

«УТВЕРЖДАЮ»

Директор Федерального
государственного учреждения
«Федеральный исследовательский центр
«Информатика и управление»
Российской академии наук»,
член-корреспондент РАН



М.А. Посыпкин

» мая 2026 г.

ОТЗЫВ

ведущей организации

**Федерального государственного учреждения
«Федеральный исследовательский центр «Информатика и управление»
Российской академии наук»
на диссертацию Тутова Андрея Владимировича
на тему «Модели и методы распределения информационных и вычислительных
ресурсов гетерогенных центров обработки данных»,
представленную на соискание ученой степени кандидата технических наук
по специальности 2.3.8 – «Информатика и информационные процессы
(технические науки)»**

Актуальность темы диссертации

В условиях экспоненциального роста объемов данных и увеличения вычислительных нагрузок современные центры обработки данных (ЦОД) становятся гетерогенными, включающими разнородные аппаратные и программные ресурсы. Помимо серверов с традиционными архитектурами процессоров, стали широко использоваться серверы с графическими (GPU) процессорами для решения задач высокопроизводительных вычислений, облачных сервисов и обработки больших данных. Это обуславливает необходимость разработки новых моделей и методов эффективного распределения информационных и вычислительных ресурсов, способных обеспечить баланс между производительностью и энергоэффективностью.

Актуальность темы также определяется потребностью в разработке отечественного программного обеспечения, в частности планировщиков ресурсов гетерогенных ЦОД. Разработка эффективных алгоритмов систем управления ресурсами позволит повысить эффективность работы ЦОД, снизить операционные затраты и обеспечить устойчивость к динамическим изменениям нагрузок.

Структура и основное содержание диссертации

Диссертация состоит из введения, пяти глав, заключения, списка литературы из 170 наименований и 5 приложений. Полный объем диссертации составляет 212 страниц, включая 67 рисунков и 25 таблиц.

Во введении соискатель обозначил актуальность исследования, сформулировал цель и основные задачи работы. Приведены новые научные результаты, полученные в ходе

исследования, а также их практическая значимость и апробация. Кроме того, сформулированы основные положения, выносимые на защиту.

В первой главе проанализированы основные проблемы развития операторов связи и облачных провайдеров, обусловленные ростом цифровых сервисов, Интернета вещей и внедрением технологий машинного обучения. Современные ЦОД становятся гетерогенными, а спрос на аренду виртуальных машин с графическими процессорами стремительно растет. Рассмотрены технологии виртуализации, включая NVIDIA MIG, позволяющую на аппаратном уровне разделять GPU между виртуальными машинами. Основное внимание уделено проблеме высокого энергопотребления ЦОД и необходимости эффективного управления ресурсами. Проведен обзор существующих подходов к размещению и миграции виртуальных машин, выявлен пробел в учете гетерогенности ЦОД и специфики GPU-ресурсов. Сформулированы цель и задачи исследования, направленные на совершенствование управления ресурсами гетерогенных облачных ЦОД.

Во второй главе автором исследуется задача статического размещения виртуальных машин (VM) в гетерогенном ЦОД с учетом энергопотребления, равномерности загрузки ресурсов и соблюдения соглашений об уровне сервиса (SLA). Предложена многокритериальная математическая модель, включающая нормированные показатели энергопотребления серверов и видеокарт, неиспользованных виртуальных ресурсов (vCPU, vRAM, vGPU, vGDDRAM) и нарушений SLA. Для решения задачи используется муравьиный алгоритм, адаптированный автором для оптимизации размещения VM на гетерогенных физических серверах. Проведено сравнение предложенного подхода с эвристиками First Fit и Best Fit на имитационной модели гетерогенного ЦОД. Результаты показали, что муравьиный алгоритм обеспечивает более сбалансированное распределение ресурсов и меньшее количество нарушений SLA, хотя и требует больших вычислительных затрат. Определены эффективные параметры алгоритма.

В третьей главе рассматривается динамическое распределение ресурсов в облачном ЦОД, включающее мониторинг, прогнозирование загрузки серверов, выбор виртуальных машин для миграции и их размещение. Для прогнозирования перегрузки/недогрузки серверов автором предложено использовать метод группового учета аргументов (МГУА), а сглаживание данных осуществлять методом скользящего окна. Для оценки эффективности алгоритмов проведено имитационное моделирование. Определены рациональные параметры обучающей выборки и размера скользящего окна. Введен коэффициент устойчивости миграции, связывающий длительность миграции и размер окна наблюдения. Также в третьей главе разработан метод расчета длительности миграции и простоя виртуальных машин на основе анализа плотности вероятности. Результаты позволяют повысить точность прогнозирования длительности миграции и минимизировать риски нарушения SLA.

В четвертой главе сформулирована задача двухкритериальной оптимизации для определения лучшего размещения выбранных для миграции VM. Задача относится к комбинаторной задаче оптимизации с бинарными переменными. Предложен метод решения, заключающийся в её сведении к закрытой транспортной задаче. Проведены три серии экспериментов: сравнение с эвристическими алгоритмами (BFD, FFD), оценка масштабируемости и имитационное моделирование крупномасштабного ЦОД. Результаты показали, что предложенный метод превосходит эвристики по комбинированному критерию ESV (энергопотребление и нарушения SLA) в среднем в 3,5 раза, а венгерский метод демонстрирует лучшую вычислительную эффективность по сравнению с симплекс-методом. Это позволяет существенно улучшить управление ресурсами гетерогенных ЦОД и обеспечить соблюдение SLA.

Пятая глава посвящена описанию прикладных задач, связанных с реализацией предложенных моделей и методов в распределенном планировщике ресурсов ЦОД.

Описана двухуровневая архитектура управления ресурсами, включающая локальные и глобальный контроллеры. Разработаны алгоритмы обнаружения недогрузки/перегрузки хостов, выбора и размещения ВМ с учётом минимизации времени миграции и накладных расходов. Разработаны рекомендации по интеграции алгоритмов в открытую облачную платформу OpenStack, что позволяет повысить эффективность управления ресурсами гетерогенных ЦОД и обеспечить соблюдение SLA-соглашений.

Степень обоснованности научных положений и выводов, новизна результатов

В диссертации корректно используются методы математической статистики, многокритериальной оптимизации, прогнозирования и имитационного моделирования. Эффективность разработанных моделей и методов подтверждена имитационным моделированием и реальным внедрением. Достоверность также подтверждается публикациями в ведущих мировых и российских рецензируемых журналах, докладами на всероссийских и международных конференциях и научных семинарах.

В работе Тутова А.В. получены новые научные результаты.

1. Разработаны метод и модель описания первоначального (статического) размещения ресурсов виртуальных машин в гетерогенном ЦОД, которые, в отличие от существующих аналогов, учитывают возможности GPU-серверов и обеспечивают сбалансированное решение по множеству критериев, в то время как применяемые на практике методы обычно ограничиваются единственным критерием.

2. Разработан новый критерий определения рациональной длительности окна наблюдения за серверами с учетом времени размещения информации. При его расчете используется метод оценки времени размещения (миграции) информации, что в совокупности позволяет повысить качество мониторинга и, как следствие, стабильность облачных сервисов.

3. Разработана архитектура программно-аппаратного комплекса инфокоммуникационной системы ЦОД, в которой реализуются методы и алгоритмы первоначального и динамического размещения виртуальных машин, прогнозирования загрузки физических серверов и миграции информационных ресурсов, что повышает энергоэффективность и стабильность облачных сервисов в среднем в 3,5 раза в отличие от используемых на практике алгоритмов;

Значимость для науки и практики полученных автором диссертации результатов

Теоретическая значимость диссертации заключается в разработке критериев, моделей и методов распределения ресурсов в гетерогенных облачных ЦОД, позволяющих расширить теорию построения автоматизированных информационных систем управления ресурсами. Практическая значимость заключается в разработанных алгоритмах, рекомендациях, архитектуре и программных модулях для практической интеграции моделей и методов программно-аппаратного комплекса распределенного планировщика информационных и вычислительных ресурсов в гетерогенных ЦОД.

Рекомендации по использованию результатов диссертации

Судя по приведенному в диссертации перечню актов результаты диссертации внедрены и апробированы в сетях крупных операторов связи и международных компаниях, располагающих собственными ЦОД.

Основные результаты диссертационного исследования, опубликованы в открытой печати. Автор выпустил 30 научных трудов, в том числе: 8 статей в журналах, индексируемых Scopus, в том числе в журналах 3 квартиля, 13 статей в рецензируемых

научных журналах, рекомендуемых ВАК при Минобрнауки России. Зарегистрированы три программы для ЭВМ. Результаты диссертации внедрены в учебный процесс Московского технического университета связи и информатики.

Замечания

По результатам анализа диссертационной работы Тутова А. В. целесообразно сформулировать следующие замечания:

1. В главах 2 и 4 диссертации критерии оптимизации (энергопотребление, SLA, неиспользованные ресурсы) сформулированы математически, но не всегда ясно, как они соотносятся с реальными бизнес-метриками (например, стоимость владения ЦОД, штрафы за нарушение SLA). Было бы целесообразно добавить экономическое обоснование выбора весовых коэффициентов в аддитивной свертке.

2. В работе подчёркивается гетерогенность ЦОД, но не всегда ясно, как именно учитываются специфические ограничения (например, совместимость vGPU разных поколений, ограничения по пропускной способности сети при миграции). Рекомендуется дополнить описание алгоритмов деталями о поддержке гетерогенных конфигураций.

3. В главе 4 приведены данные о времени работы венгерского метода, но не рассмотрены сценарии для очень крупных ЦОД, например, 10 000+ серверов. Стоит обсудить возможности распределённой обработки или аппроксимации для промышленного применения.

4. В работе сравниваются предложенные алгоритмы с классическими эвристиками (FFD, BFD), но отсутствует сравнение с последними разработками, например алгоритмами на основе глубокого обучения или обучения с подкреплением, которые стали применяться в облачных системах в последнее время. Это поможет позиционировать работу в современном контексте.

Указанные замечания не снижают ценности работы и не оказывают решающего влияния на положительную оценку диссертационной работы Тутова А. В. Она представляет собой законченную работу, в которой решена задача разработки моделей и методов распределения ресурсов в гетерогенных облачных ЦОД.

Заключение

Результаты диссертационной работы и выводы по ней, полученные Тутовым А. В., свидетельствуют о том, что соискатель провел актуальное исследование, направленное на решение теоретических и практических задач центров обработки данных. Выполненные исследования отличаются от ранее выполненных другими авторами тем, что соискатель разработал эффективные модели и методы распределения ресурсов с учетом гетерогенности ЦОД. Важнейшей частью работы является комплексный подход — от теоретического обоснования и формализации задач до практической реализации и экспериментальной проверки. Это позволило не только предложить новые алгоритмы, но и доказать их эффективность и применимость в реальных условиях. На основе вышеизложенного можно заключить, что диссертация Тутова Андрея Владимировича «Модели и методы распределения информационных и вычислительных ресурсов гетерогенных центров обработки данных» является законченным научным исследованием, полностью удовлетворяет критериям положения о присуждении ученых степеней, предъявляемых к диссертации на соискание ученой степени кандидата технических наук по специальности 2.3.8 – «Информатика и информационные процессы (технические науки)», а ее автор, Тутов Андрей Владимирович, заслуживает присуждения степени кандидата технических наук по этой специальности.

Отзыв ведущей организации обсужден и принят на заседании отдела 51 «Информационные, управляющие и телекоммуникационные системы» Отделения 5 «Информационные, управляющие и телекоммуникационные системы. Информационная безопасность. Информационные технологии в цифровой экономике» ФИЦ ИУ РАН 8 мая 2026 г., протокол № 51/1 (присутствовало: 5 чел. Результаты голосования: за - 5, против - нет, воздержались - нет).

Козлов Сергей Витальевич,
ведущий научный сотрудник, руководитель отдела 51, Отделения 5 ФИЦ ИУ РАН
кандидат технических наук (специальность ВАК 20.01.09 – 1986 г.),
старший научный сотрудник

Федеральное государственное учреждение «Федеральный исследовательский центр
«Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН)

Адрес: 119333, Москва, Вавилова, д. 44, кор. 2

<http://www.frccsc.ru/>

Тел: +7 (499) 135-62-60

E-mail: frccsc@frccsc.ru