

« 24 » ноября 2021 г.

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

на диссертационную работу Черных Андрея Николаевича
«Методы и алгоритмы решения задач оптимизации ресурсов в
нестационарных распределенных гетерогенных вычислительных средах»,
представленную на соискание ученой степени доктора физико-
математических наук по специальности 2.3.5 (05.13.11) – Математическое и
программное обеспечение вычислительных машин, комплексов и
компьютерных сетей

Диссертация Черных А.Н. носит теоретический характер и направлена на проведение исследований, связанных с определением теоретических оценок границ оптимизации ресурсов при выполнении распределенных вычислений в нестационарных гетерогенных средах с учетом различных аспектов планирования и распределения вычислительной нагрузки.

Актуальность разработки фундаментальных основ распределения нестационарных ресурсов, методов их анализа и адаптивных алгоритмов планирования вычислений с различными сценариями распределения вычислительной нагрузки обусловлена тем, что в настоящее время данные проблемы не решены в полной мере, что приводит к существенному недоиспользованию возможностей современных вычислительных сред на практике. Решение данных проблем позволяет перейти от традиционного планирования к новым адаптивным подходам, совершенствующим существующие стратегии распределения ресурсов и учитывающим стремительную эволюцию вычислительных средств. При этом новые подходы к оптимизации ресурсов в таких сложных средах должны учитывать не только требования к производительности и эффективности использования ресурсов, а также пожелания пользователей по качеству обслуживания, стоимости услуг, среднему времени ожидания и решения задач.

Цель работы заключается в разработке новых стратегий планирования распределенных вычислений в нестационарных гетерогенных средах, обеспечивающих эффективное выполнение работ (решение задач пользователей) на основе использования динамического и адаптивного планирования, для реализации которого достаточно наличие ограниченных знаний о ресурсах и работах.

Анализ содержания работы.

Во введении обосновываются актуальность темы, новизна и практическая значимость диссертационного исследования, а также

формулируются его цели и задачи. В нем также представлены основные результаты диссертации и приведена ее структура.

Первая глава диссертации посвящена рассмотрению основных концепций планирования ресурсов в нестационарных вычислительных средах. В ней приведен краткий обзор основных современных моделей и методов, используемых при решении соответствующих задач на основе теории расписаний. Описана природа неопределенностей, возникающих в процессе планирования вычислений, и приведена их классификация. Проанализированы подходы к снижению степени неопределенности применительно к облачным вычислениям. В качестве эффективных альтернатив известным технологиям детерминированной оптимизации рассмотрены различные стохастические и реактивные, а также работающие в условиях неопределенности алгоритмы планирования распределенных вычислений.

Во второй главе вводится прикладная базовая модель гетерогенная распределенной системы, именуемой в дальнейшем ГРИД. Модель включает большое число процессоров, которые объединяются в рамках нескольких машин разного размера. Выполняемые в модели работы являются независимыми. Они имеют фиксированную степень параллелизма и поступают периодически с течением времени. Работа может быть выполнена только на процессорах, принадлежащих одной и той же машине. На данной модели рассматриваются методы онлайн-планирования параллельных работ с неизвестными параметрами и без прерываний. Основной целью планирования является минимизация общей продолжительности выполнения всех работ.

Предложен алгоритм планирования, гарантирующий необходимые оценки оптимизации ресурсов. Показано, что этот алгоритм, реализуемый с использованием известного подхода «кражи работ», может быть использован в качестве основы при разработке алгоритмов планирования в реальных системах. Исследованы различные модификации данного алгоритма. Сформулирован и доказан ряд теорем относительно вышеупомянутых оценок. Продемонстрировано, что достижение необходимых оценок гарантируется для рассмотренных алгоритмов планирования.

В третьей главе проанализированы алгоритмы планирования вычислений, обеспечивающие как распределение работ по машинам, так и построение локальных расписаний их выполнения. Предложена и исследована адаптивная схема распределения работ с использованием концепции допустимого распределения. Ее основной идеей является задание ограничений по распределению маленьких работ на большие машины, а также их динамическая адаптация к различным вычислительным нагрузкам и свойствам ГРИД. Данная схема позволяет достичь снижения возможного дисбаланса нагрузки, когда машины с большим числом процессоров могут выполнять работы с небольшим параллелизмом, и сократить вынужденное ожидание при выполнении высоко параллельных работ.

Представлено подробное исследование влияния коэффициента

допустимости, включенного в политику распределения работ, на общую эффективность работы системы. Сформулирован и доказан ряд теорем, определяющих теоретические оценки такого влияния.

Показано, что с точки зрения заданных критериев оптимизации, алгоритмы с допустимым коэффициентом использования машин превосходят алгоритмы, которые используют все доступные машины для распределения работ. Экспериментально установлено, что адаптивные алгоритмы планирования являются надежными и стабильными в разнообразных условиях функционирования среды, а также успешно справляются с различными вычислительными нагрузками.

В четвертой главе представлено исследование оптимизации распределения ресурсов классическими методами, что на практике является труднодостижимым из-за разнообразия реальных параллельных и распределенных сред. Рассмотрено планирование так называемых неclairvoyantных параллельных работ, то есть работ с известными требованиями к ресурсам, но с неизвестным временем выполнения. Предложено новое семейство стратегий планирования, основанных на выполнении двух чередующихся фаз последовательного и параллельного выполнения работ.

В рамках предложенного семейства стратегий обобщены известные предельные границы производительности в наихудшем случае и предложено использование двух дополнительных параметров: штрафа за распараллеливание работ и коэффициента регулирования простоя. Доказано, что регулирование этих параметров может улучшать предельные границы производительности.

Пятая глава посвящена проблемам планирования виртуализированных вычислительных ресурсов, предоставляемых пользователям в форме облачных сервисов с различными уровнями обслуживания, которые определяются выделительными ресурсами, выделяемыми в течение определенного периода времени.

В диссертации уровень обслуживания формализован путем введения слак-фактора и цены за единицу времени обработки. Предложен ряд алгоритмов планирования вычислений на данной модели и проведен конкурентный анализ различных сценариев выполнения работ с целью достижения оценок оптимальности.

В шестой главе модель, предложенная в предыдущем разделе диссертации, распространена на случай с несколькими уровнями обслуживания, предоставляемыми провайдером облачных ресурсов. Кроме того, сформулирована постановка задачи двухкритериальной оптимизации, в рамках которой учитываются увеличение дохода провайдера и снижение энергопотребления. Рассмотрены восемь стратегий планирования, характеризующиеся типом и объемом информации, используемой для принятия решения (без знаний о системе, работах и ресурсах; с информацией об энергопотреблении; с информацией о производительности машин).

Предложена стратегия, доминирующая над другими рассмотренными стратегиями в большинстве тестов и являющаяся устойчивой при изменении условий среды, а также обеспечивающая достижение улучшения критериев и гарантии качества обслуживания.

Седьмая глава фокусируется на формулировке и решении задачи планирования в облачных системах VoIP с учетом изменения динамической нагрузки, вариативность времени запуска виртуальных машин, свойств работ, характеристик облачной инфраструктуры, наличия множества пользователей и других специфических особенностей. Сформулированная задача рассматривается как частный случай динамической упаковки в контейнеры. Ее принципиальной новизной является временное существование элементов (вызовов) в упаковке контейнера.

Разработаны алгоритмы планирования, использующие в процесс адаптации своей работы к разнообразным гетерогенным средам три настраиваемые параметра: порог утилизации, интервал времени, оставшийся до окончания периода аренды виртуальной машины, и интервал прогнозирования. С помощью экспериментов на реальных данных показано, что разработанные алгоритмы с методами прогнозирования нагрузки превосходят известные алгоритмы подобного назначения, обеспечивая высокое качество обслуживания и более низкую стоимость, а также могут быть эффективно использованы в облачной среде VoIP.

В заключении приведены краткие выводы по основным результатам диссертационной работы.

Основной текст диссертации изложен на 294 страницах, включая 72 рисунка и 49 таблиц, состоит из введения, семи глав, заключения, библиографии, включающей 223 наименования, и двух приложений. Содержание исследований изложено достаточно ясно и понятно. Текст хорошо структурирован, иллюстрирован рисунками и оформлен в соответствии с требованиями. Автореферат в полной мере отражает содержание работы. Можно отметить обширный список научных проектов и программ, а также всероссийских и зарубежных конференций в рамках которых проводились и апробировались исследования.

Представленный в диссертации подход к оптимизации ресурсов в нестационарных распределенных гетерогенных вычислительных средах является новым. В рамках этого подхода основой планирования являются разработанные алгоритмы с параметрической настройкой их работы. Обоснованность полученных в работе результатов базируется на корректном применении классических методов их доказательства, а также подтверждается результатами экспериментов.

Практическая значимость результатов диссертации заключается в повышении эффективности планирования параллельных работ в современных вычислительных системах за счет применения полученных результатов диссертационных исследований, обеспечивающих возможность адаптации алгоритмов планирования к изменению параметров рабочей нагрузки систем.

Основные результаты диссертационных исследований были использованы в рамках широкого спектра научно-технических работ, выполненных в различных российских и зарубежных научных и образовательных организациях.

Все основные результаты по теме диссертации достаточно полно опубликованы в 64 научных работах. В их числе 21 статья в журналах из списка ВАК и изданиях, проиндексированных в международные базы данных Scopus и Web of Science, включая 12 публикаций в изданиях с квартилем Q1, 1 – с квартилем Q2, 4 – с квартилем Q3, и 2 – с квартилем Q4. 42 работы опубликовано в трудах российских и международных конференций. Получены четыре свидетельства о государственной регистрации программ для ЭВМ и один патент.

Замечания.

- 1) Во Введении в Кратком содержании далеко не все значения неиспользуемых символов пояснены, что затрудняет восприятие этой части работы (в дальнейшем, в основном тексте, нужные пояснения присутствуют). То же самое касается пунктов Новизны, в части формулировки доказанных теорем.
- 2) Более детальное уточнение некоторых используемых терминов, заимствованных из англоязычных источников, например, таких как неклаирвоаянная работа, слак-фактор и др., упростило бы понимание излагаемого материала.
- 3) В тексте диссертации местами присутствует избыточное изложение некоторых понятий и положений.
- 4) Правильно было бы дать в начале или в конце работы раздел с пояснением используемых обозначений и аббревиатур, что существенно облегчило бы восприятие содержания диссертации.
- 5) В ряде экспериментов по оптимизации распределения ресурсов было бы интересно исследовать обобщение области дискретных значений варьируемых исходных параметров (характеристик работ и среды) предложенных моделей на некоторую допустимую непрерывную область значений таких параметров.
- 6) Из текста диссертации не совсем ясно влияют ли различные периоды времени (суток, месяца, года) на изменение реальной вычислительной нагрузки облачных ресурсов и, как следствие, на выбор уровней обслуживания в алгоритмах, работа которых моделируется в шестой главе.
- 7) В тексте диссертации имеются несущественные опечатки, не препятствующие восприятию изложенного материала исследований.

Выводы. Отмеченные недостатки не снижают высокого научного уровня и практической значимости диссертации. В целом диссертационная работа Черных А.Н. представляет собой самостоятельную и законченную научно-исследовательскую работу, подводящую итог многолетним и

плодотворным исследованием соискателя. Данная работа является новым крупным научным достижением в области оптимизации ресурсов в нестационарных распределенных гетерогенных вычислительных средах. Все основные результаты диссертационного исследования снабжены теоретическими доказательствами и экспериментальными расчетами, достоверность которых не вызывает сомнений. Основные результаты диссертации опубликованы в ведущих российских и зарубежных рецензируемых изданиях и прошли апробацию на крупных международных и всероссийских конференциях. Работа выполнена на высоком теоретическом уровне и свидетельствует о хорошей профессиональной квалификации автора. Содержание диссертационной работы соответствует пунктам 3, 8 и 9 области исследований из паспорта специальности 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Таким образом, диссертационная работа «Методы и алгоритмы решения задач оптимизации ресурсов в нестационарных распределенных гетерогенных вычислительных средах» отвечает требованиям ВАК, предъявляемым к диссертациям на соискание ученой степени доктора физико-математических наук по специальности 2.3.5 (05.13.11) – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей, а ее автор, Черных Андрей Николаевич, несомненно, заслуживает присуждения ему ученой степени доктора физико-математических наук.

Отзыв обсужден и одобрен на заседании Общелабораторного семинара Лаборатории информационных технологий им. М.Г. Мещерякова 19 ноября 2021г.

Сведения о ведущей организации: Международная межправительственная организация Объединенный институт ядерных исследований

Адрес: 141980, Россия, Московская обл., г. Дубна, ул. Жолио-Кюри, 6

Электронная почта: post@jinr.ru

Телефон: +7(496)216-50-59 (секретариат)

ОКПО 08626319; ОГРН 1035002200221

ИНН/КПП 9909125356/501063001

Отзыв составил:

Директор ИИТ им. М.Г. Мещерякова ОИЯИ,
доктор технических наук

/Кореньков В.В./