

ОТЗЫВ

официального оппонента 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»

Актуальность темы исследования.

Диссертационная работа Черных Андрея Николаевича направлена на создание математических основ адаптивного планирования ресурсов гетерогенных распределенных вычислительных сред в условиях высокой степени их нестационарности, как по техническим характеристикам и организации их работы, так и по интенсивности их использования. Под гетерогенными распределенными вычислительными средами в диссертации понимаются различные инфраструктуры - грид, облака и др. В диссертации такие распределенные системы обозначаются как ГРИД. Ключевой характеристикой таких систем является недостаток точной информации о компьютерных ресурсах, имеющий неустранимый характер для формирования плана выполнения вычислительных работ. Эта проблема относится к классу NP полных задач комбинаторной оптимизации. Что означает невозможность общих решений полиномиальной сложности в рамках действующей вычислительной парадигмы (машина Тьюринга). Некоторые перспективы решения таких проблем появились в связи с концепцией квантового компьютеринга, в рамках которой, в настоящее время, активно обсуждаются и задачи комбинаторной оптимизации. Но эта перспектива неопределенного будущего. А на сегодня актуальными являются эвристические подходы и решения в суженных формулировках общей проблемы, когда возможно эффективное продвижение по сложности задачи до уровней, интересных для приложений. Что и определяет высокую актуальность темы диссертационных исследований А.Н. Черных.

Теоретическая и практическая значимость полученных результатов, их научная новизна.

В диссертации в рамках вышеописанной общей проблемы поставлен ряд задач, решение которых оформлены в виде списка полученных в диссертации результатов. Среди них результаты демонстрирующие существенные отличия задачи оптимизации в ГРИД случае от классической (многопроцессорной) задачи, результаты обобщающие известные результаты по оптимизации многопроцессорной системы, а также результаты нового ГРИД типа:

- показано, что планирование в ГРИД сложнее, чем для многопроцессорной системы, в частности, списочный алгоритм не гарантирует конкурентный фактор;
- доказано, что граница списочного алгоритма Гарей-Грэхэма в ГРИД случае не работает;
- представлены решения для ГРИД планирования, обобщающие классический метод списочного расписания, на основе нескольких списков работ для каждой машины;
- получены новые результаты по аппроксимационным и конкурентным факторам, расширяющие известные результаты, с возможностью планирования без точной

информации о производительности узлов и времени выполнения работ. Эти факторы варьируются через изменение коэффициента допустимости;

- разработаны подходы на основе динамической адаптации к нестационарности системных параметров;

- разработан алгоритм планирования работ с фиксированной степенью параллелизма и неизвестным временем обработки;

- предложен адаптивный алгоритм планирования на основе концепции допустимого распределения на основе двухуровневой модели ГРИД (распределение параллельных работ – составление локальных расписаний);

- показано, что коэффициент допустимости может быть динамически скорректирован для учета динамического изменения рабочей нагрузки, улучшая конкурентный фактор;

- разработаны новые стратегии планирования с сочетанием последовательного и параллельного выполнения работ;

- предложен новый метод планирования работ с неизвестными временами выполнения за счет регулирования периодов простоя машин в контексте общей политики списочных расписаний;

- обобщены известные предельные границы производительности планирования в наихудшем случае. Введены два дополнительных параметра, штраф за распараллеливание и коэффициент регулирования простоя;

- показано, что существует компромисс между более ранним началом распараллеливания и задержкой их распараллеливания, за счет чего можно балансировать запросы пользователя с накладными расходами на параллелизацию;

- предложена новая модель облачных вычислений для выполнения и планирования работ на основе понятия уровня обслуживания;

- разработаны новые алгоритмы для сценариев с одним и несколькими уровнями обслуживания и с одной и несколькими машинами;

- разработан облачный VoIP планировщик, который учитывает динамическую рабочую нагрузку. Задача планирования рассмотрена как частный случай динамической упаковки в контейнеры. Временное существование элементов (вызовов) при упаковке является принципиальной новизной данной проблемы.

Все перечисленные результаты, полученные в рамках диссертационной работы, являются новыми и оригинальными.

Диссертант характеризует свою работу, как теоретическую, что, на наш взгляд, обосновано. При внешнем впечатлении разнородности полученных результатов, в целом они составляют хорошую теоретическую основу для дальнейшего прогресса в рамках общей проблемы адаптивного планирования ресурсов гетерогенных распределенных вычислительных сред в условиях высокой степени их нестационарности. При этом полученные результаты имеют и высокую практическую значимость, что подтверждается востребованностью в сфере информационных технологий.

Одним из важнейших, на наш взгляд, практически важных результатов связан с появлением во второй половине 00-х годов технологий виртуализации вычислительных ресурсов, сыгравших революционную роль в развитии распределенных гетерогенных вычислений. Это привело к массовому предоставлению вычислительных ресурсов пользователям в форме облачного сервиса IaaS. Технология IaaS скрывает от пользователя весь комплекс технических и организационных аспектов по вычислительным ресурсам. В результате, при

планировании вычислений (через IAAS), необходимо учитывать QoS ограничения. Также должны обеспечиваться различные уровни обслуживания SLA, чтобы гарантировать время выполнения работы в зависимости от предоставляемого уровня SLA. Это обуславливает научную важность задачи планирования вычислений с учетом качественных характеристик IaaS в условиях неопределенности нестационарной распределенной гетерогенной системы. Заметим, что это проблема успешно решена в диссертационной работе, и научная значимость этого результата состоит, во-многом, в его прикладной ценности.

Практическая и теоретическая значимость полученных результатов и вклад диссертанта в развитие соответствующей отрасли знаний подтверждается более чем внушительным количеством цитирований результатов в международных изданиях: 2512 ссылок в Google Scholar (h-index = 25) и 1412 ссылок в Scopus (h-index = 20).

Достоверность и обоснованность научных положений, выводов и рекомендаций подтверждена корректным применением выбранных методов исследования, строгими доказательствами математических утверждений и анализом эффективности разработанных моделей и алгоритмов. Результаты численных экспериментов в полной мере коррелируют с теоретическими выводами. Обоснованность положений, выдвигаемых на защиту, подтверждается не только достоверностью полученных результатов, но и соответствием современному уровню научного знания в соответствующих областях математики и информатики.

Структура диссертации. Диссертация состоит из Введения, 7-ми Глав, Заключения, библиографии из 223-х источников, и 2-х приложений. Общий объем основного текста работы 294 страницы, включая 49 таблиц и 72 рисунка.

Апробация полученных результатов.

Основные результаты работы докладывались и обсуждались на 88-ти международных и 14-ти российских научных конференциях, среди отметим следующие: Открытая конференция ИСП РАН им. В.П. Иванникова - 2020, 2019, 2018; Международная конференция «Инжиниринг & Телекоммуникации» - 2020, 2019, 2017, 2016, 2015; IPDPS–IEEE International Parallel and Distributed Processing - 2021, 2019, 2018, 2017, 2012, 2010, 2008, 2006 (CORE Rank A); CCGRID–IEEE/ACM International Symposium Cluster, Cloud and Internet Computing - 2020, 2016 (CORE Rank A); CloudCom–IEEE International Conference on Cloud Computing Technology and Science - 2015 (CORE Rank A); HPCS International Conference on High Performance Computing & Simulation - 2020, 2019, 2018, 2014, 2012 (CORE Rank B); InterCloud-HPC International Symposium on Cloud Computing and Services for High Performance Computing Systems - 2020 (CORE Rank B); PPAM International Conference on Parallel Processing and Applied Mathematics - 2017, 2011, 2007, 2005, 2003 (CORE Rank C); ISUM International Supercomputing Conference in Mexico - 2019, 2017, 2016, 2014, 2012, 2010.

Всего по теме диссертационного исследования автором было опубликовано 64 статьи, в том числе 21 статья в журналах из списка ВАК/Scopus/WoS. 42 публикации вышли в трудах российских и международных конференций. Получен 1 патент и 4 Свидетельства о государственной регистрации программ для ЭВМ. Высокий индекс цитируемости журналов (многие уровня Q1) и уровень конференций (некоторые ранга CORE Rank A), в которых опубликованы научные статьи диссертанта, говорит о высоком уровне научной значимости и качестве результатов диссертационных исследований.

Основные результаты диссертационного исследования были использованы в рамках 17-ти научно-технических работ, среди которых проекты: Министерства образования и науки Российской Федерации, РФФИ – Российского фонда фундаментальных исследований, CICESE Research Center (Мексика), ANII – Национального агентства по изучению и инновациям (Уругвай), FNR – Национального фонда научных исследований (Люксембург), BSC – Барселонского суперкомпьютерного центра (Испания) и других научно-исследовательских организаций.

Тематика работы и основные результаты диссертации соответствуют следующим областям исследований паспорта специальности ВАК 2.3.5 (05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»: п. 3 «Модели, методы, алгоритмы, языки и программные инструменты для организации взаимодействия программ и программных систем», п. 8 «Модели и методы создания программ и программных систем для параллельной и распределенной обработки данных, языки и инструментальные средства параллельного программирования», п. 9 «Модели, методы, алгоритмы и программная инфраструктура для организации глобально распределенной обработки данных».

Замечания.

В диссертации рассматриваются вычислительные ресурсы распределенных систем и проблема планирования их использования в приложениях. Таким образом в фокусе внимания два объекта планирования: вычислительные ресурсы и пользовательские задания (в диссертации используется термин *работы*). В этом отношении задача во многом схожа с проблемой планирования исполнения] пользовательских заданий в многопроцессорных вычислительных системах (HPC -high performance computing). Однако распределенные вычислительные среды создаются часто для приложений, для которых в вычислениях требуются большие объемы данных – HTC (High Throughput Computing). Для таких приложений уже не два объекта планирования, а три – добавляется еще и компьютерные ресурсы хранения (дисковые, ленточные, ...). Причем в современных распределенных средах и для таких ресурсов имеется сильная степень нестационарности. В результате проблема планирования становится сильно сложнее. В диссертации такие распределенные вычислительные среды исключены из рассмотрения. Возникает вопрос – возможно ли обобщать методы и подходы, развитые в диссертации и на такие распределенные среды? Заметим, что в задаче планирования в распределенных средах есть и четвертый объект – сетевые ресурсы (каналы передачи данных), для которых также может возникать сильная нестационарность для исполнения пользовательских заданий. Понимание того, что в планировании должны участвовать четыре объекта (*актора*) - пользовательские задания, вычислительные ресурсы, ресурсы хранения данных и сетевые ресурсы, возникло еще на рубеже 90-х -00-х годов, когда обсуждалась концепция *grid* вычислений и глобальная *grid* инфраструктура WLCG для распределенной обработки и анализа данных Большого адронного коллайдера в ЦЕРН. Правда четвертый объект, сетевые ресурсы, как оказалось, не было необходимости обсуждать в рамках проблемы планирования – просто в силу резко опережающего увеличения пропускной способности каналов передачи данных. Однако для таких распределенных инфраструктур третий объект планирования, ресурсы хранения данных, до сих пор представляют серьезное усложнение проблемы планирования. И сейчас эта проблема существует для инфраструктур, в которых *grid* и *cloud* сервисы добавляют серьезное усложнение в гетерогенность таких инфраструктур.

Заметим, однако, что указанное замечание не является критическим и не снижает научную и практическую ценность проведенных исследований.

Заключение.

Диссертационная работа Андрея Николаевича Черных представляет собой завершенное научное исследование в рамках сложнейшей проблемы выработки фундаментальных основ планирования нестационарных ресурсов, их анализ, и разработка новых адаптивных алгоритмов для различных сценариев. Совокупность разработанных теоретических положений можно квалифицировать как новое достижение в этой области. Диссертация и автореферат написаны ясным языком, имеют четкую структуру, внутреннее единство содержания, смысла и терминологии. Автореферат правильно отражает содержание диссертации.

В результате, приходим к заключению, что диссертация и автореферат соответствует требованиям ВАК РФ, предъявляемым к докторским диссертациям по специальности 2.3.5 (05.13.11) “Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей”, а ее автор, Черных Андрей Николаевич, заслуживает присуждения ученой степени доктора физико-математических наук по вышеуказанной специальности.

Официальный оппонент,
доктор физико-математических наук,
главный научный сотрудник
КК НБИКС-природоподобных технологий
НИЦ «Курчатовский институт»

Ильин Вячеслав Анатольевич
22 ноября 2021

Подпись сотрудника НИЦ «Курчатовский ин
Главный ученый секретарь НИЦ «Курчатовск

А. заверяю

И.И. Еремин