазников, А.А.** Комбинаторный алгоритм вложения параллельных программ в вычислительные системы / А.А. Пазников // Материалы XLVI Международной научной студенческой конференции “Студент и научно-технический прогресс”. – Новосибирск: НГУ, 2008. – С. 218.

29. **Пазников, А.А.** Точный алгоритм вложения параллельных программ в структуры вычислительных систем / А.А. Пазников // Материалы Российской научно-технической конференции “Информатика и проблемы телекоммуникаций”. – Новосибирск: СибГУТИ, 2008. – С.152.

30. Курносов, М.Г. Об оптимизации распределения ветвей параллельных MPI-программ по процессорным ядрам вычислительного кластера / М.Г. Курносов, **А.А. Пазников** // Материалы VII Международной конференции-семинара “Высокопроизводительные вычисления на кластерных системах”. – Нижний Новгород: ННГУ, 2007. – С. 218-225.

Личный вклад автора в совместные публикации

Результаты диссертационной работы, выносимые на защиту, принадлежат автору, что подтверждено публикациями в научных изданиях. Во всех совместных публикациях авторство неделимое.

Пазников Алексей Александрович

Алгоритмы организации функционирования мультикластерных вычислительных систем с иерархической структурой

Автореферат диссертации
на соискание ученой степени кандидата технических наук

Подписано в печать “16” апреля 2013 г.
Формат бумаги 60x84/16, отпечатано на ризографе, шрифт № 10,
изд. л.1,6, заказ № 86, тираж 100 экз., ФГОБУ ВПО “СибГУТИ”.
630102, г. Новосибирск, ул. Кирова, д. 86.