



005047734

На правах рукописи

Воробьев Андрей Игоревич

**МОДЕЛИ И МЕТОДЫ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ПРЕДОСТАВЛЕНИЯ  
ИНФОРМАЦИОННЫХ УСЛУГ В ЦЕНТРАХ ОБРАБОТКИ ДАННЫХ**

Специальность: 05.13.01 -  
Системный анализ, управление и обработка информации  
(технические системы)

**АВТОРЕФЕРАТ**  
диссертация на соискание ученой степени  
кандидата технических наук

20 ДЕК 2012

Санкт-Петербург – 2012

Работа выполнена на кафедре автоматизированных систем обработки информации и управления Санкт-Петербургского государственного электротехнического университета "ЛЭТИ" им. В.И.Ульянова (Ленина)

Научный руководитель: Академик Российской академии образования  
доктор технических наук, профессор  
**Советов Борис Яковлевич**

Официальные оппоненты: доктор технических наук, профессор  
**Водяхо Александр Иванович**  
Санкт-Петербургский государственный  
электротехнический университет "ЛЭТИ" им.  
В.И.Ульянова (Ленина), профессор кафедры  
вычислительной техники

кандидат технических наук  
**Васильев Николай Владимирович**  
ОАО «Информационные телекоммуникационные  
технологии», начальник сектора 01211

Ведущая организация:  
**Федеральное государственное автономное образовательное учреждение  
высшего профессионального образования "Санкт-Петербургский  
государственный университет аэрокосмического приборостроения"**

Защита диссертации состоится 24 декабря 2012 г. в 15:00 на заседании совета по защите докторских и кандидатских диссертаций Д 212.238.07 Санкт-Петербургского государственного электротехнического университета "ЛЭТИ" им.В.И.Ульянова (Ленина)" по адресу: 197376, Россия, Санкт-Петербург, улица Профессора Попова, дом 5.

С диссертацией можно ознакомиться в библиотеке Санкт-Петербургского государственного электротехнического университета "ЛЭТИ" им.В.И.Ульянова (Ленина)".

Автореферат разослан 23 ноября 2012 г.

Ученый секретарь  
диссертационного совета, к.т.н.



Цехановский В.В.

## **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

**Актуальность темы диссертации.** Фундаментом построения ИТ-инфраструктуры является Центр обработки данных (ЦОД). Предназначение ЦОД – обеспечение безотказной работы бизнес-процессов, надежное хранение данных и гарантированный доступ к ним, предоставление информационных услуг.

Центры обработки данных востребованы государственными службами (электронное правительство, госуслуги), крупными организациями, такими как, банки, страховые и торговые корпорации, предприятия добывающей отрасли, телекоммуникационные компании (биллинговые системы, хостинг, всевозможные Web-сервисы и социальные службы). Все они используют сложные бизнес-приложения, и их деятельность зависит от надежности функционирования ИТ-инфраструктуры.

Важнейшими преимуществами создания ЦОД являются консолидация вычислительных мощностей и систем хранения. В тоже время растущие потребности бизнеса необходимо удовлетворить в условиях ограниченных ресурсов: надо хранить больше данных в выделенном пространстве, быстрее производить вычисления, не потребляя и не выделяя слишком много энергии, передавать больше информации по имеющимся каналам связи, обеспечивать максимальную степень готовности эксплуатируемых ИТ-систем. И всё это с ограниченным финансированием. В таких условиях грамотное проектирование ЦОД является ключевым звеном для достижения эффективной работы предприятия, а указанные ограничения определяют выбор используемых технологий и оборудования.

Известно, что в области ИТ-систем высокой доступности ситуация постоянно и стремительно меняется. Отсутствие единого подхода к организации ЦОД, стандартов проектирования и эксплуатации различных серверных помещений и центров обработки ставят проблему выработки системного подхода к реализации инфраструктуры, а также разработки моделей и методов повышения эффективности предоставления информационных услуг в ЦОД. Появление опыта в виде стандартов, моделей и методов позволит унифицировать реализацию ЦОД и упростить внесение изменений в его инфраструктуру, способствовать тиражируемости и масштабируемости решений. Очевидно, что рациональным является выбор такой инфраструктуры, которая обеспечивала бы долгосрочную защиту инвестиций и позволяла компании выполнять текущие задачи и развиваться.

Особенное значение данная проблема приобретает в условиях экономического кризиса, который усилил дефицит ресурсов. Предприятия сокращают бюджеты на ИТ, меньше вкладывают в долговременные проекты вообще и в ЦОД в частности. Первостепенной становится задача создания легких, недорогих и в то же время эффективных решений, отвечающих требованиям бизнес-заказчиков. Основная проблема состоит в том, как в условиях постоянного увеличения стоимости ресурсов получать максимальную отдачу от их эксплуатации. Отсюда возникают ключевые вопросы, которые необходимо решать при реализации ЦОД, - каким образом добиться увеличения основных показателей (производительности, надежности и т.д.) при минимизации затрат (энергопотребление, администрирование), учесть возможный рост нагрузки, предусмотреть восстановление после сбоя и высокую доступность критически важных приложений и сервисов.

Сказанное говорит об актуальности разработки моделей и методов повышения эффективности предоставления информационных услуг в центрах обработки данных. Важно также учитывать технологии, базирующиеся на открытых отраслевых стандартах, минимизирующие проблемы взаимодействия по мере расширения ЦОД.

При оценке характеристик ЦОД большое значение имеют математическое моделирование и вычислительный эксперимент на модели.

Теория моделирования сложных систем и их элементов развита такими отечественными учеными, как В.А. Ершов, Г.Л. Захаров, А.П. Кулешов, О.И. Кутузов, В.Г. Лазарев, И.А. Мизин, Ю.И. Рыжиков, Б.Я. Советов, С.А. Яковлев. Оценка вероятностно-временных характеристик систем, оптимизация и распределение ресурсов рассматривались в работах зарубежных ученых: Л. Клейнрок, Д. Мартин, У. Столлингс, Э. Таненбаум, М. Шварц и другие. Работы этих и ряда

других ученых составляют теоретическую базу моделирования систем, в том числе отдельных узлов, кластеров, центров обработки данных, сетей хранения данных и их элементов.

Сказанное говорит о важности научной задачи разработки моделей и методов повышения эффективности функционирования ЦОД при предоставлении информационных услуг, имеющей большое значение для экономики страны.

**Целью работы** является повышение эффективности функционирования центров обработки данных, предоставляющих информационные услуги.

Для достижения названной цели необходимо решение следующих задач:

1. Разработать структурно-функциональную модель центра обработки данных.
2. Разработать комплекс моделей оценки качества предоставления информационных услуг в центрах обработки данных.
3. Разработать методы повышения эффективности предоставления информационных услуг в центрах обработки данных.

**Объектом исследования** являются центры обработки данных, выполненные в виде консолидации кластеров и систем хранения данных.

**Предметом исследования** являются процессы функционирования центров обработки данных при предоставлении информационных услуг.

**Методы исследования.** Для решения поставленных задач, в работе используются математические методы системного анализа, теории вероятности и математической статистики, эволюционных алгоритмов и теории массового обслуживания.

**Научная новизна работы** состоит в разработке моделей и методов повышения эффективности предоставления информационных услуг в ЦОД, позволяющих обеспечить соглашение о качестве обслуживания клиентов. Новые научные результаты:

1. Структурно-функциональная модель центра обработки данных отличается выделением набора структурных элементов и их функционального назначения, присущих большинству современных ЦОД, что позволяет более полно описывать процессы предоставления информационных услуг.
2. Комплекс моделей оценки качества предоставления информационных услуг в центрах обработки данных отличается учетом многоэтапности обработки клиентского запроса, что дает возможность оценить необходимые требования к производительности ЦОД.
3. Методы повышения эффективности предоставления информационных услуг в центрах обработки данных учитывают зависимость показателей эффективности от характеристик ЦОД, что позволяет обеспечивать необходимое качество обслуживания клиентов.

**Научные положения, выносимые на защиту:**

1. Структурно-функциональная модель центра обработки данных.
2. Комплекс моделей оценки качества предоставления информационных услуг в центрах обработки данных.
3. Методы повышения эффективности предоставления информационных услуг в центрах обработки данных.

**Достоверность научных результатов и выводов** подтверждается: корректностью математических выкладок, обоснованностью используемых ограничений, корректностью интерпретации в предметной области, результатами моделирования.

**Научная и практическая ценность** диссертационной работы заключается в том, что модели и методы повышения эффективности предоставления услуг в ЦОД можно рассматривать как единую технологию обеспечения необходимого качества обслуживания клиентов. Полученные результаты моделирования представляют практическую ценность при выборе архитектуры центров обработки данных.

Результаты работы внедрены в учебный процесс СПбГЭТУ (ЛЭТИ) и используются при проведении лабораторных работ, а также в курсовом и дипломном проектировании по темам, связанным с функционированием ИТ-систем, на кафедре автоматизированных систем обработки информации и управления.

**Апробация результатов работы.** Основные положения и результаты диссертационной работы представлялись на XII международной конференции «Информатика: проблема, методология, технологии» (Воронеж, 9-12 февраля 2012 г.); XIII международной конференции «Региональная информатика (РИ-2012)» (Санкт-Петербург, 24-26 октября 2012 г.); международной конференции «Региональная информатика», (Санкт-Петербург, 1-3 марта 2005 г.); второй всероссийской конференции «Абонентский доступ в сетях следующего поколения» (Санкт-Петербург, 2005 г.) и на научно-технических конференциях профессорско-преподавательского состава научных сотрудников и аспирантов СПбГУТ им. проф. М.А. Бонч-Бруевича в 2006 – 2012 гг.

**Публикации.** По теме диссертационной работы опубликовано 12 научных работ, из них – 3 статьи из перечня изданий, рекомендованных ВАК, и 9 – работ в научных трудах международных и всероссийских конференций.

**Структура и объем диссертации.** Диссертация состоит из введения, четырех глав, заключения, списка литературы, включающего 57 наименований. Основная часть работы изложена на 142 страницах машинописного текста. Работа содержит 41 рисунок, 12 таблиц.

### КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность темы диссертации, формулируется цель исследования и решаемые задачи, определена научная новизна и практическая ценность основных результатов работы.

В первой главе рассмотрена организация ИТ-инфраструктуры центра обработки данных. Предложена структурно-функциональная модель ЦОД, которая представлена на трех уровнях иерархии: структурные модули ЦОД, назначение и выполняемые функции модулей ЦОД, взаимодействие элементов ЦОД в единой системе и с внешней средой (1):

$$CM=f(\Pi, \Phi, G, G^{доп}), \quad (1)$$

где  $\Pi$  – структурные модули ЦОД,  
 $\Phi$  – множество функций, выполняемых ЦОД,  
 $G$  – множество характеристик ЦОД,  
 $G^{доп}$  – параметры внешней среды.

Современный ЦОД состоит из следующих функциональных модулей:

– Серверного комплекса (СК), включающего различные группы кластеров. Кластер рассматривается как совокупность однотипных серверов, объединенных коммутационной системой. Каждый сервер имеет один процессор и память ограниченного объема. Обозначим СК как множество  $M1 = \{CK_{i=1, A; j=1, B}, i - \text{число серверов в кластере } j\text{-й группы}\}$ .

– Системы хранения данных (СХД), предназначенной для организации надёжного хранения информационных ресурсов и предоставления к ним доступа со стороны серверов. Обозначим как множество  $M2 = \{СХД_{i=1, N; j=1, M}, i - \text{число СХД } j\text{-го типа}\}$ .

– Сетевой инфраструктуры, включающей оборудование маршрутизации и коммутации, оборудование организации оптических каналов связи; оборудование подключения пользователей; средства обеспечения информационной безопасности. Обозначим как множество  $M3 = \{CO_{i=1, I; j=1, J}, i - \text{наименований сетевого оборудования (СО) } j\text{-го типа}\}$ .

– Инженерных систем, отвечающих за предоставление необходимых условий для надежной эксплуатации вычислительного оборудования. Обозначим как множество  $M4$ .

– Системы управления, обеспечивающей контроль и администрирование всех функциональных модулей ЦОД. Обозначим как множество  $M5$ .

Таким образом, структурно-функциональная модель ИТ-инфраструктуры ЦОД представлена в виде следующего кортежа элементов:

$$\text{ЦОД} = \left\langle M1 \left\{ \bigcap_{i,j=1}^{i,j=A,B} \text{СК}_{ij} \right\}, M2 \left\{ \bigcap_{i,j=1}^{i,j=N,M} \text{СХД}_{ij} \right\}, M3 \left\{ \bigcap_{i,j=1}^{i,j=I,J} \text{СО}_{ij} \right\}, M4, M5 \right\rangle.$$

ЦОД структурируется из функциональных модулей в зависимости от масштаба корпоративной сети, поступающей нагрузки, особенностей запросов и бизнес-приложений. С увеличением нагрузки по определенному классу запросов производительность ЦОД обеспечивается расширением серверного комплекса через добавление одного или нескольких серверов, появление новых бизнес-приложений может привести к появлению нового вычислительного кластера. При необходимости резервного копирования, зеркалирования и других процессов, требующих значительных объемов памяти все системы хранения являются съемными и взаимозаменяемыми. При необходимости можно добавлять дополнительные типы систем хранения – более современные или объемные.

С точки зрения качества предоставляемых ЦОД информационных услуг интерес вызывают подсистемы М1и М2.

Основными показателями качества ЦОД являются следующие его характеристики: пропускная способность, время ответа на клиентский запрос, надежность функционирования, управляемость и эффективность.

*Пропускная способность*  $\Lambda$  отражает объем данных, обработанных ЦОД в единицу времени (2) и характеризует качество выполнения основной функции ЦОД – обработки запросов пользователей.

$$\Lambda = \lim_{\tau \rightarrow \infty} \frac{N(\tau)}{\tau} \text{ запросов /с,} \quad (2)$$

где  $N(\tau)$  – число запросов, обслуживаемых ЦОД за интервал времени (0,  $\tau$ ).

*Время ответа*  $\tilde{t}_{\text{отв}}$  – длительность промежутка времени от момента поступления запроса пользователя в ЦОД до момента окончания его выполнения. В общем случае время ответа – случайная величина и чаще всего оценивается средним значением (3):

$$\bar{t}_{\text{отв}} = M[\tilde{t}_{\text{отв}}], \text{ с,} \quad (3)$$

где  $M[\tilde{t}_{\text{отв}}]$  – математическое ожидание случайной величины  $\tilde{t}_{\text{отв}}$ .

*Надежность* – свойство системы выполнять возложенные на нее функции в заданных условиях эксплуатации с заданными показателями качества. Показано, что у сложных систем, типа ЦОД основной характеристикой надежности является отказоустойчивость – способность ЦОД работать в условиях отказа некоторых его элементов.

*Управляемость* ЦОД подразумевает возможность централизованно контролировать состояние функциональных модулей, входящих в состав ЦОД, выявлять и разрешать проблемы, возникающие при их работе, выполнять анализ производительности и планировать развитие ЦОД. Задача анализа производительности ЦОД сводится к частным задачам: оценке интегральных и дифференциальных характеристик ЦОД и исследованию характера зависимости этих характеристик от входной нагрузки, внешних условий и внутренних параметров ЦОД. Решению вопросов развития инфраструктуры ЦОД и выявления элементов, нуждающихся в модернизации, способствует оценка использования ресурсов ЦОД, основанная на оценке коэффициента эффективности использования ресурсов и коэффициента дисбаланса с последующей интерпретацией результатов в виде гистограмм и графов. По результатам анализа представляется возможным сделать выводы относительно степени влияния негативных

факторов на эффективность функционирования ЦОД; доли неэффективно используемых средств, вложенных в различные модули ЦОД; «узких» и потенциально «узких» мест ЦОД; необходимости дозагрузки или перераспределение ресурсов между серверами, кластерами и системами хранения ЦОД. Данный подход может применяться не только для оценки эффективности использования различных модулей ЦОД, но и характер этой задачи является ресурсосберегающим и создающим источник дополнительных инвестиций.

*Эффективность* – основной показатель качества работы ЦОД, характеризующий степень его способности выполнять свои функции по назначению. Данный показатель используется как для сравнения процессов самого ЦОД, с целью выбора его оптимальных параметров, так и для сравнительной оценки с другими ЦОД.

Процесс предоставления услуг центром обработки данных характеризуется появлением соглашений о качестве обслуживания (Service Level Agreement - SLA). SLA-соглашение – это часть договора между провайдером и клиентом, в котором описывается упорядоченная система показателей для контроля качества предоставляемых услуг, а также штрафы за нарушение соглашения и вознаграждения в случае успешного его выполнения. Наиболее распространенными показателями качества, задаваемыми в SLA-соглашениях, являются среднее время ответа на запрос и максимальное время ответа для заданной доли запросов.

Клиентский запрос информационной услуги ЦОД представляет собой задание, состоящее из последовательности задач и времени на обдумывание между этими задачами, т.е. временем, прошедшим между завершением выполнения задачи на каком-либо узле, например сервере кластера и поступлением следующей задачи на другой узел ЦОД, например системы хранения данных, в пределах одного задания. Таким образом, задание пользователя можно описать тройкой  $Z_i = \{t_i, A_i, T_i\}$ , где  $t_i$  – момент инициирования задания,  $A_i$  – атрибуты задания, определяющие имена источника, инициировавшего задание, пользователя задания, режим обработки задания, приоритет и др., и  $T_i$  – трасса задания. Трасса выполнения задания – последовательность задач (событий), связанных с изменением состояния задания. Трасса представляется в виде упорядоченного множества событий  $T_i = \{S_1, S_2, \dots, S_M\}$ , имевших место в моменты времени  $t_1, t_2, \dots, t_M$ , причем  $t_1 \leq t_2 \leq \dots \leq t_M$ . К событиям относятся моменты старта задания, начала и завершения обработки отдельных задач, начала и окончания выполнения процессов в системах хранения данных ЦОД, начала использования и освобождения ресурсов ЦОД, предоставляемых заданию, и др.

Задача диссертационного исследования сформулирована как многопараметрическая и многокритериальная экстремальная задача с ограничениями: при заданных значениях элементов множества параметров  $\Theta$ , задающих технические характеристики элементов ЦОД, найти необходимое множество параметров  $\Pi$ , описывающих структуру ЦОД (число серверов в кластере, число кластеров, систем хранения и т.п.), такое, чтобы выполнялись условия:

$$0 < g_i(\Theta, \Pi) \leq g_i^{\text{доп}}, \quad g_i \in \bar{G} \quad (4)$$

$$\phi \rightarrow \min f(\Theta, \Pi) \quad (5)$$

где  $\phi$  – аддитивная функция стоимости ЦОД, включающая стоимость технических средств и программного обеспечения.

$\bar{G}$  – вектор показателей качества ЦОД;

$g_i^{\text{доп}}$  – допустимое значение  $i$ -го показателя качества, определенное в SLA-соглашении.

Решение сформулированной задачи распадается на два этапа: на первом – на соответствующих моделях находятся зависимости показателей качества  $\bar{G}$ , на втором – решается задача, определяемая выражениями (4)–(5).

Определены подсистемы ЦОД и процессы предоставления информационных услуг, оказывающие существенное влияние на показатели качества. К ним относятся:

1) Подсистема  $M1$  – серверный комплекс. Чтобы учесть влияние серверного комплекса на

показатели качества необходимо разработать:

Модель оценки характеристик производительности ЦОД и запасов по пропускным способностям:

$$\bar{t}_{\text{отв}} = f(\lambda, T_{\text{обс}}), \Lambda(t_{\text{отв}}), D=f(\lambda),$$

где  $\lambda$  – нагрузка, поступающая в ЦОД, [запросов/с];

$T_{\text{обс}}$  – время обслуживания клиентского запроса в ЦОД, [с];

$\bar{t}_{\text{отв}}$  – время ответа на запрос пользователя;

$D$  – запас по пропускным способностям ЦОД.

Модель балансировки нагрузки в кластере ЦОД.

$$p(k_1, k_2, k_N) = f(K, N, \mu),$$

где  $p(k_1, k_2, \dots, k_N)$  – вероятность распределения запросов пользователей по серверам кластера;

$K$  – число запросов пользователей к  $i$ -му серверу кластера;

$N$  – количество серверов в кластере;

$\mu_i$  – производительность  $i$ -го сервера кластера, [запросов/с].

Модель оценки надежности ЦОД.

$$P_{\text{цод}} = f\{P_1, \dots, P_n, \theta(t)\},$$

где  $P_{\text{цод}}$  – вероятность безотказного функционирования ЦОД;

$P_i$  – вероятности безотказного функционирования  $i$ -го модуля ЦОД в процессе решения заданий клиентов ЦОД.

$\theta$  – множество всех возможных состояний ЦОД в момент времени  $t$ .

2) Подсистема М2 – хранилища данных. Чтобы учесть влияние систем хранения данных на показатели качества ЦОД необходимо разработать:

Модель размещения ресурсов в системах хранения данных

$$t_{\text{отв}} = f(C, \lambda, \mu, m),$$

где  $t_{\text{отв}}$  – время ответа на запрос пользователя;

$\lambda$  – интенсивность заданий, поступающих к ресурсу, [заданий/с];

$C$  – пропускная способность ресурса, [операций/с];

$\mu$  – среднее число операций, необходимых для выполнения задания;

$m$  – количество ресурсов.

Во второй главе предложен комплекс моделей оценки качества предоставления информационных услуг в центрах обработки данных.

Математическим аппаратом для оценки качества предоставления информационных услуг в центрах обработки данных выбраны безприоритетные разомкнутые экспоненциальные сети массового обслуживания (СеМО). Переход заявок между узлами СеМО (кластерами ЦОД) происходит мгновенно в соответствии с переходными вероятностями  $p_{ij}$ ,  $i, j = \overline{1, N}$ ,  $p_{ij}$  – вероятность того, что заявка после обслуживания в кластере  $i$  перейдет в кластер  $j$ . СеМО задается следующими параметрами: числом  $N$  СМО – кластеров ЦОД; числом  $K_1, \dots, K_N$  каналов в каждой  $i$ -й СМО,  $i = 1, \dots, N$ ; матрицей  $P = \| p_{ij} \|$  вероятностей передач,  $i = 1, \dots, N; j = 0, \dots, N$ ; интенсивностями  $I_1, \dots, I_N$  входных потоков заявок; средними временами обслуживания  $\bar{T}_{\text{обс}1}, \dots, \bar{T}_{\text{обс}N}$  заявок в каждой СМО (кластере ЦОД).



С точки зрения выполняемых ЦОД функций наиболее важными характеристиками СеМО, являются системные характеристики, т.к. они отражают свойства СеМО в целом, т.е. ЦОД рассматривается как единая, неделимая на части система. Важными системными характеристиками СеМО являются.

1) Среднее время  $\bar{T}_{\text{пр}}$  пребывания в сети. Временем пребывания в сети называется время между приходом заявки в сеть и ее уходом из сети. В случае обработки задачи клиента, это среднее время ответа  $\bar{T}_{\text{отв}}$

$$\bar{t}_{\text{отв}} = \frac{1}{I} \sum_{j=1}^N \lambda_j \bar{T}_{\text{пр}j}, \quad (6)$$

где  $I = I_1 + \dots + I_N$ ;

$\bar{T}_{\text{пр}j}$  – среднее время пребывания заявки в  $j$ -й СМО, т.е. обработки запроса в  $j$ -м звене (кластере) ЦОД.

2) Передаточные коэффициенты  $\alpha_{ij}$ ,  $i, j = \overline{1, N}$ . Пусть заявка входит в сеть из  $i$ -го входного потока. Ее маршрут в сети случаен, поэтому случайно и число приходов в  $j$ -ю СМО за время пребывания в СеМО. Среднее значение  $\alpha_{ij}$  этого числа приходов называют передаточным коэффициентом. Он однозначно определяется для любых  $i, j$ , матрицей  $P$  вероятностей передач.

3) Входовые средние времена  $F_1, \dots, F_N$  пребывания в сети. Величина  $F_j$  определяется как среднее время пребывания в СеМО заявки, поступающей из  $j$ -го входного потока ( $j = \overline{1, N}$ ). В случае ЦОД, с учетом разных видов заданий клиентов эта характеристика позволит определить среднее время ответа для разных категорий (классов) запросов, к которым могут предъявляться разные требования по качеству обслуживания.

4) Абсолютные пропускные способности  $A_j$ . Предположим, что в заданной СеМО интенсивности всех входных потоков, кроме  $j$ -го, заменены на нулевые, а  $I_j$  заменена на предельное значение, при котором сеть ещё стационарна. Это значение  $A_j$  называется абсолютной пропускной способностью по  $j$ -му входу.

5) Условные пропускные способности  $B_1, \dots, B_N$ . Предположим, что в заданной СеМО значение интенсивности  $I_j$  заменено на максимальное значение, при котором сеть ещё стационарна. Это значение  $B_j$  называется условной пропускной способностью по входу  $j$ . Данная характеристика имеет большее значение для развития планирования ЦОД. Если реальная пропускная способность стремительно приближается к условной, то необходимо планировать меры по масштабированию ЦОД.

6) Запасы  $D_1, \dots, D_N$  по пропускным способностям. Запас  $D_j$  показывает, насколько может быть увеличена интенсивность прихода заявок на  $j$ -м входе (при заданных остальных) без нарушения условия стационарности. Запас  $D_j = B_j - J_j$ ,  $j = \overline{1, N}$ .

Названные характеристики могут дать ценную информацию о свойствах ЦОД. Например, среднее время пребывания  $\bar{T}_{\text{пр}}$  характеризует среднее время обработки клиентского запроса, а запасы  $D$  выражают готовность ЦОД продолжать устойчивое функционирование при увеличении нагрузки (интенсивности запросов) по тому или иному входу.

Поскольку в SLA-соглашениях задаются ограничения на качество обслуживания для запросов различных видов, в работе предлагается модель ЦОД, которая учитывает современную структуру входного потока заявок. Поведение пользователей характеризуется их сессиями. Сессия пользователя состоит из последовательности заданий, а также времени на обдумывание, т.е. временем, прошедшим с момента отправки ответа сервером какого-либо кластера и до поступления нового задания в

систему в рамках одной сессии. Типичная сессия может длиться несколько минут. Поэтому в модели необходимо учитывать длительное время жизни сессий. Для моделирования процесса обдумывания в ЦОД, имеющего  $c = 1, \dots, C$  кластеров введен виртуальный сервер ( $c=0$ ) с бесконечным числом параллельных независимых каналов обслуживания, позволяющий отразить в модели независимость времени на обдумывание пользователей от времени обработки задания в ЦОД.

Предлагается оценивать время ответа  $T_k$  на запрос класса  $k$  по формуле (7), как суммарное время прохождения запроса через все звенья ЦОД

$$M[T^k] = \sum_{c=1}^C \frac{\lambda_c^k}{\lambda^k} \cdot \frac{m_c^k}{S_c - q_c}, \quad (7)$$

где  $C$  – число кластеров ЦОД;

$S_c$  – число серверов в кластере  $c$ ;

$q_c$  – номинальная загрузка кластера  $c$ ;

$m_c^k$  – среднее время ответа на запрос класса  $k$  сервером кластера  $c$ ;

$\frac{\lambda_c^k}{\lambda^k}$  – среднее число посещений кластера  $c$  запросом класса  $k$  за время его нахождения в ЦОД.

Предложена модель сервиса балансировки нагрузки в ЦОД. Предполагается, что имеется поток заданий к ЦОД и все задания ведут себя статистически одинаково, причем каждое из них требует некоторого числа операций от приложения, вычислительных ресурсов (серверов) и ресурсов хранения (систем хранения данных), обслуживающих  $M$  пользователей. Формально постановка задачи балансировки нагрузки проведена с использованием модели замкнутой СеМО с центральным обслуживающим прибором, которая описывает работу мультипрограммной вычислительной системы с фиксированным числом узлов, в которую допускается точно  $K$  заданий. Эти задания циркулируют в системе, коллективно используя  $N$  ресурсов. Центральный обслуживающий прибор представляет собой корневой узел кластера с сервисом балансировки нагрузки, а остальные  $(N-1)$  узлов – периферийные узлы, собственно выполняющие обработку клиентских запросов. Исследование данной системы с точки зрения пользователя позволило найти решение, выражаемое через параметры  $\mu_i$  – производительность  $i$ -го узла кластера и  $p_i$  – вероятность перехода заявки на  $i$ -й узел кластера. Для частного случая замкнутой Марковской сети с центральным обслуживающим прибором решение по распределению задач пользователей по серверам кластера представлено формулами (8) и (9).

$$P(k_1, k_2, \dots, k_N) = \frac{1}{G(K)} \prod_{i=1}^N \left( \frac{\mu_i p_i}{\mu_i} \right)^{k_i}, \quad (8)$$

где

$$G(K) = \sum_{k \in A} \prod_{i=2}^N \left( \frac{\mu_i p_i}{\mu_i} \right)^{k_i}. \quad (9)$$

Если  $A_i$  – стационарная вероятность того, что  $i$ -й узел не пуст. Тогда

$$A_i = \begin{cases} \frac{G(K-1)}{G(K)}, & i=1, \\ \frac{\mu_1 p_i}{\mu_i} A_1, & i=2,3,\dots,N. \end{cases} \quad (10)$$

Из (10) получено, что  $A_1 \mu_1 p_1 = A_i \mu_i$  ( $\geq 2$ ), т.е. интенсивность, с которой задания поступают в  $i$ -й узел, равна интенсивности, с которой они покидают этот узел. Таким образом, установлено, что сбалансированная система – это такая система, для которой выполняется равенство (11), которое показывает, что мера, определяющая, насколько узкое место создается в  $i$ -м узле, пропорциональна скорости изменения производительности в зависимости от роста интенсивности обслуживания в этом узле.

$$\frac{d}{d\mu_i} A_1 \mu_1 p_1 = \frac{d}{d\mu_j} A_1 \mu_1 p_1, \quad 1 \leq i, j \leq N. \quad (11)$$

Рост объемов данных, расширение спектра задач, связанных с их обработкой, критичность доступа к данным приводят к необходимости выбора модели эффективного размещения ресурсов в системах хранения данных ЦОД. В связи с этим, рассмотрена работа некоторой системы хранения, при чем среднее время ответа  $T$  для нее больше, чем требуется и имеются некоторые возможности изменения параметров системы, приводящие к снижению  $T$ . Исследованы следующие случаи:

- совокупность  $m$  ресурсов, каждый из которых имеет пропускную способность  $C/m$  [операций/с], что характерно для случая «новая задача – новый сервер» или систем хранения типа DAS. Моделируется набором из  $m$  СМО  $G|G|1$  с суммарной пропускной способностью  $C$ ;

- одна очередь ко всему набору  $m$  ресурсов с суммарной интенсивностью  $\lambda$ . Моделируется СМО  $G|G|m$ . Подобный случай характерен для использования систем хранения типа NAS (консолидация систем хранения без виртуальных сред);

- система с объединенной очередью и объединенными ресурсами. Моделируется СМО  $G|G|1$  с интенсивностью потока на входе  $m\lambda$  и пропускной способностью ресурса  $mC$ . Такая система может быть реализована путем использования кластера вычислительных устройств с развернутыми виртуальными средами или системой хранения типа SAN.

В третьей главе предложены методы повышения эффективности предоставления информационных услуг в центрах обработки данных: метод управления нагрузкой, поступающей на ЦОД, метод отказоустойчивого размещения задач в кластере серверного комплекса ЦОД, метод распределения памяти в ЦОД.

Эффективность ЦОД  $E_{f_{\text{ЦОД}}}$  представлена функцией (12).

$$E_{f_{\text{ЦОД}}} = F(\Lambda, P_{\text{отк}}, t_{\text{отв}}, P_{\text{пот}}), \quad (12)$$

где  $\Lambda$  – пропускная способность ЦОД;

$P_{\text{отк}}$  – вероятность безотказного функционирования (отказоустойчивость ЦОД);

$t_{\text{отв}}$  – время ответа на запрос пользователя;

$P_{\text{пот}}$  – вероятность потерь ЦОД.

Предложен метод управления нагрузкой, поступающей на ЦОД, который позволяет увеличить  $\Lambda$  и уменьшить  $P_{\text{пот}}$

$$\Lambda_{\text{упр}} = f(\lambda), \quad P_{\text{пот}}^{\text{упр}} = f(\lambda), \quad (13)$$

где  $\Lambda_{\text{упр}}$  – пропускная способность ЦОД при использовании механизма управления,  $\Lambda_{\text{упр}} > \Lambda$ ;

$\lambda$  – интенсивность поступления нагрузки на вход ЦОД;

$P_{\text{пот}}^{\text{упр}}$  – вероятность потерь ЦОД при использовании управления,  $P_{\text{пот}}^{\text{упр}} < P_{\text{пот}}$ .

Нагрузка на ЦОД с течением времени меняется и если выделять необходимые для решения задач пользователей ресурсы только на основании значения пиковой интенсивности, то часть ресурса не будет востребована в процессе выполнения запроса. Следовательно, используя статистические свойства нагрузки, администрация ЦОД может без потери качества осуществить статистическое мультиплексирование нагрузки, что позволяет предоставить суммарный общий кредит всем клиентам, превышающий общую пропускную способность ЦОД. Для обеспечения выполнения SLA-соглашения предлагается выдавать клиентам ЦОД так называемый «кредит», т.е. разрешение на передачу данных со скоростью, не превышающей заданную таким «кредитом». Для уменьшения числа потерянных заявок введен буфер ограниченного объема, куда будут помещаться заявки в случае превышения порогового значения счетчика. Эти ячейки будут повторно поставлены на передачу, когда нагрузка на вход ЦОД уменьшится. Тогда потеря заявок будет происходить только в случае переполнения буфера. Для осуществления такого управления потоками заявок был применен игровой метод, когда управляющее устройство состоит из коллектива вероятностных автоматов, закрепленных за виртуальными каналами, по которым передаются заявки от соответствующего пользователя. Работа автомата задается действиями  $\gamma_1, \dots, \gamma_n$ . Действие означает выбор  $i$ -го приращения к заданному кредиту с вероятностью  $m_i$  из  $n$  возможных при поступлении соответствующей заявки. Выбор действия заявки задается вектором-строкой

$$M = (m_1, \dots, m_n), \quad \text{где } m_i \geq 0$$

$$\sum_{i=1}^n m_i = 1. \quad (14)$$

Изменение структуры автомата  $M$  происходит следующим образом. При поступлении клиентских запросов автомат сравнивает значение трафика с пропускной способностью ЦОД. Если трафик не превышает пропускную способность или превышение такое, что заявки записываются в буфер без потерь, то действие  $\gamma_i$  автомата считается успешным, автомат поощряется, элемент  $m_i$  увеличивается умножением на коэффициент  $\beta \geq 1$  и вся строка нормируется. Если буфер переполнен и поступающие заявки теряются, то автомат штрафуются, при этом элемент  $m_i$  уменьшается умножением на  $0 \leq \alpha \leq 1$  и вся строка нормируется. Если совершенно действие  $\gamma_i$ , то новое значение элементов автомата

$$m_i' = \frac{m_i}{1 + (z - 1)}, \quad (15)$$

где  $z = \alpha$  при штрафе и  $z = \beta$  при поощрении.

Поощрение или штраф соответствуют передаче источнику заявок сигнала об увеличении или снижении скорости обработки.

Предложен метод отказоустойчивого размещения задач в кластере ЦОД, который позволяет снизить  $P_{\text{отк}}$ . Результатом является получение отказоустойчивого плана размещения задач по серверам кластера ЦОД

$$P_{\text{отк}} = f(P_i, Q_j), \quad i=1, \dots, k, \quad j=1, \dots, h; \quad D = f(U, p-N, o-M), \quad (16)$$

где  $P_{\text{отк}}$  – вероятность безотказного функционирования ЦОД;

$P_i$  – вероятность исправного состояния  $i$ -го сервера в кластере ЦОД;

$Q_j$  – вероятность неисправного состояния  $j$ -го сервера в кластере ЦОД;

$D$  – план размещения задач по серверам кластера ЦОД;  
 $U$  – число задач, решаемых в кластере ЦОД одновременно;  
 $p-N$  – количество работоспособных серверов кластера ЦОД;  
 $o-N$  – количество отказавших серверов кластера ЦОД.

Кластер выполняет фиксированное задание  $\Gamma$ , которое представляет собой известное множество задач  $\Gamma = \{U_1, U_j, \dots, U_L\}$  с заданными требованиями к порядку их выполнения и взаимосвязями. Каждая задача  $U_j$  характеризуется следующими показателями:  $b_j$  – вес задачи (величина, определяющая важность задачи  $U_j$  для системы);  $v_j$  – объем оперативной памяти, требуемый для хранения и выполнения задачи;  $\tau_j$  – время выполнения задачи.

В процессе работы кластера возможны отказы серверов. Состояние кластера определяется как  $s_v = \sigma_1, \dots, \sigma_n$ , где  $\sigma_i = 0$ , если  $M_i$  – работоспособный сервер (р-Сервер) и  $\sigma_i = 1$ , если  $M_i$  – отказавший сервер (о-Сервер);  $s^0 = 00 \dots 0$  – начальное состояние кластера; все состояния  $s_v \neq s^0$  называются искаженными.

Постановка задачи отказоустойчивого размещения следующая: для каждого искаженного работоспособного состояния  $s_\omega \in S^\omega$  найти такой план  $D_\omega = \left| d_{ji}^\omega \right|$  распределения задач, при котором достигается максимальное значение функциональной мощности серверного комплекса ЦОД в состоянии  $s_\omega$

$$E_\omega^f = \sum_{U_j \in \Omega_\omega^f} \sum_{M_i \in H_\omega} d_{ji}^\omega b_j \rightarrow \max, \quad (17)$$

где  $\Omega_\omega^f$  – множество собственных задача отказавших серверов;

$H_\omega$  – множество р-Серверов для состояния  $s_\omega$ , при ограничениях:

– на суммарное время выполнения сервером  $M_i$  всех задач, назначенных ему в состоянии  $s_\omega$  (ограничение 1) и

– на объем памяти каждого р-Сервера  $M_i$ , доступный для размещения копий задач, дополнительно назначаемых этому серверу для выполнения в искаженном состоянии  $s_\omega$  (ограничение 2).

Задача (17) относится к классу задач дискретной оптимизации с булевыми переменными, для решения которых могут применяться точные методы математического программирования. Поскольку для многих задач практической сложности применение таких методов затруднено ввиду их чрезвычайно большой трудоемкости, то предложено использовать эвристические методы, дающие не точное, а некоторое «хорошее» решение. К ним относятся различные модели эволюционных вычислений, в частности генетические алгоритмы.

Предложен генетический алгоритм, формирующий план  $D_\omega$ . Последовательность выполнения генетического алгоритма, состоит из следующих шагов.

1. Создание начальной популяции  $P^0$  состоит в выполнении операции инициализации для каждой особи из ее  $N_0$  хромосом, т.е. случайного распределения задач по работоспособным серверам  $p-N$  некоторого данного состояния с одновременной проверкой заданных ограничений на объем памяти сервера и на максимальное суммарное время выполнения задач на каждом сервере.  $N_0$  задается пользователем. Алгоритм инициализации одной хромосомы:

– формирование случайного списка задач;

– выбор очередной задачи из списка, если он не пуст, в противном случае окончание работы алгоритма;

– попытка размещения выбранной задачи в  $i$ -й р-Сервер данного состояния ( $i=1, \dots, g$ ,  $g$  – число р-Серверов), начиная с  $i=1$  при проверке ограничений 1 и 2. Данная задача назначается для

решения в  $i$ -й сервер, если ограничения 1 и 2 для него не нарушены, иначе – попытка поместить задачу в  $(i+1)$ -й сервер; процедура продолжается до наибольшего номера  $p$ -Сервера;

– если данную задачу не удастся разместить ни на один сервер – отбрасывание этой задачи и переход к шагу 2.

2. Селекция (отбор) хромосом для скрещивания. В данной реализации алгоритма выбираем метод пропорциональной селекции хромосом непосредственно из предшествующей популяции (на первой итерации селекция осуществляется из начальной популяции). Из них случайным образом формируются пары различных хромосом, и над каждой из этих пар с заданной вероятностью выполняется операция кроссинговера.

3. Выполнение операторов генетического алгоритма. Над данной парой хромосом с заданной вероятностью  $P_{кр}$  выполняется операция кроссинговера. Реализуется операция следующим образом. Если случайно сгенерированное число  $r < P_{кр}$ , то над данной парой выполняется кроссинговер. Потомки проверяются на ограничения и сравниваются с родительскими хромосомами. Если потомок не удовлетворяет ограничениям или если полезность потомка меньше полезности родительской хромосомы, то потомок отбрасывается и ищется новая родительская пара для кроссинговера. Если  $r > P_{кр}$ , то над данной парой кроссинговер не выполняется, а «несостоявшиеся родители» переходят в следующий родительский пул  $R$  для выполнения следующего генетического оператора – мутации. Если потомок удовлетворяет ограничениям и полезность потомка больше полезности родителя, то потомок переходит в промежуточный родительский пул  $R$ , а родители отбрасываются. В данной реализации генетического алгоритма используется одноточечный кроссинговер. Операция одноточечной мутации выполняется для каждого элемента (бита) каждой хромосомы из популяции  $R$  с заданной вероятностью  $P_{мут}$ .

Пусть  $P_i^c$  – дочерняя популяция, полученная из популяции  $P_i$  в результате кроссинговера её особей и последующей мутации. В следующую (новую) популяцию  $P_{i+1}$  могут попасть как особи только дочерней популяции  $P_i^c$ , так и лучшие особи из дочерней  $P_i^c$  и родительской  $P_i$  популяций.

4. Создание новой популяции  $P_{i+1}$ : хромосомы, полученные в результате операций отбора, кроссинговера и мутации над хромосомами текущей популяции  $P_i$ , переходят в  $P_{i+1}$ .

5. Вычисление качества популяции. Данный этап необходим для ведения статистики и для определения значения критерия останова: после каждого шага генетического алгоритма вычисляется полезность  $F(A_k)$  каждой хромосомы и полезность всей популяции, как сумма всех  $F(A_k)$  при  $k=1 \dots N$ .

6. Замена предыдущей популяции на новую и проверка критерия останова. В реализации генетического алгоритма для данной задачи использованы следующие критерии останова:

а) заданное максимальное количество популяций;

б) заданное максимальное значение  $F(A_k)$  полезности хромосомы – заданный процент от суммы весов всех задач начального множества.

Самая полезная хромосома в популяции, полученная перед остановом, является решением задачи.

Предложен метод распределения памяти в ЦОД, который позволяет сократить число запросов к часто используемым ресурсам и тем самым уменьшить  $t_{отв}$ . Результатом является план распределения памяти с ресурсами по серверам кластера ЦОД

$$t_{отв} = f(Q); Q = f(U, p, \lambda_{ij}), \quad (18)$$

где  $Q$  – план распределения памяти по серверам кластера ЦОД;

$U$  – количество задач, решаемых ЦОД;

$\lambda_{ij}$  – интенсивность запроса  $i$ -го ресурса для выполнения  $j$ -й задачи.

Задача распределения памяти в кластере ЦОД сведена к задаче математического программирования (19).

$$\sum_{j=1}^m a_j u_{ij} \leq c_i, \quad i=1, \dots, n, \quad (19)$$

где  $a_j$  – объем (в битах)  $j$ -го блока информации,  $j = 1, \dots, m$ ;

$u_{ij}$  – двоичная переменная, выражающая наличие ( $u_{ij}=1$ ) или отсутствие ( $u_{ij}=0$ ) в памяти  $i$ -го сервера данных  $j$ -го блока информации;

$c_i$  – объем собственной памяти  $i$ -го сервера  $i = 1, \dots, n$ .

Определена вероятность попадания задачи  $l$ -го потока на  $i$ -й сервер обработки (20) при следующей дисциплине обслуживания: «направлять задачу туда, где больше необходимой информации».

$$p'_{li}(U) = \frac{\sum_{j=1}^m u_{ij} P_{ij}}{\sum_{i,j} u_{ij} P_{ij}}. \quad (20)$$

Определить вероятность того, что задача  $l$ -го потока, направленная на  $i$ -й сервер обработки (21), найдет там необходимую для решения информацию, т.е. не придется обращаться к хранилищу данных.

$$p''_{li} = p_{li}(U). \quad (21)$$

Располагая интенсивностями потоков решаемых задач  $\lambda_1, \dots, \lambda_k$  и производительностями каждого сервера  $\mu_1, \dots, \mu_n$ , сформулирована задача определения оптимального расположения блоков информации по серверам кластера ЦОД, минимизирующая обращаемость к хранилищу данных в процессе функционирования ЦОД:

$$Q(U) = \sum_{i=1}^n \sum_{l=1}^k \lambda_{li} [1 - p''_{li}(U)] p'_{li}(U) \rightarrow \min_{U \in S} \rightarrow U^*, \quad (22)$$

$$\text{где } S: \begin{cases} \sum_{l=1}^k \lambda_{li} p_{li}(U) < \mu_i \\ \sum_{i=1}^m a_i u_{ij} \leq c_i \end{cases},$$

$U$  – двоичный вектор, определяющий распределение блоков памяти по кластерам ЦОД,  $U = (u_{11}, \dots, u_{nm})$ .

Четвертая глава посвящена апробации предложенных в диссертационной работе моделей и методов повышения эффективности предоставления информационных услуг в ЦОД.

Выполнен вычислительный эксперимент по оценке производительности ЦОД и запасов по пропускным способностям. Результаты эксперимента показали, что разные классы запросов требуют разного времени обслуживания в ЦОД. Модель оценки производительности ЦОД учитывает структуру входного потока заявок и соответствует реальным SLA-соглашениям, в которых задаются ограничения на качество обслуживания для каждой информационной услуги.

Приведена оценка влияния балансировки нагрузки на производительность ЦОД. Результаты вычислительного эксперимента показывают, что для устранения возможности появления узкого места необходим сервис балансировки нагрузки:

– с ростом нагрузки наличие узкого места не позволяет кластеру в целом справиться с поступающими запросами пользователя, в то время как кластер с сервисом балансировки нагрузки продолжает функционировать;

Эксперимент по сравнению вариантов размещения ресурсов в системах хранения данных ЦОД показал, что:

– система типа DAS эффективна только для офисных ЦОД;

– системы типа NAS эффективнее систем типа DAS, но не позволяют использовать мощности свободных ресурсов для ускорения работы остальных занятых ресурсов;

– система хранения данных типа SAN позволяет эффективно разделять имеющиеся ресурсы путем использования кластера с развернутыми виртуальными средами.

Представлена оценка влияния механизма управления обработкой запросов пользователей на пропускную способность ЦОД и на вероятность потерь заявок пользователей ЦОД. Исследования показали, что применение механизма управления позволяет увеличить пропускную способность ЦОД на 15-20%. При большем числе входящих виртуальных каналов, можно получить большее увеличение пропускной способности.

Приводится описание программного обеспечения, разработанного на языке C++, которое реализует генетический алгоритм, предложенный для решения задачи отказоустойчивости кластеры ЦОД. Машинный эксперимент по оценке отказоустойчивости ЦОД выполнен для двух вариантов: без учета приоритета задач пользователя и обработка с приоритетом. Результатом эксперимента является:

а) количество серверов в кластере, необходимое для выполнения заданий пользователя с обеспечением оптимального времени обслуживания заданий;

б) план отказоустойчивого размещения задач пользователя.

Результаты проведенных экспериментов свидетельствуют о высокой степени доверия к предложенным моделям и методам повышения эффективности предоставления информационных услуг в ЦОД.

В заключении сформулированы основные результаты исследования, представленные в диссертационной работе.

## **ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ**

В рамках диссертационной работы проведено исследование и получено решение важной научно-технической задачи разработки моделей и методов повышения эффективности предоставления информационных услуг в центрах обработки данных.

Основные результаты работы заключаются в следующем.

1) Предложена структурно-функциональная модель центра обработки данных, включающая описания:

инфраструктуры ЦОД на уровне структур организации серверного комплекса и технологий организации хранения данных;

характеристик качества обслуживания клиентов ЦОД;

вариантов применения ЦОД;

2) Разработан комплекс моделей оценки качества предоставления информационных услуг в центрах обработки данных, включающий:

модель оценки характеристик производительности ЦОД и запасов по пропускным способностям, в том числе с разделением запросов на классы;

модель балансировки нагрузки в кластере серверного комплекса ЦОД;

модель размещения ресурсов в системах хранения данных ЦОД.

3) Предложены методы повышения эффективности предоставления информационных услуг в центрах обработки данных:



метод управления нагрузкой, поступающей на ЦОД;  
метод обеспечения отказоустойчивости серверного комплекса ЦОД;  
метод распределения памяти в кластере ЦОД.

4) Полученные экспериментальные соотношения вероятностно-временных характеристик показывают, влияние различных внешних и внутренних факторов на эффективность предоставления информационных услуг в ЦОД. Эти соотношения подтверждают теоретические положения, выносимые на защиту.

## **СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ**

### *Публикации в изданиях, рекомендованных ВАК России:*

1. Воробьев А.И., Колбанев М.О., Нестеренко В.Д. Анализ характеристик инфокоммуникационной системы с неординарным потоком заявок на обслуживание// Труды учебных заведений связи, 2006. № 175. С.31-36.

2. Воробьев А.И., Колбанев М.О., Татарникова Т.М. Модель балансировки нагрузки в вычислительном кластере центра обработки данных// Информационно-управляющие системы, №3, 2012. С. 37-41.

3. Советов Б.Я., Воробьев А.И. Применение методов оптимизации в задачах структуризации корпоративного центра обработки данных// Известия СПбГЭТУ «ЛЭТИ» №8, 2012. С. 41-46.

### *Другие статьи и материалы конференций:*

4. Воробьев А.И., Колбанев М.О., Матяшов П.А. Обзор технологий абонентского доступа инфокоммуникационных сетей// Санкт-Петербургская международная конференция «Региональная информатика», 1-3 марта 2005. – СПб, 2005.

5. Воробьев А.И. Оценка вероятностей блокировки заявок в мультисервисных сетях// 59-я НТК студентов, аспирантов и молодых специалистов СПбГУТ им. проф. М.А. Бонч-Бруевича, 23 - 27 мая 2005 г. – СПб., 2005.

6. Воробьев А.И., Колбанев М.О., Матяшов П.А. Модель в мультисервисной сети абонентского доступа// Вторая всероссийская конференция «Абонентский доступ в сетях следующего поколения», 2005. – СПб., 2005

7. Воробьев А.И., Колбанев М.О., Матяшов П.А. Анализ характеристик инфокоммуникационной системы с неординарным потоком заявок на обслуживание// 58-я НТК профессорско-преподавательского состава научных сотрудников и аспирантов СПбГУТ им. проф. М.А. Бонч-Бруевича, 23 - 27 января 2006 г. – СПб., 2006.

8. Воробьев А.И. Метод расчёта запаса по пропускным способностям центра обработки данных// XII международная конференция «Информатика: проблема, методология, технологии», Воронеж, 9-12 февраля 2012 г.

9. Воробьев А.И., Татарникова Т.М. Реализация сервиса балансировки нагрузки в центре обработки данных// XII международная конференция «Информатика: проблема, методология, технологии», Воронеж, 9-12 февраля 2012 г.

10. Воробьев А.И., Татарникова Т.М., Яготинцева Н.В. Применение генетического алгоритма для решения задачи обеспечения отказоустойчивости вычислительного кластера// Межвузовский сборник научных трудов: Управление, экономика, транспорт, право. №1(9), 2011 г. С. 3-9.

11. Воробьев А.И., Колбанев М.О., Коршунов И.Л., Омелян А.В. Предложения по развитию ИТ-инфраструктуры образовательной среды СПбГУСЭ// XIII международная конференция «Региональная информатика (РИ-2012)». Санкт-Петербург, 24-26 октября. СПб.: СПОИСУ, 2012. С. 303.

12. Воробьев А.И., Колбанев М.О., Омелян А.В. Современный подход к организации учебного процесса в вузе// XIII международная конференция «Региональная информатика (РИ-2012)». Санкт-Петербург, 24-26 октября. СПб.: СПОИСУ, 2012. С. 303-304.

Подписано в печать 22.11.2012. Формат 60x84/16  
Отпечатано с готового оригинал-макета в типографии ЗАО «КопиСервис».  
Печать ризографическая. Заказ № 1/1122.  
П. л. 1.0. Уч.-изд. л. 1.0. Тираж 100 экз.

---

ЗАО «КопиСервис»  
Адрес: 197376, Санкт-Петербург, ул. Проф. Попова, д. 3.  
тел.: (812) 327 5098