

На правах рукописи

Задорожный Владимир Николаевич

**МЕТОДЫ АНАЛИТИКО-ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ
СИСТЕМ С ОЧЕРЕДЯМИ И СТОХАСТИЧЕСКИХ СЕТЕЙ**

Специальность: 05.13.18 – «Математическое моделирование,
численные методы и комплексы программ»

Автореферат диссертации
на соискание ученой степени
доктора технических наук

Санкт-Петербург – 2011

Работа выполнена в Государственном образовательном учреждении высшего профессионального образования «Омский государственный технический университет»

Научный консультант: доктор технических наук, профессор
Никонов Александр Васильевич

Официальные оппоненты: доктор технических наук, профессор
Григорьев Юрий Дмитриевич
доктор технических наук, профессор
Рыжиков Юрий Иванович
доктор технических наук, профессор
Шапцев Валерий Алексеевич

Ведущая организация: Омский филиал института математики
им. С.Л. Соболева сибирского отделения РАН

Защита диссертации состоится 16 ноября 2011 г. в 15 час. 00 мин. на заседании совета по защите докторских и кандидатских диссертаций Д 212.238.01 Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» им. В.И. Ульянова (Ленина) по адресу: 197376, Санкт-Петербург, ул. проф. Попова, 5.

С диссертацией можно ознакомиться в библиотеке университета.

Автореферат разослан «___» _____ 2011 г.

Ученый секретарь совета по защите докторских
и кандидатских диссертаций
Д 212.238.01

Н. Л. Щеголева

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Информационные технологии стали в последние десятилетия одним из основных ресурсов и инструментов развития ведущих стран мира. Развитие цифровых технологий, вычислительных систем и сетей особенно важно для Российской Федерации с ее обширной территорией и богатым природным потенциалом. Непрерывная модернизация информационных технологий и повсеместное внедрение компьютерных сетей требуют опережающего развития инженерных методов и теории, направленных на сокращение сроков и повышение качества проектирования информационно-вычислительных сетей (ИВС). Теория вычислительных систем, формируемая в последние десятилетия, решает задачи эффективного комплексирования и использования аппаратных, информационных и программных ресурсов, и основывается на концептуальных и математических моделях, учитывающих стохастический характер поступления данных и запросов на их обработку, а также недетерминированное время обработки запросов вычислительными узлами и каналами связи. Математической базой теории вычислительных систем является теория массового обслуживания (ТМО), позволяющая решать разнообразные задачи анализа и синтеза ИВС путем определения технико-экономических показателей эффективности функционирования систем в целом при известных технических параметрах их отдельных узлов. Широкую известность в последние годы приобрели фундаментальные учебники по теории вычислительных систем и компьютерных сетей, использующие ТМО (*Л. Клейнрок, 1976; С.А. Майоров, Г.И. Новиков, Т.И. Алиев и др., 1978; Д. Феррари, 1978, Ю.И. Рыжиков, 1996; В.М. Вишневецкий, 2003*). Решаемые методами ТМО задачи включают расчет вероятностно-временных характеристик функционирования центральных процессоров и узлов коммутации, анализ буферной памяти и алгоритмов маршрутизации, расчет потерь данных и загрузки линий связи, обеспечение требуемого времени ответа ИВС на запросы конечных пользователей и т.д.

Адекватными математическими моделями ИВС и их узлов в ТМО являются **системы с очередями**, т.е. системы массового обслуживания (СМО) и сети массового обслуживания (СеМО). В настоящее время изучено большое число СМО и СеМО с различными законами поступления заявок, структурами систем и способами обслуживания. Разработаны точные аналитические методы для решения задач расчета, анализа и синтеза широкого класса СМО и ряда СеМО. Однако предположения, лежащие в основе точных аналитических методов и связанные с применением преобразований Лапласа-Стилтьеса или производящих функций, выполняются далеко не во всех важных для задач проектирования ИВС случаях. Приближенные аналитические методы ТМО также не всегда реализуемы, и могут приводить к неприемлемым погрешностям, либо к чрезмерному объему вычислений. В качестве примера, типичного для задач проектирования ИВС, можно привести одноканальную или многоканальную СМО с приоритетами и не экспоненциальными законами распределения вероятностей, характеризующими интервалы поступления заявок и время их обслуживания. Использование в подобных случаях приоритетных дисциплин еще более усложняет дело. Так, если в процессорных узлах ИВС используется типичная для них дисциплина абсолютных приоритетов,

то даже при пуассоновском потоке задач аналитическое выражение средних длин очередей к процессорам дается в ТМО лишь для однопроцессорных узлов. Особенно трудно поддаются аналитическим методам задачи оптимизации немарковских СеМО. Разнообразие вероятностных распределений, описывающих время обслуживания заявок в узлах, и корреляция интервалов между заявками в потоках, смешиваемых на входах узлов, делают оптимизацию немарковских сетей аналитическими методами практически невозможной (за исключением, быть может, специальных частных случаев).

Ограниченные возможности применения точных и приближенных аналитических методов ТМО в задачах проектирования ИВС приводят к широкому использованию современных технологий и программных средств имитационного моделирования (ИМ). При этом ИМ, облегчая широкое внедрение в практику проектирования ИВС формализмов и концепций ТМО, приводит, в то же время, к ряду специфических вычислительных проблем, обусловленных стохастическим характером отклика имитационных моделей и необходимостью корректной обработки этого отклика (например, в целях использования градиентных алгоритмов оптимизации). Проблема использования градиентов – это лишь одна из широко известных общих проблем ИМ систем с очередями. Другие известные проблемы связаны с существованием «критических точек» – таких значений параметров моделируемой системы, вблизи которых происходит резкий рост затрат машинного времени на моделирование. Типичные примеры попадания в окрестности критических точек – это ИМ СМО с близкими к единице коэффициентами загрузки или ИМ событий, имеющих малую вероятность. Подобная ситуация возникает и при моделировании приоритетных систем, в которых интенсивности потоков заявок разных классов различаются на порядки (проблема разномасштабных интенсивностей). Затраты машинного времени возрастают в этой ситуации также на порядки, поскольку для накопления необходимой статистики по классу редко поступающих заявок приходится вместе с ними имитировать и на порядки большее количество заявок других классов.

Наряду с классическими задачами теории ИВС, решаемыми на базе формализмов ТМО, в последнее десятилетие возникло большое число задач иной природы – задач, связанных с беспрецедентным ростом глобальных информационных сетей, в первую очередь – сети Интернет. Быстро развивающиеся большие **стохастические сети**, подобные сети Интернет, приобретают специфические свойства, обусловливаемые взаимодействием детерминированных «правил» и стохастических факторов роста сетей. Необходимость адекватного математического описания структурных особенностей стохастических сетей, состоящих из сотен тысяч и миллионов узлов и связей, привела к созданию и быстрому в течение последних лет развитию теории случайных неклассических графов. Ее наиболее развитую ветвь представляет теория сетей (графов) предпочтительного связывания. Такие графы выращиваются из небольших графов-затравок с помощью простых алгоритмов (генераторов), воспроизводящих способы развития моделируемых сетей. Наиболее широко известен граф Барабаши-Альберт (БА-граф), растущий неограниченно в результате циклического добавления к нему приращенного графа – новой вершины с фиксированным числом m инцидентных ей ребер

(Barabási A., Albert R., 1999). Графы БА обладают свойствами, принципиально отличающими их от классических случайных графов Эрдеша и Реньи (Erdős P., Rényi A., 1959). На основе графов БА аналитическими и/или имитационными методами исследуют разнообразные позитивные и негативные процессы, происходящие в реальных сетях, и разрабатывают оптимальные стратегии влияния на эти процессы. Однако, несмотря на лавинообразный рост числа публикаций, посвященных сетям предпочтительного связывания, нерешенными остаются многие важные вопросы, требующие развития аналитической теории таких сетей.

Таким образом, в целом в современной теории и инженерной практике проектирования ИВС широко и успешно используются аналитическое и имитационное моделирование ИВС как *систем с очередями* и как *стохастических сетей*. Однако возникающие при этом проблемы, перечисленные выше, существенно затрудняют развитие методов проектирования ИВС в соответствии с растущими потребностями практики и обуславливают острую необходимость разработки методов для эффективного решения этих проблем.

В развитие методов аналитико-имитационного моделирования (АИМ) СМО и СеМО большой вклад внесли многие отечественные и зарубежные авторы: В.М. Вишнеvский, В.А. Жожикашвили, Д.Л. Иглхарт, В.В. Калашников, Дж. Клейнен, И.Н. Коваленко, Н.Ю. Кузнецов, О.И. Кутузов, Б.И. Плакс, Ю.Г. Полляк, Ю.И. Рыжиков, А.Л. Толмачев, Д.С. Шедлер, R.V. Gabriel, M. Reinaldo, R.Y. Rubinstein, R. Suri, M. Zazanis и другие. Методы АИМ СМО и СеМО основываются на ТМО, связанной с именами А.К. Эрланга, А.А. Боровкова, Б.В. Гнеденко, Л. Клейнрока, И.Н. Коваленко, А.Н. Колмогорова, Ф. Поллачека, Ю.И. Рыжикова, Т. Саати, А.Я. Хинчина и других ученых.

В развитие методов АИМ больших стохастических сетей значительный вклад внесли О.И. Кутузов, Ю.Л. Павлов, Ю.Ю. Тарасевич, А.Л. Эфрос, R. Albert, A. Barabási, B. Bollobas, S.N. Dorogovtsev, P. Erdős, P.L. Krapivsky, J. Leskovec, M.E.J. Newman, D. Price, S. Redner и A. Rényi.

Объектами исследования в диссертации являются ИВС и другие технические системы, выполняющие массовую обработку заявок (материалов, данных, денежных средств, документов, клиентов, транспорта и т.д.), представимые в виде СМО и СеМО, а также большие стохастические сети (БСС) типа Интернет.

Предмет исследования: приоритетные СМО, однородные немарковские СеМО, графы предпочтительного связывания и *методы их АИМ*, разрабатываемые для решения задач, возникающих в проектировании ИВС.

Цель работы: разработка, обоснование, исследование, тестирование и программная реализация математических моделей и методов АИМ, предназначенных для эффективного решения задач анализа и синтеза СМО, СеМО и БСС, возникающих в теории и инженерной практике проектирования ИВС.

Задачи исследования:

1. Разработка ускоренных аналитико-имитационных методов моделирования СМО с дисциплиной абсолютных приоритетов при существенно различающихся интенсивностях поступления заявок разных приоритетных классов.

2. Вывод точных и асимптотических соотношений, позволяющих существенно (на порядки) ускорять ИМ при решении задач анализа и синтеза приоритетных

СМО и сводить эти задачи или отдельные этапы их решения к задачам, решаемым аналитическими или элементарными численными методами.

3. Разработка эффективных аналитико-имитационных градиентных и квази-градиентных методов оптимизации однородных немарковских СеМО.

4. Тестирование разработанных методов и алгоритмов АИМ СМО и СеМО, разработка рекомендаций по их использованию и их программная реализация.

5. Разработка эффективных методов АИМ БСС. Развитие теории графа БА.

6. Создание основ общей теории случайных графов с нелинейным правилом предпочтительного связывания (графов с НППС).

Графы с НППС включают графы БА в качестве частного случая. Теория графов с НППС должна решать в общем виде задачу анализа и задачу синтеза генераторов, выращивающих графы с требуемыми структурными характеристиками, соответствующими структурным характеристикам моделируемых реальных БСС. Основы теории графов с НППС требуется реализовать в виде комплекса эффективных программных средств для моделирования реальных больших сетей.

Методы исследования. Решение поставленных в работе задач осуществляется методами ТМО, теории восстановления, теории вероятностей, математической статистики, теории случайных графов, численными методами оптимизации, исследования операций и аналитико-имитационного моделирования.

Основные положения и результаты, выносимые на защиту:

1. *Методы асимптотического анализа цикла обслуживания неприоритетной заявки*, основанные на изучении вложенных в цикл обслуживания процессов восстановления и сопряженных с ними процессов накопления. На защиту выносятся два метода – метод суммирования периодов занятости (метод СПЗ) и метод суммирования периодов обслуживания (метод СПО).

2. *Метод декомпозиции систем $GI_2|GI_2|n$ и $GI_2|GI_2|n||$ с абсолютными приоритетами на бесприоритетные системы* и полученные на его основе аналитические выражения, ускоренные методы моделирования и новый класс моделей с очередями – многоканальные циклические системы обслуживания (ЦСО) с пакетированием неприоритетных заявок. (Нижний индекс 2 в обозначениях СМО указывает на наличие заявок двух классов – приоритетных и неприоритетных. В системах $GI_2|GI_2|n||$ неприоритетная работа разделяется между всеми n каналами).

3. *Двухуровневые двухэтапные аналитико-имитационные методы оптимизации* однородных немарковских СеМО. На защиту выносятся входящий в состав этих методов оригинальный метод «направляющих гипербол», а также аппроксимационный ускоренный метод оптимального распределения каналов по узлам СеМО и расширенный метод редукции (РМР) графов полумарковских процессов.

4. *Новые фундаментальные аналитические результаты, развивающие теорию графа БА.* В их число входят: уравнения баланса для вероятностных характеристик стационарного графа БА; точное общее решение уравнений баланса; аналитическое выражение для совместного распределения вероятностей концевых степеней дуги (ребра) графа БА; маргинальные распределения вероятностей концевых степеней дуг/ребер графа БА; асимптотически степенной с показателем 0.5 закон роста максимальной степени вершин в ходе эволюции графа БА; аналитические выражения для коэффициента кластеризации графа БА; метод сепарабель-

ной калибровки графа БА по коэффициенту кластеризации.

5. *Основы общей теории случайных графов с НППС*, включающие: уравнения баланса для стационарного распределения вероятностей локальной степени вершин; точные решения этих уравнений; ускоренный метод генерации случайных графов с НППС; точное распределение степени вершин в графах с НППС; методы калибровки генераторов графов при фиксированном и при случайном числе ребер в приращениях графа; методика калибровки генераторов графов по эмпирическим распределениям степени связности; рекомендации по практическому использованию графов с НППС; рекомендации по дальнейшему развитию теории.

Научная новизна диссертационной работы заключается в следующем.

1. При решении проблемы ИМ процессов с разномасштабными интенсивностями применительно к СМО $GI_2|GI_2|n$ с абсолютными приоритетами и дообслуживанием заявок выявлены и изучены две системы процессов восстановления и накопления, вложенных в цикл обслуживания неприоритетной заявки. На их основе путем использования известных в теории восстановления асимптотических результатов (Cox D.R., Lewis P.A.W., 1966) разработан для анализа цикла обслуживания метод суммирования периодов занятости (СПЗ) и метод суммирования периодов обслуживания (СПО). Метод СПЗ имеет сходство с методом Гейвера (Gaver D.P.), но свободен от предположения о пуассоновском характере потоков заявок. Это существенно расширяет область его практического применения. Метод СПО является развитием метода СПЗ и позволяет выразить асимптотические характеристики цикла обслуживания через известные параметры интервалов поступления и времени обслуживания заявок приоритетных классов.

2. Обнаруженные новые возможности применения теории восстановления к анализу цикла обслуживания неприоритетной заявки позволили найти ряд неизвестных ранее асимптотических и точных выражений для систем $GI_2|GI_2|n$ с абсолютными приоритетами и для систем $GI|GI|n$. В отличие от *традиционных методов*, основанных на преобразованиях Лапласа-Стилтьеса и производящих функций, методы СПЗ и СПО не связаны с предположениями об экспоненциальном распределении интервалов поступления и/или времени обслуживания заявок.

3. Разработанный на основе методов СПЗ и СПО метод приближенной декомпозиции системы $GI_2|GI_2|n$ на вложенную систему, сохраняющую только приоритетный поток заявок, и фоновую систему, сохраняющую только неприоритетный поток заявок, эффективно решает проблему разномасштабных интенсивностей при ИМ систем $GI_2|GI_2|n$. По сравнению с известным методом усреднения вклада интенсивных прерываний (Максимей И.В., 1979) значительно снижается погрешность результатов моделирования (благодаря учету моментов второго порядка).

4. Посредством разработанных методов декомпозиции для систем $GI_2|GI_2|n$ и $GI_2|GI_2|n||$ найден ряд аналитических приближений, получаемых на основе решений, известных для бесприоритетных систем $GI|GI|1$ и $GI|GI|n$. В частности, получены и протестированы приближенные формулы для расчета средней длины очереди неприоритетных заявок для случая сочетания разномасштабных интенсивностей с высокой нагрузкой (в системе $GI_2|GI_2|n||$) и для случая пуассоновского потока неприоритетных заявок (в системе $GI_2|GI_2|1$).

5. Разработанные в диссертации методы декомпозиции *обобщены* на системы

$GI_k|GI_k|n$ (с k классами заявок), в том числе на системы с вложенными прерываниями. Тестирование обобщенных методов показало их приемлемую точность и хорошие перспективы их развития применительно к случаям наличия в системах $GI_k|GI_k|n$ интенсивных прерываний.

6. Логика развития разрабатываемых методов АИМ приоритетных СМО привела к разработке и исследованию в диссертации *нового класса систем с очередями*, названных циклическими системами обслуживания. Для ЦСО найдены асимптотические выражения первичных технико-экономических показателей и разработаны эффективные методы АИМ. Эти системы отражают особенности функционирования многих реальных систем обслуживания и характеризуются наличием циклов накопления пакетов неприоритетных заявок для последующего их обслуживания в фоновом режиме.

7. Для решения задачи оптимального распределения непрерывного ресурса (стоимости, производительности) по узлам однородной немарковской СеМО разработан *двухуровневый аналитико-имитационный градиентный метод оптимизации*, ядром которого является оригинальный метод «*направляющих гипербол*» (НГ). Метод НГ использует настраиваемую по результатам ИМ локальную нелинейную аналитическую аппроксимацию поверхности отклика, что позволяет достичь принципиальных преимуществ по сравнению с известными методами оптимизации. Разработанный двухуровневый метод решает актуальную задачу минимизации среднего времени ответа СеМО. Решаемая задача отличается от известной задачи оптимизации замкнутых однородных марковских СеМО (*Жожикивили В.А., Вишневецкий В.М., 1988*) общностью постановки: оптимизируются не только замкнутые и не только марковские СеМО.

8. Разработанный расширенный метод редукции графа случайных задержек позволяет рассчитывать математическое ожидание и дисперсию длительности описываемого графом полумарковского процесса с марковской цепью, вложенной в моменты смены состояний. От прототипа – метода Байцера (*Beizer B., 1970*) он *отличается встроенным в ход редукции точным расчетом коэффициентов чувствительности* вычисляемых характеристик к вариациям параметров графа, включая вариации переходных вероятностей. Агрегирование двухуровневого метода оптимизации СеМО с РМР позволяет определять чувствительность оптимальных решений к возмущениям параметров оптимизируемых СеМО.

9. Разработанная двухуровневая аналитико-имитационная структура процесса оптимизации применена к другой задаче – *задаче оптимального распределения каналов по узлам однородной немарковской СеМО*. В результате получены эффективные методы и для решения этой задачи. Оценка перспектив распространения двухуровневой аналитико-имитационной структуры процесса оптимизации на решение других задач позволяет охарактеризовать эту структуру как основу и предмет самостоятельного направления исследований в области АИМ.

10. В отличие от известных асимптотических (*Barabási A., Albert R., 1999*) и частных точных подходов (*Dorogovtsev S.N., Mendes J.F.F., Samukhin A.N., 2000; Krapivsky P.L., Redner S., 2001*) в диссертации для определения точного совместного распределения конечных степеней ребра графа БА *используется техника составления уравнений баланса и их решения в общем виде*. На основе найденного

решения впервые получены аналитические выражения ряда важнейших характеристик, которые ранее рассчитывались трудоемкими приближенными методами ИМ. Разработан метод калибровки графа БА по коэффициенту кластеризации.

11. Предложенная в диссертации *общая теория случайных графов с НППС полностью оригинальна* и содержит множество новых результатов, часть которых в качестве своих частных случаев охватывает ранее известные или впервые полученные в диссертации результаты, установленные для графа БА.

Личный вклад соискателя. В список положений, выносимых на защиту, включены лишь те результаты, которые получены соискателем единолично.

Теоретическая значимость. Полученные в работе теоретические результаты являются существенным вкладом в развитие двух научных направлений:

1) в разработку эффективных аналитико-имитационных методов решения задач анализа и синтеза *сложных систем с очередями*;

2) в разработку эффективных аналитико-имитационных методов исследования *больших стохастических сетей*.

В рамках первого направления в диссертации получены новые аналитические решения, представляющие самостоятельную ценность для ТМО, разработаны эффективные методы АИМ и определены перспективы развития полученных результатов. В рамках второго направления существенно усилена новыми точными методами и решениями теория графа БА, разработаны основы общей теории случайных графов с НППС и определены перспективы ее дальнейшего развития.

Достоверность и обоснованность результатов подтверждается строгими математическими выкладками, а также численными и имитационными тестами. Работоспособность и эффективность предложенных методов подтверждается имитационным моделированием и численными примерами. Пригодность разработанной общей теории случайных графов с НППС, ее преимущества и высокая эффективность ее методов подтверждены имитационным моделированием реальных сетей, содержащих сотни тысяч узлов. Формальными проверками новых аналитических результатов, полученных в диссертации при исследовании СМО, СеМО, графа БА и графов с НППС установлено, что эти результаты включают в качестве частных случаев множество ранее известных результатов.

Практическая ценность работы заключается в следующем.

1. Ускоренные алгоритмы моделирования систем $GI_2|GI_2|n$ и $GI_2|GI_2|n|$ позволяют в случае значительно различающихся интенсивностей у потоков разных приоритетных классов сокращать на порядки затраты машинного времени на имитационные эксперименты.

2. Соотношения, полученные в результате асимптотического анализа ЦСО, позволяют рассчитывать показатели качества соответствующих реальных СМО, не прибегая к ИМ. В частности, когда циклы накопления пакетов заданы фиксированной длительностью, эти соотношения можно использовать для анализа и оптимизации систем со сменной, суточной или сезонной организацией работы. Задачи оптимизации ЦСО можно на основе этих соотношений сводить к алгебраическим уравнениям, решаемым элементарными численными методами.

3. Двухуровневые аналитико-имитационные методы оптимизации однородных немарковских СеМО позволяют с высокой точностью находить оптимальные ре-

шения для сетей, состоящих из сотен узлов и тысяч каналов. Это открывает новые возможности для оптимизации проектов ИВС.

4. Расширенный метод редукции и двухуровневый метод оптимизации СеМО могут эффективно использоваться в задачах проектирования структур ИВС, разработки программного обеспечения ИВС, маршрутизации сообщений и во многих других задачах, в которых исследуемый процесс может быть представлен полумарковским процессом с вложенной марковской цепью.

5. Новые аналитические результаты, развивающие теорию графа БА, позволяют находить аналитические выражения для ряда характеристик графа, которые ранее описывались лишь асимптотическими (по степени k связности вершин) выражениями, или оставались неизвестными (закон роста максимальной степени вершин, коэффициент кластеризации и др.). Поскольку граф БА широко используется для моделирования быстро растущих сетей типа сети Интернет, это позволяет углубить и уточнить практические результаты, получаемые во всех соответствующих приложениях графа БА.

6. Теория случайных графов с НППС дает инструменты калибровки случайных графов, используемых для моделирования реальных сетей. Применяя такие калиброванные модели, можно прогнозировать развитие этих сетей, исследовать методом ИМ разнообразные процессы, происходящие или планируемые для реализации в этих сетях, оптимизировать стратегии влияния на эти процессы, эффективно противостоять негативным сетевым процессам.

Практическая ценность диссертационной работы значительно усиливается тем обстоятельством, что разработанные в ней математические модели, аналитико-имитационные и численные методы могут быть использованы не только в проектировании ИВС. Их можно использовать и для исследования множества других систем, которые могут рассматриваться как системы с очередями или как БСС. Как системы с очередями можно рассматривать транспортные сети, связь, производственные и торговые комплексы, коммунальные службы и службы здравоохранения, учреждения юстиции, науки и образования, оборонные структуры, системы управления объектами с потенциально опасной технологией и т.д. Теория графов с НППС позволяет воспроизводить в моделях наиболее существенные свойства таких сетей, как сети социальных контактов, сети сотовой связи, сети цитирования, сети финансовых потоков, сети террористических организаций, сети участия белков в химических реакциях организма человека и т.д.

Реализация и внедрение результатов работы. Тематика научных исследований, выполненных в диссертации, связана с проводившимися в ОмГТУ научно-исследовательскими работами: «Система имитационного моделирования вычислительных комплексов (СИМВК)» (СМОФАП Киевского ПКБ АСУ, № ГОСФАП ПОО7643), «Система имитационного моделирования вычислительного процесса в АСУ (научно – исследовательская тема № 546, № ГР 76090017, № инв. Б853603), «Анализ эффективности организации банка данных в АСУ реального времени» (№ ГР 81088775, инв. № 02860047403), «Разработка и исследование моделей и методов оптимизации оперативного управления механообрабатывающим участком в условиях АСУ» (№ ГР 01. 0031165), «Методология управления нелинейными динамическими фазовыми системами и информационно-вычислительными

комплексами в условиях неопределенности» (инв. № 022.007 01269) и др. Полученные в диссертации результаты нашли практическое применение на следующих предприятиях: ООО «АйПро», г. Омск; Омский филиал института математики РАН; ООО «Фарфалле» и используются в учебном процессе при чтении лекций, в курсовом и дипломном проектировании, проведении студенческих НИР и в лабораторных работах на кафедре «Автоматизированные системы обработки информации и управления» ОмГТУ. Издано 7 учебных пособий, в том числе 2 пособия с рекомендациями УМО вузов по университетскому политехническому образованию, и ряд методических указаний. Практическое использование результатов диссертационной работы подтверждено соответствующими актами о внедрении, представленными в приложении к диссертации.

Публикации. По теме диссертации опубликовано более 60 научных работ. Основные научные результаты представлены в 57 публикациях, в число которых входят: 2 монографии, 20 статей в журналах «Перечня ведущих рецензируемых научных журналов и изданий», 12 статей в других изданиях, 3 свидетельства программ ЭВМ, 20 работ в материалах международных и всероссийских научно-технических конференций.

Апробация работы. Основные научные положения и результаты диссертационной работы докладывались, обсуждались и были одобрены на следующих международных, всероссийских и региональных конференциях: II Всесоюзной конференции по перспективам и опыту внедрения статистических методов в АСУ ТП, Смоленск, 11-13 мая 1984; Всесоюзном научно-техническом семинаре «Информационное обеспечение автоматизированных систем управления нефтеперерабатывающих предприятий», Омск, 18-20 сентября 1984; Региональной научно-практической конференции «Перспективные направления развития автоматизированных систем управления и их компонентов», Омск, 1989; Межрегиональной научно-практической конференции «Омский регион: исторический опыт, проблемы и пути экономического развития в современных условиях», Омск, 1994; Межрегиональной научно-практической конференции «Омский регион: исторический опыт, проблемы и пути экономического развития в современных условиях», Омск, 1999; V Международной научно-технической конференции «Динамика систем, механизмов и машин», 16-18 ноября 2004; Второй Всероссийской научно-практической конференции по имитационному моделированию и его применению в науке и промышленности «Имитационное моделирование. Теория и практика» (ИММОД-2005), Санкт-Петербург, 19-21 октября 2005; Конференции-конкурсе работ студентов, аспирантов и молодых ученых «Технологии Microsoft в теории и практике программирования», Новосибирск, 22-24 февраля 2006; Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР-2006», 4-7 мая 2006; Третьей всероссийской научно-практической конференции по имитационному моделированию и его применению в науке и промышленности «Имитационное моделирование. Теория и практика» (ИММОД-2007), Санкт-Петербург, 17-19 октября 2007; Международной конференции «Идентификация систем и задачи управления» (SICPRO'08), Москва, 28-31 января 2008, Институт проблем управления им. В.А. Трапезникова РАН; Всероссийской научно-технической конференции «Россия молодая: передо-

вые технологии – в промышленность», Омск, 2008; Межвузовской научно-практической конференции «Информационные технологии и автоматизация управления», Омск, 20-24 апреля 2009; Четвертой всероссийской научно-практической конференции по имитационному моделированию и его применению в науке и промышленности «Имитационное моделирование. Теория и практика» (ИММОД-2009), Санкт-Петербург, 21-23 октября 2009; VII Международной научно-технической конференции «Динамика систем, механизмов и машин» (10-12 ноября 2009); II Межвузовской научно-практической конференции «Информационные технологии и автоматизация управления», Омск, 20-24 апреля 2010; III Региональной научно-практической конференции «Информационные технологии и автоматизация управления», Омск, 9-12 апреля 2011 г.

Структура и объем работы. Диссертация состоит из введения, семи глав, заключения, семи приложений и списка литературы, включающего 286 наименований. Основная часть работы изложена на 300 страницах машинописного текста. Работа содержит 54 рисунка и 27 таблиц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность темы исследования, определяются направление и цель диссертационной работы, отмечаются новизна, теоретическая и практическая ценность работы, кратко излагается содержание работы.

В первой главе выполняется обзор методов аналитического и имитационного моделирования, используемых в теории и инженерной практике проектирования современных ИВС для решения задач системного уровня. Вводятся необходимые понятия и обсуждаются достоинства АИМ. Уточняется цель, и определяются задачи диссертационного исследования, направленного на разработку эффективных аналитико-имитационных методов решения актуальных задач исследования, анализа и синтеза систем с очередями и стохастических сетей.

Во второй главе исследуются системы с относительно интенсивными прерываниями, т.е. СМО $GI_2|GI_2|n$ и $GI_2|GI_2|n|$ ($n \geq 1$) с абсолютными приоритетами и дообслуживанием и с относительной интенсивностью $\alpha = \lambda/\lambda' \gg 1$, где λ и λ' – интенсивности приоритетного и неприоритетного потоков заявок соответственно. Анализ таких систем является задачей, трудной как для аналитического решения, так и для решения методом ИМ. Если при ИМ для достижения заданной точности оценок, характеризующих неприоритетную очередь, моделируется прохождение через систему N неприоритетных заявок, то вместе с этим моделируется и прохождение в среднем αN приоритетных заявок. И т.к. длительность прогона модели примерно пропорциональна числу имитируемых заявок, то по сравнению со случаем $\alpha = 1$ равных интенсивностей время моделирования с ростом α возрастает в $(N + \alpha N)/(N + N) = (1 + \alpha)/2 \approx \alpha/2$ раз. С ростом α затраты времени на ИМ, и без того значительные, могут возрасти на порядки. Эта ситуация типична, например, для ИМ центральных процессоров ИВС, где управляющие процессы обладают абсолютным приоритетом по отношению к прикладным процессам.

Для решения проблемы ИМ систем с интенсивными прерываниями обосновывается возможность *приближенной* при больших α *декомпозиции* системы

$GI_2|GI_2|n$ (или $GI_2|GI_2|n||$) на вложенную приоритетную систему S , получаемую из исходной системы устранением потока неприоритетных заявок, и фоновую систему S^* , получаемую устранением приоритетного потока и корректировкой времени обслуживания неприоритетных заявок.

Разрабатываются методы асимптотического анализа систем $GI_2|GI_2|n$ и $GI_2|GI_2|n||$ при $\alpha \rightarrow \infty$. Выводятся асимптотические соотношения, характеризующие цикл обслуживания неприоритетной заявки. Условие $\alpha \rightarrow \infty$ специфицируется путем: а) фиксации функций распределения вероятностей (ф.р.) $A_1(t)$ и $B_1(t)$ интервалов поступления и, соответственно, времени обслуживания *неприоритетных* заявок, и б) масштабного с коэффициентом α преобразования ф.р. $A(t)$ и $B(t)$ интервалов поступления и, соответственно, времени обслуживания *приоритетных* заявок: $A(t) = A_0(\alpha t)$, $B(t) = B_0(\alpha t)$, где A_0, B_0 – ф.р., заданные при $\alpha = 1$. Первые два момента перечисленных распределений полагаются конечными.

В одноканальной СМО $GI_2|GI_2|1$ выполнен анализ цикла обслуживания неприоритетной заявки при фиксированном чистом времени $x' = T$ ее обслуживания. Выявлены и изучены две системы процессов восстановления и накопления, вложенные в цикл обслуживания. Разработаны два метода асимптотического по α анализа цикла обслуживания. Оба метода – основанный на суммировании периодов занятости *метод СПЗ* и основанный на суммировании периодов обслуживания *метод СПО* – используют известные предельные теоремы теории восстановлений, установленные при статистическом анализе последовательности событий. С целью применения соответствующих выражений асимптотически нормальной при $T \rightarrow \infty$ совместной двумерной ф.р. для числа восстановлений на отрезке $[0, T]$ и суммы случайных приращений, происходящих в моменты восстановлений, условие $T \rightarrow \infty$ трансформируется в условие $T = \text{const}$, $\alpha \rightarrow \infty$. При этом сумма приращений трансформируется в суммарное время прерываний, а число восстановлений – либо в число прерывающих периодов занятости (метод СПЗ), либо в число прерывающих заявок (метод СПО) (*Задорожный В.Н., АиТ, 2008*).

Методы СПЗ и СПО позволяют корректно реализовать идею декомпозиции системы $GI_2|GI_2|1$ на вложенную и фоновую системы, и тем самым решить для системы $GI_2|GI_2|1$ проблему разномасштабных интенсивностей. Метод СПО позволяет также найти асимптотические по α выражения условных по T характеристик цикла обслуживания непосредственно через известные параметры интервалов поступления и времени обслуживания заявок приоритетного класса.

Методом СПЗ установлено, что в системе $GI_2|GI_2|1$ при фиксированном чистом времени $x' = T$ обслуживания неприоритетной заявки совместное распределение суммарного времени Z_T прерываний ее обслуживания и числа γ_T прерывающих периодов занятости (числа прерываний) является асимптотически нормальным распределением с параметрами:

$$\bar{Z}_T \sim \frac{\rho}{1-\rho} \cdot T, \quad \bar{\gamma}_T \sim \frac{T}{\bar{\psi}}, \quad \text{corr}(Z_T, \gamma_T) \sim \frac{C_\psi - rC_\pi}{\sqrt{C_\pi^2 - 2rC_\pi C_\psi + C_\psi^2}},$$

$$C_{Z_T}^2 \sim \frac{\bar{\psi}}{T} (C_\pi^2 - 2rC_\pi C_\psi + C_\psi^2), \quad C_{\gamma_T}^2 \sim \frac{\bar{\psi}}{T} \cdot C_\psi^2, \quad (\alpha \rightarrow \infty), \quad (1)$$

где \bar{Z}_T – среднее суммарное время обслуживания приоритетных заявок (суммарное время прерываний) в цикле обслуживания неприоритетной заявки,
 $\rho = \lambda \bar{x}$ – приоритетный коэффициент загрузки (\bar{x} – среднее время обслуживания приоритетной заявки),

$\bar{\gamma}_T$ – среднее число прерываний обслуживания неприоритетной заявки,

$\bar{\psi}$ – средняя длительность периода незанятости системы приоритетными заявками (приоритетного периода незанятости),

C_{Z_T} – коэффициент вариации (к.в.) времени прерываний Z_T ,

C_π, C_ψ – к.в. периода занятости π и, соответственно, периода незанятости ψ системы приоритетными заявками,

C_{γ_T} – к.в. числа прерываний обслуживания неприоритетной заявки,

$corr(Z_T, \gamma_T)$ – коэффициент корреляции между Z_T и γ_T ,

r – коэффициент корреляции периода занятости системы приоритетными заявками и последующего приоритетного периода незанятости.

Характеристики $\bar{\psi}$, C_π , C_ψ и r в правых частях соотношений (1) определяются по отношению к вложенной системе S , получаемой путем устранения потока неприоритетных заявок, и в общем случае рассчитываются с помощью ИМ.

Методом СПО установлено, что в системе $GI_2|GI_2|1$ при фиксированном числе времени $x' = T$ обслуживания неприоритетной заявки совместное распределение времени Z_T прерываний и числа ν_T приоритетных заявок в цикле обслуживания является асимптотически нормальным распределением с параметрами:

$$\bar{Z}_T \sim \frac{\rho}{1-\rho} \cdot T, \quad \bar{\nu}_T \sim \frac{T}{\bar{\tau}} \cdot \frac{1}{(1-\rho)},$$

$$C_{Z_T}^2 \sim \frac{\bar{\tau}}{T} \cdot \left(\frac{C_\tau^2 + C_x^2}{1-\rho} \right), \quad C_{\nu_T}^2 \sim \frac{\bar{\tau}}{T} \cdot \left(\frac{C_\tau^2 + \rho^2 C_x^2}{1-\rho} \right), \quad (2)$$

$$corr(Z_T, \nu_T) \sim \frac{1 + \rho Q^2}{\sqrt{1+Q^2} \cdot \sqrt{1+\rho^2 Q^2}}, \quad (\alpha \rightarrow \infty), \quad (3)$$

где $\bar{\nu}_T$ – среднее число прерывающих заявок, обслуживаемых в цикле,

$\bar{\tau}$ – средняя длительность интервала поступления приоритетных заявок,

C_{ν_T} – к.в. числа прерывающих заявок,

C_τ, C_x – к.в. интервалов поступления и, соответственно, времени обслуживания приоритетных заявок,

$corr(Z_T, \nu_T)$ – коэффициент корреляции между Z_T и ν_T ,

$Q = C_x/C_\tau$ – относительная вариация времени обслуживания приоритетной заявки.

Найдены асимптотические характеристики цикла обслуживания неприоритетной заявки в **многоканальной системе $GI_2|GI_2|n$** с разделяемой каналами неприоритетной работой. В этой системе обслуживание неприоритетной заявки идеально распараллеливается между свободными каналами, и новая неприоритетная заявка не начинает обслуживаться, пока не завершится обслуживание предыдущей неприоритетной заявки. Когда какой-либо канал освобождается от приоритетных заявок, на него тут же переносится часть незавершенного обслуживания уже вы-

полняемой неприоритетной заявки (если такая имеется). В этой системе при фиксированном чистом времени $x' = T$ обслуживания неприоритетной заявки время Z_T прерываний и соответствующее число ν_T прерывающих заявок имеют асимптотически нормальное двумерное распределение вероятностей с параметрами

$$\begin{aligned}\bar{Z}_T &\sim \frac{\rho}{1-\rho} \cdot T, & \bar{\nu}_T &\sim \frac{T}{n\bar{\tau}} \cdot \frac{1}{(1-\rho)}, \\ C_{Z_T}^2 &\sim \frac{n\bar{\tau}}{T} \cdot \left(\frac{C_\tau^2 + C_x^2}{1-\rho} \right), & C_{\nu_T}^2 &\sim \frac{n\bar{\tau}}{T} \cdot \left(\frac{C_\tau^2 + \rho^2 C_x^2}{1-\rho} \right),\end{aligned}\quad (\alpha \rightarrow \infty),$$

где $\rho = \bar{x}/(n\bar{\tau})$ – коэффициент загрузки системы приоритетными заявками. Коэффициент корреляции $\text{corr}(Z_T, \nu_T)$ описывается выражением (3).

Для систем $GI_2|GI_2|1$ и $GI_2|GI_2|n$ **найлены асимптотические выражения первых двух безусловных моментов** календарного времени у обслуживания неприоритетной заявки и числа ν заявок, прерывающих ее обслуживание:

$$\bar{y} \sim \frac{\bar{x}'}{n(1-\rho)}, \quad \bar{\nu} \sim \frac{\bar{x}'}{n\bar{\tau}(1-\rho)}, \quad (4)$$

$$C_y^2 \sim C_{x'}^2 + \frac{n\rho}{(1-\rho)} \cdot (C_\tau^2 + C_x^2) \cdot \frac{\bar{x}}{\bar{x}'}, \quad C_\nu^2 \sim C_{x'}^2 + \frac{n\bar{\tau}}{\bar{x}'} \cdot \left(\frac{C_\tau^2 + \rho^2 C_x^2}{1-\rho} \right), \quad (5)$$

где \bar{y} – среднее календарное время обслуживания неприоритетной заявки,

$\bar{\nu}$ – среднее число прерывающих заявок в цикле обслуживания,

\bar{x}' – среднее значение чистого времени x' (трудоемкости) обслуживания неприоритетной заявки,

$C_{x'}$ – к.в. времени x' .

Соответствующая плотность распределения вероятностей (п.в.) календарного времени y выражается интегральным асимптотическим приближением

$$f_y(t) \sim \int_0^\infty \frac{b_1(T)}{\sqrt{2\pi} \cdot \sigma(T)} \cdot \exp\left\{-\frac{[t-a(T)]^2}{2\sigma^2(T)}\right\} dT, \quad (6)$$

где $b_1(t) = \frac{dB_1(t)}{dt}$ – п.в. времени x' , $\sigma(T) = \left(\frac{(C_\tau^2 + C_x^2)\rho^2\bar{\tau}}{n(1-\rho)^3} \cdot T \right)^{1/2}$, $a(T) = \frac{T}{n(1-\rho)}$,

и масштабным асимптотическим приближением

$$f_y(t) \sim n(1-\rho) \cdot b_1(n(1-\rho) \cdot t). \quad (7)$$

С использованием ИМ **выполнены испытания приближений** (6), (7) и сформулированы рекомендации по их практическому применению. В общем случае масштабное приближение (7) рекомендуется применять для оценки формы п.в. $f_y(t)$ в области значений t , сравнимых с \bar{y} (рис. 1). Интегральное приближение (6) позволяет с хорошей точностью оценивать хвосты распределения $f_y(t)$.

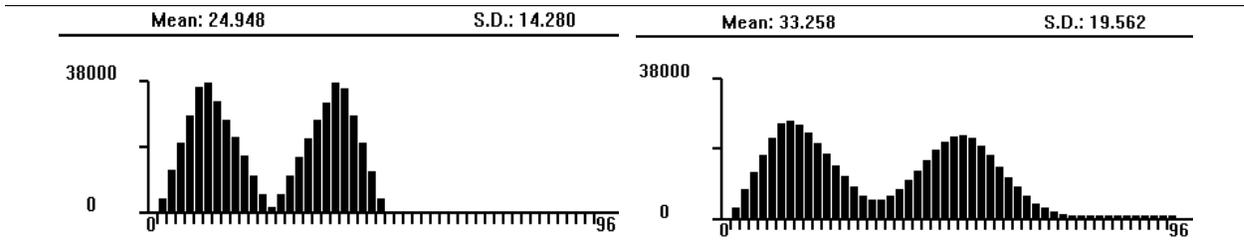


Рис. 1. Трансформация прерываниями двухмодальной п.в. чистого времени x' обслуживания неприоритетной заявки (слева) в п.в. длительности у цикла обслуживания

Гистограммы на рис. 1 получены путем ИМ одной из тестовых СМО (одноканальной) при среднем числе прерываний $\bar{\nu} = 8,33$. В этой СМО для сл.в. x' на отрезке $(0, 50)$ задана двухмодальная п.в., симметричный график которой составлен боковыми сторонами двух одинаковых треугольников. Основания треугольников лежат на отрезках $(0, 25)$ и $(25, 50)$, вершины – в точках с координатами $(10; 0,04)$ и $(40; 0,04)$. Здесь $\bar{x}' = 25$, $C_{x'}^2 = 0,3267$. Сл.в. x , τ и τ' экспоненциальные. Среднее число прерываний $\bar{\nu}$, определяемое формулой (4), при $\rho = 1/4 = \text{const}$ изменялось путем одновременного изменения \bar{x} и $\bar{\tau}$. При $\bar{\nu} = \bar{\nu}_1 = 8,33$ и $\bar{\nu} = \bar{\nu}_2 = 83,3$ путем ИМ вычислены соответствующие вероятности $p_1 = \mathbf{0,022}$ и $p_2 = \mathbf{0,0021}$ хвоста $y > 66,67$. Из (6) для вероятности хвоста $y > T_0$ выводится интегральная оценка

$$p = P\{y > T_0\} \sim 1 - \int_0^{\infty} b_1(T) \cdot \Phi\left(\frac{T_0 - a(T)}{\sigma(T)}\right) dT,$$

(где Φ – стандартная нормальная ф.р.), дающая для p_1 и p_2 приближения $p_1 \approx \mathbf{0,019}$ и $p_2 \approx \mathbf{0,0021}$. С ростом $\bar{\nu}$ точность интегральной оценки возрастает.

Сопоставлением асимптотических представлений (1) и (2) показателя $C_{z\tau}^2$ **установлено точное соотношение** для систем GI|GI|1, названное C-соотношением:

$$\bar{\psi} \cdot (C_{\pi}^2 - 2rC_{\pi}C_{\psi} + C_{\psi}^2) = \bar{\tau} \cdot \left(\frac{C_{\tau}^2 + C_x^2}{1 - \rho}\right).$$

Найденное C-соотношение позволяет контролировать правильность результатов, получаемых при исследовании периодов занятости СМО, или упрощать анализ периодов занятости и незанятости при имитационном моделировании за счет вычисления оценки для r через оценки других моментов. В доказательстве C-соотношения используется тот факт, что любую систему GI|GI|1 можно рассматривать как приоритетную систему, вложенную в систему GI₂|GI₂|1.

В приложении А диссертации C-соотношение тестируется и подтверждается аналитическими проверками, численными методами и имитационным моделированием. **В приложении Б** в связи с повышенными требованиями к проводимым тестовым имитационным экспериментам выполняется углубленное тестирование датчиков случайных чисел и определяются пределы их корректного использования, которые учитываются во всех последующих имитационных экспериментах.

В третьей главе метод приближенной декомпозиции систем с абсолютными приоритетами на вложенную систему S и фоновую систему S* используется вместе с установленными в главе 2 асимптотическими соотношениями для решения задач расчета и моделирования систем GI₂|GI₂|n и GI₂|GI₂|n|. Приближенные **фор-**

мулы для расчета систем $GI_2|GI_2|n$ и $GI_2|GI_2|n||$ выводятся на основе приближенных или точных формул, известных для систем $GI|GI|n$, $n \geq 1$. Вложенная система S рассчитывается непосредственно как система $GI|GI|n$ с заданными известными параметрами. Аналогично (как система $GI|GI|n$) рассчитывается фоновая система S^* , время x' обслуживания заявок в которой предварительно заменяется длительностью x^* цикла обслуживания неприоритетных заявок в исходной системе.

Этот подход позволил найти формулы для расчета системы $GI_2|GI_2|n||$ при большой нагрузке. При ИМ СМО близкий к единице коэффициент загрузки приводит к большим погрешностям оценок средних характеристик очередей (теоретически это рассмотрено в приложении Б). Погрешности ИМ еще более возрастают при сочетании большой нагрузки и разномасштабных интенсивностей потоков. В таких условиях вместо ИМ может применяться формула, выведенная для расчета средней длины \bar{q}' очереди неприоритетных заявок в СМО $GI_2|GI_2|n||$:

$$\bar{q}' \sim \frac{C_{\tau'}^2 + \rho^{*2} C_{x^*}^2}{2(1 - \rho^*)}, \quad (\alpha \rightarrow \infty), \quad (8)$$

где $\rho^* = \lambda \bar{x}^* \rightarrow 1$, $\bar{x}^* = \frac{\bar{x}'}{1 - \rho}$, $C_{x^*}^2 = C_{x'}^2 + \frac{n \bar{\tau} \rho^2}{\bar{x}'(1 - \rho)}(C_{\tau'}^2 + C_x^2)$, $C_{\tau'}$ – к.в. длительности τ' интервала поступления неприоритетных заявок. Условие $\alpha \rightarrow \infty$ эквивалентно условию $\bar{\tau} / \bar{x}' \rightarrow 0$. Проверка формулы (8) путем ИМ СМО $GI_2|GI_2|1$ с равномерным распределением случайных величин (сл.в.) τ' , x' , τ и усеченным гиперболическим распределением сл.в. x (при $C_x^2 = 27,24$) показала при $\alpha = 10$ следующее (табл.1). Точность оценки (8) (графа «расчет») с увеличением суммарной загрузки ρ^* возрастает (варьировались \bar{x} и \bar{x}' при $\bar{x} / \bar{x}' = const$).

Таблица 1

Сравнение результатов расчета и моделирования тестовой СМО

ρ^*	\bar{x}^*		C_{x^*}		\bar{q}'_0	\bar{q}'	
	расчет	ИМ	расчет	ИМ		расчет	ИМ
0,50	5,00	5,00	1,31	1,31	0,15	0,76	0,82
0,70	7,00	6,99	1,50	1,50	0,50	2,40	2,70
0,80	8,00	7,99	1,59	1,59	0,99	4,90	5,17
0,90	9,00	8,99	1,68	1,68	2,63	13,07	13,71
0,95	9,50	9,50	1,72	1,72	5,79	29,98	30,52

Величина \bar{q}'_0 в табл. 1 представляет собой альтернативную оценку средней длины очереди, получаемую при использовании известного метода усреднения вклада интенсивных прерываний. Из табл. 1 видно, что погрешность расчета параметров \bar{x}^* и C_{x^*} цикла обслуживания гораздо ниже погрешности расчета величины \bar{q}' и лежит в пределах долей процента. Это позволяет заключить, что основным источником погрешностей в данном случае является используемая формула Кингмана-Келлерстрема, но не метод декомпозиции. Отсюда следует, что применение в методе декомпозиции для приближенного расчета фоновой систе-

мы S^* точных результатов ТМО должно приводить к более точным решениям.

Для расчета \bar{q}' в системе $GI_2|GI_2|1$ с пуассоновским неприоритетным потоком выводится (с применением для расчета системы S^* формулы Поллачека-Хинчина) следующая приближенная формула:

$$\bar{q}' \sim \frac{\rho^*}{2(1-\rho_\Sigma)} \left[\rho'(1+C_x^2) + \frac{1}{\alpha} \frac{\rho^2}{\rho'} (C_\tau^2 + C_x^2) \right], \quad (9)$$

где $\rho_\Sigma = \rho + \rho'$ – суммарный коэффициент загрузки, $\rho' = \lambda \bar{x}'$. Тестирование точности формулы (9) посредством ИМ системы с равномерными и эрланговскими распределениями сл.в. x, x', τ показало следующее (табл. 2). При $\alpha = 10$ в широком диапазоне изменения загрузки погрешность формулы (9) лежит в пределах 2,5%. И даже при $\alpha = 1$ погрешность формулы (при $\rho_\Sigma \geq 0,5$) не превышает 5-6%.

Таблица 2

Результаты приближенного расчета тестовой СМО и ее моделирования при $\alpha = 10$

ρ^*		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
\bar{q}'	расчет	0,009	0,039	0,101	0,209	0,392	0,705	1,28	2,51	6,35
	ИМ	0,009	0,040	0,102	0,209	0,394	0,708	1,32	2,51	6,39

На основе асимптотических результатов главы 2 методом декомпозиции выводятся приближенные формулы для первых двух моментов времени обслуживания пакетов неприоритетных заявок в фоновом режиме. Посредством ИМ показывается хорошая точность полученных формул, возрастающая с ростом среднего числа β прерываний обслуживания пакета. Формулы можно использовать для практических расчетов уже при $\beta = 1-10$ и, тем более, при $\beta > 10$. Развитие этого результата приводит к разработке и исследованию (в гл. 4) моделей ЦСО.

На основе метода приближенной декомпозиции разрабатывается и тестируется ускоренный метод ИМ систем с интенсивными прерываниями. Здесь использование декомпозиции не требует применения расчетных приближений для очередей и поэтому характеризуется малыми погрешностями, сопоставимыми с погрешностями оценки моментов для длительности цикла обслуживания неприоритетной заявки. Так, при относительной интенсивности потока приоритетных заявок $\alpha > 5$ погрешность ускоренного метода составляет доли процента (в смысле систематической погрешности, характеризующей оценку средней длины очереди неприоритетных заявок). При большой нагрузке для погрешности известного метода, использующего усреднение вклада прерываний, выводится формула

$$\frac{\bar{q}'_0 - \bar{q}'}{\bar{q}'} \approx \frac{1}{\alpha} \cdot \frac{\rho^2}{(1-\rho)^2} \cdot Q_{np},$$

где \bar{q}'_0 – математическое ожидание получаемой этим методом оценки средней длины \bar{q}' неприоритетной очереди, $Q_{np} = (C_\tau^2 + C_x^2)/(C_\tau^2 + C_x^2)$ – показатель, характеризующий относительную вариабельность приоритетной нагрузки.

Оценены перспективы обобщения разработанного метода ускоренного моделирования на системы $GI_k|GI_k|n$ при числе классов заявок $k > 2$, в том числе на

случай многоуровневых прерываний. Тестирование точности обобщенных версий ускоренного метода ИМ показало ее сопоставимость с точностью ускоренного метода ИМ систем GI₂|GI₂|1.

В четвертой главе разрабатываются методы АИМ многоканальных ЦСО – циклических систем обслуживания, характеризуемых наличием циклов накопления пакетов неприоритетных заявок и обслуживания этих пакетов в фоновом режиме. Отличие ЦСО от рассмотренных в главе 2 систем GI₂|GI₂|n|| с параллельным обслуживанием неприоритетных заявок заключается в том, что неприоритетные заявки, поступления которых в ЦСО образуют рекуррентный поток событий, предварительно накапливаются в пакеты. Длительность циклов накопления пакетов определяется одним из трех способов (Т-, N- и X-пакетирование). Накопленный пакет обслуживается одновременно всеми каналами в промежутках времени, свободных от обслуживания приоритетных заявок. Период обслуживания пакета – период повышенной нагрузки (ППН) – должен заканчиваться до момента поступления следующего пакета. Если к этому моменту времени обслуживание пакета не заканчивается, то он остается обслуженным не полностью. Возможный вариант – ускорение работ по обслуживанию пакета. Оба варианта или их комбинация отрицательно сказываются на показателях эффективности ЦСО.

Основная задача анализа ЦСО заключается в определении первичных технико-экономических показателей эффективности ЦСО. К ним относятся моментные и квантильные характеристики календарной длительности h обслуживания пакета, и моменты числа V приоритетных заявок, прерывающих обслуживание пакета.

Рассмотрены **типичные способы пакетирования** неприоритетных заявок:

- формирование пакета по заданному времени T_0 накопления неприоритетных заявок (Т-пакетирование);
- формирование пакета по заданному числу N заявок (N-пакетирование);
- формирование пакета по заданной его трудоемкости X_m (X-пакетирование).

При любом из этих способов пакетирования ЦСО можно рассматривать как систему GI₂|GI₂|n||, в которую в качестве потока неприоритетных заявок поступает поток сформированных пакетов. Для этого нужно определить соответствующие характеристики интервала θ поступления пакетов, объема N пакета и его трудоемкости ξ (т.е. чистого суммарного времени обслуживания заявок в пакете).

Найдены **асимптотические характеристики потоков пакетов**.

При Т-пакетировании интервал θ поступления пакетов фиксирован: $\theta = T_0$, а число N заявок в пакете и его трудоемкость ξ случайны. Если $T_0 \rightarrow \infty$, то:

$$\bar{N} \sim \frac{T_0}{\bar{\tau}'}, \quad D(N) \sim \frac{D(\tau')}{\bar{\tau}'^3} \cdot T_0, \quad C_N^2 \sim \frac{D(\tau')}{\bar{\tau}' \cdot T_0};$$

$$\bar{\xi} \sim \frac{T_0}{\bar{\tau}'} \cdot \bar{x}', \quad D(\xi) \sim T_0 \cdot \left(\frac{\sigma_{x'}^2}{\bar{\tau}'} + \frac{\sigma_{\tau'}^2 \cdot \bar{x}'^2}{\bar{\tau}'^3} \right), \quad C_\xi^2 \sim \frac{\bar{\tau}'}{T_0} \cdot (C_{x'}^2 + C_{\tau'}^2),$$

$$\text{corr}(\xi, N) \sim \frac{C_{\tau'}}{\sqrt{C_{\tau'}^2 + C_{x'}^2}} = \frac{1}{\sqrt{1+Q^2}},$$

(где $Q = C_{x'}/C_{\tau'}$ – относительная вариация времени x' обслуживания заявки), а совместное распределение вероятностей сл.в. N и ξ асимптотически нормальное.

При X-пакетировании формирование пакета заканчивается в момент перескока его трудоемкостью ξ заданного уровня X_m . Если $X_m \rightarrow \infty$, то $\bar{\xi} \sim X_m$, $C_\xi^2 \rightarrow 0$. С учетом остаточного превышения трудоемкостью ξ заданного верхнего уровня

$$\bar{\xi} \sim X_m + \frac{(x')^2}{2\bar{x}'} = X_m + \bar{x}'(1 + C_{x'}^2)/2.$$

Число N заявок в пакете и интервал θ поступления пакетов распределены асимптотически нормально, причем

$$\begin{aligned} \bar{N} &\sim \frac{X_m}{\bar{x}'}, & D(N) &\sim \frac{D(x')}{\bar{x}'^3} \cdot X_m, & C_N^2 &\sim \frac{D(x')}{\bar{x}' \cdot X_m}; \\ \bar{\theta} &\sim \frac{X_m}{\bar{x}'} \cdot \bar{\tau}', & D(\theta) &\sim X_m \cdot \left(\frac{\sigma_{x'}^2}{\bar{x}'} + \frac{\sigma_{x'}^2 \cdot \bar{\tau}'^2}{\bar{x}'^3} \right), & C_\theta^2 &\sim \frac{\bar{x}'}{X_m} \cdot (C_\tau^2 + C_{x'}^2); \\ \text{corr}(\theta, N) &\sim \frac{C_{x'}}{\sqrt{C_{x'}^2 + C_\tau^2}} = \frac{1}{\sqrt{1 + Q^{-2}}}. \end{aligned}$$

При N-пакетировании

$$\bar{\xi} = N \cdot \bar{x}', \quad \bar{\theta} = N \cdot \bar{\tau}', \quad D(\xi) = N \cdot D(x'), \quad D(\theta) = N \cdot D(\tau'),$$

распределения сл.в. ξ и θ асимптотически нормальные, и сл.в. ξ и θ независимы.

На основе характеристик потоков пакетов и результатов, найденных в гл. 3 для систем $GI_2|GI_2|n||$ получены **асимптотические выражения первичных показателей эффективности ЦСО**. В частности, для длительности h ППН найдено, что

$$\bar{h} \sim \frac{\bar{N}}{n} \cdot \frac{\bar{x}'}{(1-\rho)}, \quad C_h^2 \sim \frac{1}{\bar{N}} \cdot \left[C^2 + \frac{n\rho}{(1-\rho)} \cdot (C_\tau^2 + C_x^2) \cdot \frac{\bar{x}}{\bar{x}'} \right], \quad (10)$$

где $\rho = \lambda \bar{x}/n$, а среднее число заявок в пакете \bar{N} и фактор C^2 определяются для различных способов пакетирования следующим образом:

$$\begin{aligned} - \text{для T-пакетирования} & \quad \bar{N} = T_0 / \bar{\tau}', & C^2 &= C_\tau^2 + C_x^2; \\ - \text{для X-пакетирования} & \quad \bar{N} = X_m / \bar{x}' + (1 + C_{x'}^2)/2, & C^2 &= 0; \\ - \text{для N-пакетирования} & \quad \bar{N} = N, & C^2 &= C_{x'}^2. \end{aligned}$$

Для вероятности p превышения величиной h заданного порога T получены выражения, учитывающие характеристики (10) и использующие интегральное приближение (6), в котором параметры неприоритетных заявок заменяются соответствующими параметрами пакетов. Аналогично с помощью выражений для ν (4), (5) получены выражения для числа заявок, прерывающих обслуживание пакета.

С помощью ИМ выполнено **исследование точности асимптотических результатов** (10). Основным показателем, определяющим возможность практического использования (10) в качестве приближенных формул, является показатель $\beta = (\bar{N} \bar{x}')/(\bar{\tau} n)$, характеризующий среднее число приоритетных заявок, прерывающих обслуживание пакета. При фиксированном β основными факторами, влияющими на точность приближений (10), являются к.в. $C_i^2 \in \{C_\tau^2, C_\tau'^2, C_x^2, C_{x'}^2\}$ сл.в. τ , τ' , x и x' . Точность приближений (10) практически не зависит от других параметров ЦСО, в том числе от n . Установлено, что если все C_i^2 лежат в пределах

от 0,02 до 30, то формулы (10) вполне можно использовать для практических расчетов. Даже при небольших значениях показателя β , т.е. при $0,2 < \beta < 10$, погрешности формул (10) лежат в пределах 1-3%. С ростом β погрешности снижаются.

Исследованы **возможности регулирования аритмичности нагрузки**. В качестве показателя аритмичности используется к.в. C_h . Показано, что при объединении n одноканальных ЦСО в одну многоканальную аритмичность может быть снижена в \sqrt{n} раз. В ситуациях, где это целесообразно, например, при стрессовых испытаниях ИВС, тренировке спортсменов, операторов технических установок и т.д. аритмичность может повышаться неограниченно (без изменения коэффициентов загрузки) за счет увеличения к.в. C_τ и C_x , характеризующих приоритетные заявки. Требуемые C_τ и C_x можно обеспечить применением датчиков сл.в.

Показана **перспективность использования ЦСО** как одной из типовых моделей для проектирования ИВС, анализа и оптимизации почтовых служб (включая электронную почту), вычислительных серверов, городских транспортных сетей, учебных заведений, систем мониторинга и т.д.

В пятой главе разрабатываются двухуровневые аналитико-имитационные методы оптимизации однородных немарковских СеМО, использующие на каждой итерации одно обращение к имитационной модели сети для оценки значения целевой функции в текущей точке факторного пространства. Проблема вычисления градиента (или квазиградиента) по стохастическому отклику решается за счет применения сепарабельной локальной (в окрестностях текущей точки) нелинейной аппроксимации поверхности отклика. Аппроксимация используется для оценки оптимального направления и оптимальной длины шага и в следующей точке перенастраивается по результатам следующего обращения к модели.

Задача оптимального распределения непрерывного ресурса (стоимости, производительности) по узлам однородной немарковской СеМО формулируется следующим образом. Для открытой сети задана ф.р. $A(t)$ интервалов поступления заявок в сеть, определяющая интенсивность Λ рекуррентного входного потока. Для замкнутой сети фиксировано число R заявок в сети (ф.р. $A(t)$ не задается). Известны ф.р. $B_i(t)$ времени обслуживания заявок в узлах i сети, $i = \overline{1, n}$, и числа K_i каналов в узлах. Вероятности p_{ij} ($i, j = \overline{0, n}$) переходов между узлами (индекс $i = 0$ соответствует внешней среде) заданы неразложимой стохастической матрицей $\mathbf{P} = \|p_{ij}\|$. Ресурс M сети, определяемый вектором $\vec{\mu} = (\mu_1, \dots, \mu_n)$ интенсивностей обслуживания в узлах, фиксирован: $M(\vec{\mu}) = \sum_{i=1}^n c_i \mu_i^{\beta_i} = M^* = const$, где c_i – стоимостные коэффициенты, $\beta_i > 0$ – коэффициенты нелинейности.

Среднее время $E = E(\vec{\mu})$ прохождения заявки через сеть (среднее время ответа, среднее время цикла) можно представить в виде:

$$E = \sum_{i=1}^n \alpha_i u_i = \sum_{i=1}^n \alpha_i (w_i + b_i) = \sum_{i=1}^n \alpha_i \left(w_i + \frac{1}{\mu_i} \right), \quad (11)$$

где α_i – среднее число посещений i -го узла заявкой за время цикла,

μ_i – среднее время пребывания заявки в i -м узле,

w_i – среднее время ожидания заявки в очереди i -го узла,

b_i – среднее время обслуживания заявки в i -м узле,

$\mu_i = 1/b_i$ – интенсивность обслуживания заявки каналом i -го узла.

Коэффициенты α_i однозначно определяются из системы уравнений баланса:

$$\alpha_i = \sum_{j=0}^n \alpha_j p_{ji} \quad i = \overline{0, n}, \quad \alpha_0 \equiv 1.$$

Для открытой сети через α_i определяются интенсивности $\lambda_i = \Lambda \alpha_i$ входных потоков узлов, коэффициенты загрузки узлов $\rho_i = \lambda_i / (\mu_i / K_i)$, и проверяются условия стационарности $\rho_i \leq 1$ (или $\mu_i \geq \lambda_i / K_i$), $i = \overline{1, n}$. Значения w_i в (11) определяются посредством ИМ.

В задаче требуется найти вектор $\vec{\mu} = \vec{\mu}_{opt}$, доставляющий минимум среднего времени ответа $E = E(\vec{\mu})$:

$$E = E(\vec{\mu}) \rightarrow \min_{\vec{\mu}}, \quad (12)$$

и принадлежащий области допустимых решений (ОДР):

$$M(\vec{\mu}) = \sum_{i=1}^n c_i \mu_i^{\beta_i} = M^*, \quad \mu_i \geq \mu_{i \min}, \quad (i = \overline{1, n}), \quad (13)$$

где для открытой сети $\mu_{i \min} = \lambda_i / K_i$ (граница области стационарности), а для замкнутой сети $\mu_{i \min} = 0$. Для ресурса M^* выполняется условие $M^* \geq M_{\min}$, где для открытой сети $M_{\min} = \sum_{i=1}^n c_i \mu_{i \min}^{\beta_i} = \sum_{i=1}^n c_i (\lambda_i / K_i)^{\beta_i}$, а для замкнутой $M_{\min} = 0$.

В задаче (12), (13) имеется в виду, что изменение любой интенсивности μ_i приводит к изменению среднего $b_i = 1/\mu_i$ и к соответствующему масштабному изменению ф.р. $B_i(t)$. Вид ф.р. $B_i(t)$ не изменяется, поскольку ассоциируется со случайной трудоемкостью заявок на соответствующих этапах обслуживания сетью. Варьируемый параметр μ_i взаимно однозначно связан с ресурсом $c_i \mu_i^{\beta_i}$ i -го узла.

Для решения задачи (12), (13) разработан **двухуровневый двухэтапный аналитико-имитационный метод**, включающий следующие два этапа. Этап I – ускоренный градиентный поиск точки $\vec{\mu}_{opt}$ оригинальным **методом «направляющих гипербол»**, использующим ИМ сети и локальную аппроксимацию целевой функции (название метода НГ отражает роль и форму компонент аппроксимации). Этап II (не обязательный) – уточнение найденного решения ускоренным методом циклического покоординатного спуска, не использующим аппроксимаций.

Аппроксимация $E^{ap}(\vec{\mu})$ среднего времени ответа $E(\vec{\mu})$ на каждой итерации $k \geq 2$ поиска оптимального решения $\vec{\mu}_{opt}$ методом НГ формируется по результатам ИМ сети в точках $\vec{\mu} = \vec{\mu}^{k-1}$ и $\vec{\mu} = \vec{\mu}^k$, и применяется для определения следующей точки $\vec{\mu} = \vec{\mu}^{k+1}$. **Метод НГ использует следующие опорные элементы.**

1°. **Центр** $\vec{\mu}_c$ ОДР (13) при ресурсе $M^* > M_{\min}$ определяется условием равной загрузки узлов: $\rho_i = \lambda_i / (\mu_i K_i) = \alpha_i \lambda_0 / (\mu_i K_i) = \rho_c$, ($i = \overline{1, n}$), где λ_0 – интенсивность на терминальной дуге (для открытой сети $\lambda_0 = \Lambda$). Отсюда $\mu_i = (\alpha_i \lambda_0) / (K_i \rho_c)$, $\mu_i / \mu_1 = (\alpha_i / K_i) (K_1 / \alpha_1)$, $\mu_i = (\alpha_i / K_i) (K_1 / \alpha_1) \mu_1$. Подставляя последнее выражение в (13), получаем уравнения:

$$\sum_{i=1}^n c_i \left(\frac{\alpha_i}{K_i} \cdot \frac{K_1}{\alpha_1} \cdot \mu_1 \right)^{\beta_i} = M^*, \quad \mu_i = (\alpha_i / K_i) (K_1 / \alpha_1) \mu_1, \quad i = \overline{2, n}. \quad (14)$$

Из первого уравнения численными методами определяется единственный положительный корень μ_i , затем из остальных – все другие координаты μ_i центра $\bar{\mu}_c$. Если сеть открытая, то в центре $\bar{\mu}_c$ сразу вычисляются все $\rho_i = \rho_c < 1, i = \overline{1, n}$.

Если все $\beta_i = 1$, то из (14) координаты центра ОДР можно выразить явно:

$$\mu_i = M^* \cdot (\alpha_i / K_i) \cdot \left(\sum_{j=1}^n c_j \cdot \alpha_j / K_j \right)^{-1}, \quad i = \overline{1, n}.$$

2°. *Диаметр* D ОДР определяется как длина максимального из диапазонов варьирования переменных μ_i : $D = \max\{l_i\}$, где $l_i = \mu_{i\max} - \mu_{i\min}$ и, согласно (13), $\mu_{i\max} = [(M^* - \sum_{j \neq i} c_j \mu_{j\min}^{\beta_j}) \cdot c_i^{-1}]^{1/\beta_i}, i = \overline{1, n}$.

3°. *Малый шаг* размером, например, $D \cdot 10^{-4}$, применяется при построении и сканировании траекторий на поверхности ограничений, определяемой уравнением (13). Шаг выбирается с учетом требований к точности оптимизации.

4°. *Аппроксимация* $E^{ap}(\bar{\mu})$ целевой функции $E(\bar{\mu})$, используемая для приближенного вычисления градиента, представляет собой сепарабельную функцию варьируемых переменных μ_i :

$$E^{ap}(\bar{\mu}) = \sum_{i=1}^n \alpha_i \left(W_i(\mu_i) + \frac{1}{\mu_i} \right), \quad \text{где } W_i(\mu_i) = \begin{cases} \frac{R_i}{\mu_i - S_i}, & \text{если } \hat{w}_i^k \neq \hat{w}_i^{k-1}, \\ \hat{w}_i^k, & \text{если } \hat{w}_i^k = \hat{w}_i^{k-1}, \end{cases}$$

которая на каждом шаге k оптимизации заново настраивается (посредством коэффициентов R_i и S_i) по оценкам \hat{w}_i^{k-1} и \hat{w}_i^k среднего времени ожидания, найденным для узлов $i = \overline{1, n}$ при ИМ сети в точках $\bar{\mu} = \bar{\mu}^{k-1}$ и $\bar{\mu} = \bar{\mu}^k$. При $\hat{w}_i^k \neq \hat{w}_i^{k-1}$ выражение $R_i / (\mu_i - S_i)$, входящее в $E^{ap}(\bar{\mu})$, должно аппроксимировать соответствующую функцию $w_i = w_i(\bar{\mu})$ в (11) так, чтобы его значение в точках $\bar{\mu} = \bar{\mu}^{k-1}$ и $\bar{\mu} = \bar{\mu}^k$ совпадало с имеющимися оценками $\hat{w}_i^{k-1} \approx w_i(\bar{\mu}^{k-1})$ и $\hat{w}_i^k \approx w_i(\bar{\mu}^k)$. Отсюда, требуя выполнения равенств $R_i / (\mu_i^{k-1} - S_i) = \hat{w}_i^{k-1}$, $R_i / (\mu_i^k - S_i) = \hat{w}_i^k$, находим:

$$S_i = \frac{\hat{w}_i^k \mu_i^k - \hat{w}_i^{k-1} \mu_i^{k-1}}{\hat{w}_i^k - \hat{w}_i^{k-1}}, \quad R_i = \hat{w}_i^{k-1} \cdot (\mu_i^{k-1} - S_i), \quad (i = \overline{1, n}),$$

(верхний индекс здесь везде соответствует шагу оптимизации). Такая настройка функции $E^{ap}(\bar{\mu})$ обеспечивает ее совпадение с $E(\bar{\mu})$ в точках $\bar{\mu}^{k-1}$ и $\bar{\mu}^k$ (с точностью до стохастической погрешности оценок ИМ). В других точках $\bar{\mu}$ точность аппроксимации $E^{ap}(\bar{\mu})$ тем хуже, чем они более удалены от $\bar{\mu}^{k-1}$ и $\bar{\mu}^k$, и чем «менее сепарабельна» $E(\bar{\mu})$, т.е. чем сильнее изменение интенсивностей обслуживания μ_i в одних узлах влияет на среднее время ожидания w_j в других ($i \neq j$).

5°. *Градиент* $\nabla E^{ap}(\bar{\mu})$ в точке $\bar{\mu} = \bar{\mu}^k$ есть аппроксимация градиента $\nabla E(\bar{\mu})$ в этой точке, и вычисляется с помощью выражения, полученного дифференцированием $E^{ap}(\bar{\mu})$:

$$\nabla E^{ap}(\bar{\mu}^k) = \left(\alpha_1 \frac{\partial W_1}{\partial \mu_1} - \frac{\alpha_1}{(\mu_1)^2}, \dots, \alpha_n \frac{\partial W_n}{\partial \mu_n} - \frac{\alpha_n}{(\mu_n)^2} \right),$$

где

$$\frac{\partial W_i}{\partial \mu_i} = \begin{cases} \frac{-R_i}{(\mu_i^k - S_i)^2}, & \hat{w}_i^k \neq \hat{w}_i^{k-1}, \\ 0, & \hat{w}_i^k = \hat{w}_i^{k-1}. \end{cases} \quad (i = \overline{1, n}).$$

6°. **Валидная часть** $[L]$ исходящей из $\bar{\mu}^k$ траектории L поиска точки $\bar{\mu}^{k+1}$ лежит между $\bar{\mu}^k$ и первой на L точкой $\bar{\mu}$, у которой какая-либо координата μ_i достигает границы ОДР $\mu_i = \mu_{i \min}$ или полюса $\mu_i = S_i$ аппроксимации $E^{ap}(\bar{\mu})$.

Разработаны пошаговые алгоритмы метода НГ и двухэтапного метода в целом (Zadorozhnyi V.N., *Automation and Remote Control*, 2009). Их программная реализация включена в систему ИМ и внедрена на предприятиях.

Идеи двухуровневой оптимизации, опирающейся на локальную нелинейную сепарабельную аппроксимацию целевой функции, **использованы в задаче оптимального распределения каналов**, которая формулируется следующим образом.

Однородная немарковская СеМО усиливается путем ввода в ее состав $M \geq 1$ дополнительных каналов. Все каналы технически идентичны, любой из них можно добавлять в любой узел. Различия между ф.р. $B_i(t)$ времени обслуживания заявок в узлах $i = \overline{1, N}$ связано с разной трудоемкостью работ, выполняемых разными узлами. Среднее время b_i обслуживания в i -м узле может зависеть от числа n_i его каналов. При увеличении n_i интенсивность $\mu_i = 1/b_i = \mu_i(n_i)$ обслуживания в канале узла может снижаться или повышаться. Функции $\mu_i(n_i)$ монотонные. Целью усиления сети в частном случае является снижение среднего времени ответа E : при заданных функциях $\mu_i(n_i)$ и заданном распределении $\mathbf{n}_0 = (n_{01}, \dots, n_{0N})$ имеющихся каналов по узлам $1, \dots, N$ требуется найти такое распределение $\mathbf{m} = (m_1, \dots, m_N)$ дополнительных M каналов (где $m_i \geq 0$ – число каналов, добавляемых в i -й узел), которое минимизирует E . В общем случае целью является минимизация некоторой выпуклой функции F от линейной комбинации средних длин очередей:

$$F\left(\sum_{i=1}^N c_i \cdot q_i(\mathbf{n}_0 + \mathbf{m})\right) \rightarrow \min_{\mathbf{m}}, \quad (15)$$

$$\|\mathbf{m}\| = M, \quad \mathbf{m} \geq 0,$$

где $q_i(\mathbf{n}_0 + \mathbf{m})$ – средняя длина очереди в i -м узле, c_i – стоимостные коэффициенты (штрафы), $\|\mathbf{m}\| = m_1 + \dots + m_N$.

Для решения этой задачи предложен ускоренный метод последовательного размещения дополнительных каналов. Для случая открытой сети с производительностью каналов в узлах, не зависящей от n_i , разработан аппроксимационный метод решения задачи. Метод основан на следующем эмпирическом приближении зависимости среднего времени ожидания w в изолированной СМО GI|GI|n от числа n ее каналов (Рыжиков Ю.И., 2003):

$$w(n) \approx \frac{w(1)}{n^r}, \quad \text{или, равносильно,} \quad q(n) \approx \frac{q(1)}{n^r}, \quad n \geq 1,$$

где r – числовой параметр, который для фиксированных ф.р. $A(t)$ длительности

интервалов поступления заявок и ф.р. $B(t)$ времени их обслуживания можно подобрать с помощью ИМ. Аппроксимационный метод использует сепарабельную аппроксимацию зависимостей $q_i(\mathbf{n}_0 + \mathbf{m})$ в целевой функции (15) выражениями, распространяющими приближение Ю.И. Рыжикова на узлы СеМО:

$$q_i(\mathbf{n}_0 + \mathbf{m}) \approx q_i(n_{i0}) \left(\frac{n_{i0}}{n_{i0} + m_i} \right)^{r_i}, \quad i = \overline{1, N}, \quad (16)$$

где r_i определяются по результатам ИМ сети в двух точках: при исходном распределении $\mathbf{n}_0 = (n_{01}, \dots, n_{0N})$ и при произвольном распределении $\mathbf{n}_0 + \mathbf{m}' = (n_{01} + m'_{11}, \dots, n_{0N} + m'_{N1})$, в котором все $m'_{ij} \geq 1$. Получаемая после подстановки (16) в (15) задача легко решается элементарными численными методами.

Разработаны пошаговые алгоритмы последовательного и аппроксимационного методов распределения каналов. Испытания показали высокую эффективность этих методов. Их программная реализация включена в систему ИМ. Предусмотрен этап проверки и уточнения получаемых решений методом циклического покоординатного спуска.

В контексте задач оптимизации СеМО разработан численный **расширенный метод редукции**, позволяющий без применения численного дифференцирования рассчитывать *коэффициенты чувствительности* (КЧ) моментов длительности полумарковского процесса (с марковской цепью, вложенной в моменты смены состояний) к вариациям переходных вероятностей. **В приложениях В, Г** приводятся формулы РМР. Его вычислительная сложность оценивается **в приложении Д**. Показано, что он требует выполнения $O(n^2)$ операций при затратах памяти не более, чем $O(n^2)$, где n – размер редуцируемого графа полумарковского процесса (степень связности вершин графа ограничена сверху). Метод НГ агрегирован с РМР для эффективного вычисления КЧ среднего времени ответа СеМО к возмущениям ее переходных вероятностей. Определяются три вида КЧ: *частные* (рассчитываются при условии отсутствия очередей), *виртуальные* (не учитывают чувствительности очередей к изменению переходных вероятностей) и *полные* КЧ. Метод РМР реализован программно, внедрен и используется в организациях.

В шестой главе аналитико-имитационный подход к задачам, трудным для раздельно применяемых аналитического и имитационного методов, распространяется на исследование больших стохастических сетей.

Существенно развит аналитический аппарат теории графа БА – случайного бесконечного графа с линейным правилом предпочтительного связывания, широко используемого для моделирования сетей типа Интернет. Граф БА выращается из небольшого графа-затравки путем неограниченного добавления *фиксированных приращений графа* – новых вершин с одним и тем же числом ребер $m \geq 1$. Свободный конец каждого нового ребра присоединяется к той или иной имеющейся i -й вершине графа, которая выбирается случайно с вероятностью p_i , пропорциональной локальной степени связности k_i этой вершины:

$$p_i = k_i / \sum_j k_j. \quad (17)$$

Бесконечный граф БА характеризуется следующими стационарными вероятностями Q_k того, что случайно выбранная вершина имеет степень связности k

(Dorogovtsev S.N., Mendes J.F.F., Samukhin A.N., 2000):

$$Q_k = \frac{2m(m+1)}{k(k+1)(k+2)}, \quad k \geq m. \quad (18)$$

Недостаточная изученность более тонких структурных свойств графа БА требует существенного развития точных аналитических методов его исследования.

Для ориентированной версии графа БА **выведены уравнения баланса для стационарного совместного распределения вероятностей Q_{lk} конечных степеней l, k связности случайно выбранной дуги графа. Найдено его решение в общем виде, представленное в рекурсивной форме:**

$$Q_{l,k} = \begin{cases} 0, & l \geq m, k = m, \\ \frac{2}{(m+2)(2m+3)}, & l = m, k = m+1, \\ \frac{2(m+1)}{k(k+1)(k+m+2)} + \frac{(k-1)}{(k+m+2)} Q_{l,k-1}, & l = m, k \geq m+2, \\ \frac{(k-1)Q_{l,k-1} + (l-1)Q_{l-1,k}}{(l+k+2)}, & l \geq m+1, k \geq m+1. \end{cases} \quad (19)$$

Расчет матрицы $\mathbf{Q} = \| Q_{l,k} \|$ выполняется, начиная с заполнения крайнего левого столбца (столбца $k = m$) нулями и расчета элементов верхней строки (строки $l = m$): сначала по формуле (19) вычисляется ее второй слева элемент $Q_{m,m+1}$, затем, через него – следующие вправо элементы $Q_{m,k}$, $k \geq m+2$. После этого построчно, слева направо, рассчитываются элементы следующих строк матрицы \mathbf{Q} .

Результат (19) установлен впервые. В качестве частного случая он определяет следующее решение, известное для $m = 1$ (Krapivsky P.L., Redner S., 2001):

$$Q_{lk} = \frac{4(k-1)}{l(l+1)(l+k)(l+k+1)(l+k+2)} + \frac{12(k-1)}{l(l+k-1)(l+k)(l+k+1)(l+k+2)}, \quad (l, k \geq 1).$$

Найдены маргинальные распределения конечных степеней дуг графа:

$$Q_{k,*} = Q_k = \frac{2m(m+1)}{k(k+1)(k+2)}, \quad Q_{*,k} = \frac{2(k-m)(m+1)}{k(k+1)(k+2)}, \quad l, k \geq m,$$

где $Q_{k,*}$ ($Q_{*,k}$) – вероятность того, что начало дуги (конец дуги) имеет степень k . В среднем степень начала дуги равна $2m$, степень конца дуги в среднем бесконечна. Концевые степени дуги – зависимые сл.в. Выведенные отсюда для неориентированного графа БА маргинальные вероятности конечных степеней ребра («начало» и «конец» которого выбираются случайно) составляют:

$$\Theta_{k,*} = \Theta_{*,k} = \frac{Q_{k,*} + Q_{*,k}}{2} = \frac{(m+1)}{(k+1)(k+2)}.$$

Асимптотическое по l и k совместное распределение конечных степеней ребра

$$\Theta_{l,k} = \Theta_{k,l} \sim \frac{m(m+1)}{l(l+1)k(k+1)}, \quad \text{или, компактнее, } \Theta_{l,k} \sim m(m+1)l^{-2}k^{-2}$$

с ростом m сходится к распределению независимых сл.в.

Установлен асимптотический закон квадратичной пропорциональности между максимальной степенью K вершин графа БА и временем (числом шагов) N его выращивания: $K_{\text{ср}} = M(\max\{k_i\}) \propto N^{1/2}$ (рис. 2). С учетом этого закона с помощью полученных выше асимптотических распределений конечных степеней дуг выведено ранее для графа БА неизвестное аналитическое выражение важной структурной характеристики графа – коэффициента кластеризации. Коэффициент кластеризации C в общем случае определяется как утроенное отношение среднего числа n_{Δ} треугольников в графе к среднему числу n_V вилок (т.е. путей длины 2): $C = 3n_{\Delta}/n_V$. Для графа БА

$$C \sim \frac{(m-1)(\ln N)^2}{8N} + (m-1)c_m \frac{\ln N}{N}, \quad m \geq 2, \quad (20)$$

где коэффициент c_m можно вычислять путем ИМ. Так, при $m = 2, 3, \dots, 8$ его значения приблизительно равны 0.5, 0.35, 0.275, 0.18, 0.12, 0.10 и 0.07 соответственно (с ростом m коэффициент c_m убывает по экспоненте). Второе слагаемое в (20) можно отбросить, т.к. его отношение к первому сходится к нулю. Формула (20) иллюстрируется соответствующими спрямленными зависимостями (рис. 3).

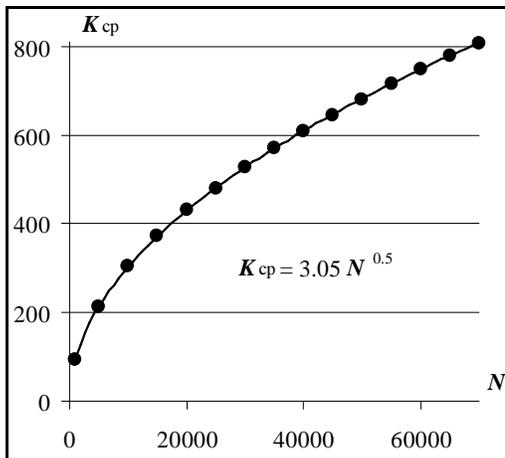


Рис. 2. Зависимость $K_{\text{ср}}$ от N при $m = 2$, построенная по данным ИМ (маркеры)

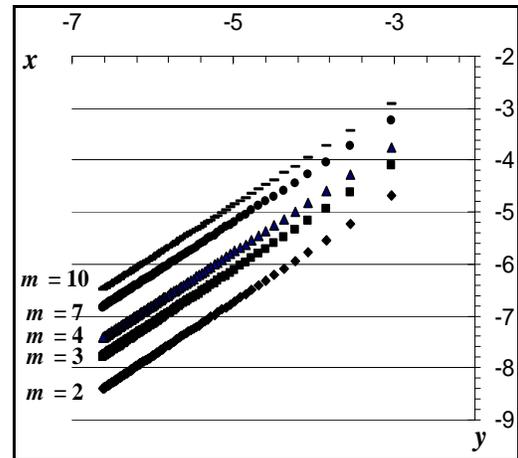


Рис. 3. Данные ИМ графов БА: зависимости показателя $y = \ln C$ от $x = \ln[(\ln N)^2 / N]$

Для улучшения согласования выращиваемых графов БА с моделируемыми реальными сетями разработан метод сепарабельной калибровки выращиваемого графа по коэффициенту кластеризации. Такая калибровка не влияет на распределение степени связности вершин (РСС) графа, но позволяет в широких пределах изменять коэффициент кластеризации. Так, при $N = 10^3 \div 10^6$ этот метод позволяет формировать граф с любым значением коэффициента в пределах от 0 до $(20 \div 3000)C$ соответственно, где C – номинальное его значение (20).

Разработана общая теория случайных графов с нелинейным правилом предпочтительного связывания. Граф БА, выращиваемый в соответствии с правилом (17), имеет при числе вершин $N \rightarrow \infty$ асимптотически степенное по k РСС (18). Однако в больших сетях типа Интернет РСС часто не являются асимптотически-степенными (Clauset A., Shalizi C. R., Newman M. E. J., 2009). Для по-

строения адекватных моделей таких сетей в диссертации предлагаются случайные графы с НППС. Вероятность p_i связывания нового ребра с i -й вершиной графа определяется нелинейным правилом предпочтительного связывания в виде

$$p_i = f(k_i) / \sum_j f(k_j), \quad (21)$$

где $f(k) = f_k > 0$ – вес (предпочтение) вершины со степенью k – произвольная функция от k , ($g \leq k \leq M \leq \infty$). При этом число ребер в приращениях графа может задаваться как независимая сл.в. x с распределением $r_k = P\{x = k\}$, $g \leq k \leq h$, ($h \leq M$). Т.о., алгоритм генерации (генератор) графа задается параметрами $\{f_k\}$ и $\{r_k\}$.

Выведены уравнения баланса для стационарных вероятностей Q_k степеней k вершин случайного графа с НППС. Найдено их решение в рекурсивной форме:

$$Q_g = \frac{r_g \langle f \rangle}{\langle f \rangle + m f_g}, \quad Q_k = \frac{r_k \langle f \rangle + m f_{k-1} Q_{k-1}}{\langle f \rangle + m f_k}, \quad k \geq g+1, \quad (22)$$

где $\langle f \rangle = \sum_k Q_k f_k$ – средний вес вершины, $m = M(x)$ – среднее число ребер в приращении графа. В частном случае, при $x \equiv m = g$, $f(k) = k$ из (22) вытекает известное распределение (18). Разработан метод быстрого численного расчета вероятностей Q_k в (22) с одновременным определением среднего веса $\langle f \rangle$ вершины.

Найдено условие существования стационарного графа: $M + 1 \geq 2m$. При $M + 1 = 2m$ генерируются *псевдорешетки* – бесконечные случайные графы, степень вершин которых с вероятностью единица равна $2m = M + 1$ (рис. 4).

Разработаны методы синтеза генераторов графов с НППС, позволяющие выращивать случайные графы с заранее заданными РСС.

Задача синтеза: для заранее заданного РСС $\{\tilde{Q}_k\} = \tilde{Q}_g, \dots, \tilde{Q}_{M+1} > 0$ со средним $\langle k \rangle \geq 2g$ найти вероятности $\{r_k\} = r_g, \dots, r_h$ и веса $\{f_k\} = f_g, \dots, f_M > 0$, при которых выращивается граф с РСС $\{Q_k\} = \{\tilde{Q}_k\}$.

Решение задачи выводится из (22):

$$f_g = \frac{r_g}{\tilde{Q}_g} - 1, \quad f_k = \frac{\tilde{Q}_{k-1}}{\tilde{Q}_k} f_{k-1} + \frac{r_k}{\tilde{Q}_k} - 1, \quad (k = g+1, \dots, M), \quad f_{M+1} = 0, \quad \langle f \rangle = m, \quad (23)$$

где $m = \langle x \rangle = \sum_k k r_k$. При $r_g = 1$ (фиксированное приращение графа: $x \equiv g = m$) реализуется наименьшее для данного g среднее $\langle k \rangle = 2g$. Вероятности r_g, \dots, r_h при заданном $\{\tilde{Q}_k\}$ необходимо задавать так, чтобы выполнялось условие $\langle k \rangle = 2m$. Если генератор использует вычисленные по заданному $\{\tilde{Q}_k\}$ неотрицательные веса (23), выполнено условие $M + 1 \geq 2m$ и $\langle k \rangle = 2m \geq 2g$, то генератор реализует требуемое распределение $\{Q_k\} = \{\tilde{Q}_k\}$ и при этом $\langle f \rangle = m$. Справедливость этого утверждения доказывается подстановкой формул (23) в формулы (22).

Примечания. 1). Значения r_g, \dots, r_h , выбираемые с учетом условия $\langle k \rangle = 2m$, должны быть такими, чтобы веса (23) были неотрицательными. Для этого необходимо и достаточно при всех $k \leq h$ соблюдать условие $\sum_{i=g}^k r_i \geq \sum_{i=g}^k \tilde{Q}_i$. Это ус-

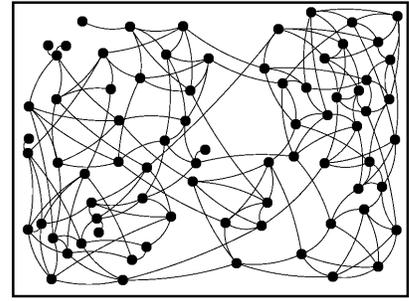


Рис. 4. Начальный фрагмент псевдорешетки, выращиваемой из 10-вершинной почочки при $f_1, \dots, f_4 = 1, 10, 100, 1000$, $M = 4, m = 2,5$

ловие позволяет обеспечивать выполнение равенства $\langle k \rangle = 2m$ при любом $\langle k \rangle \geq 2g$.

2). Стационарный режим существует, если $M + 1 \geq 2m$, однако вероятность его реализации при немедленном связывании поступающих приращений графа может быть меньше единицы. Так, при $r_1 = \dots = r_4 = 1/4$ и любом наборе весов $f_1, \dots, f_4 > 0$ (здесь $m = 2,5$, $M + 1 = 5 = 2m$) бесконечная псевдорешетка выращивается с нулевой вероятностью. Здесь после конечного числа шагов неизбежно возникает ситуация, когда после связывания нескольких ребер очередного приращения графа степень всех вершин становится равна *пяти*. Но поскольку $f_5 = 0$, остальные ребра приращения становится невозможно связать ни с одной из вершин. В подобных случаях для достоверной реализации стационарного режима те приращения графа, которые не могут быть добавлены на данном шаге, можно помещать в очередь для последующего их добавления при первой возможности. Тогда при $t \rightarrow \infty$ если $M + 1 > 2m$, то средняя длина $L(t)$ очереди приращений графа сходится к нулю, а если $M + 1 = 2m$, то $L(t) \rightarrow \infty$, но $L(t)/N \sim L(t)/t \rightarrow 0$.

Разработана методика калибровки генераторов графов с НППС по эмпирическим РСС реальных сетей, состоящих из сотен тысяч узлов и связей (*Задорожный В.Н., Проблемы управления, 2010*). Аппарат теории графов с НППС реализован в системе ИМ. Имеется свидетельство о ее официальной регистрации.

В седьмой главе приводятся основные прикладные результаты диссертационного исследования, включающие решение тестовых задач разработанными аналитико-имитационными методами.

1. Решается задача оптимизации измерений производительности ЭВМ, осуществляемых с помощью выборочного микропрограммного измерителя. Задача состоит в определении максимальной интенсивности выборок, т.е. максимальной дополнительной загрузки ρ процессора измерителем, при которой погрешность измерения средней длины очереди прикладных программ, обусловленная загрузкой ρ , не превышает заданной величины δ . Адекватным представлением рассматриваемой системы является СМО $GI_2|GI_2|1$ с пуассоновским потоком неприоритетных заявок. С помощью асимптотик, найденных в главе 3, задача сводится к алгебраическому уравнению второй степени и решается в общем виде:

$\rho = -p/2 - \sqrt{p^2/4 - q}$, где

$$p = -(2 - \rho') - \lambda' \bar{x} \cdot \frac{(1 - \rho')}{\rho'^2} \cdot \left(\frac{C_x^2 + C_x^2}{1 + C_x^2} \right) \cdot \frac{1}{(1 + \delta)}, \quad q = (1 - \rho') \cdot \frac{\delta}{1 + \delta}.$$

Параметры приоритетных заявок здесь соответствуют операциям измерителя, реализующим выборки, а параметры неприоритетных заявок – прикладным программам. Проверка полученного аналитического решения ρ путем ИМ системы показывает его вполне достаточную для инженерной практики точность. Так, в условиях изменения от 0,1 до 0,8 загрузки ρ' процессора прикладными программами найденная предельная его загрузка ρ измерителем действительно обуславливает погрешности измерений 4,8-4,9% , близкие к заданной границе $\delta = 5\%$.

2. Решается задача оптимизации режима работы серверного центра. Посредством асимптотик, полученных для ЦСО в главе 4, задача сводится к системе

алгебраических уравнений и решается элементарными численными методами. Задача представляет собой комплексный тест для аналитико-имитационных методов, разработанных в главах 2-4. В моделируемой системе (рис. 5) сочетается ряд типичных особенностей, не позволяющих эффективно использовать аналитическое или непосредственное имитационное моделирование:

- не экспоненциальные распределения и многоканальные узлы;
- относительно интенсивные потоки заявок с абсолютными приоритетами;
- пакетирование неприоритетных заявок;
- циклическое ежесуточное изменение интенсивностей входных потоков;
- медленный тренд интенсивностей на больших периодах времени;
- малые значения вероятности – показателя, который нужно минимизировать.

Ежесуточно обслуживается два пакета: пока один накапливается в коммутационной ЭВМ, другой обслуживается многопроцессорным узлом. Накопление пакетов осуществляется непрерывно. Задача состоит в выборе оптимального времени T_1 начала часа профилактики (ЧП), в течение которого пакеты не обслуживаются (один пакет обслуживается от начала суток до T_1 , другой – от $T_1 + 1$ ч до начала следующих суток). Необходимо минимизировать вероятность того, что обслуживание хотя бы одного из двух пакетов не будет завершено вовремя.

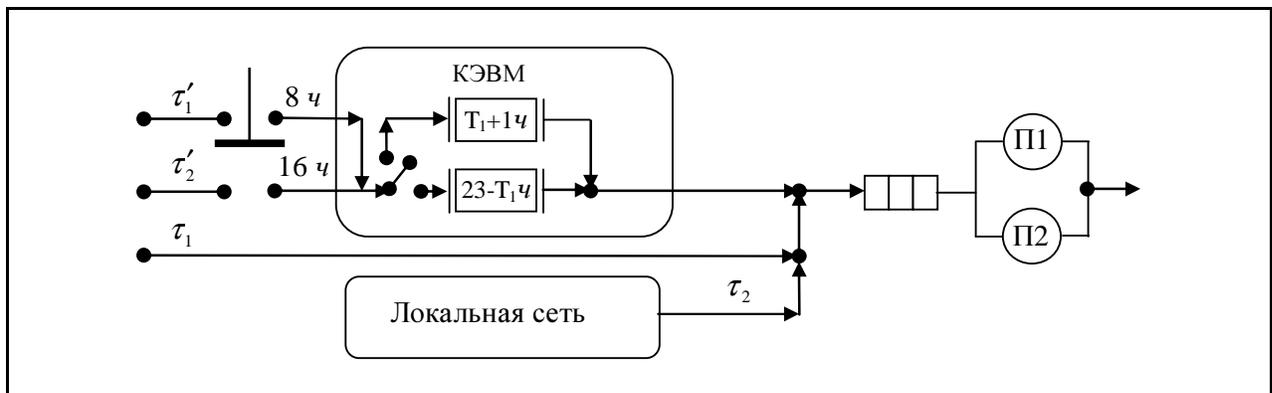


Рис. 5. Структура серверного центра

Задача сведена к системе алгебраических уравнений для времени T_1 начала ЧП:

$$\frac{60 \cdot (23 - T_1) - \bar{y}_1}{\bar{y}_1 \cdot C_{y1}} = \frac{60 \cdot T_1 - \bar{y}_2}{\bar{y}_2 \cdot C_{y2}}, \quad \bar{y}_1 = \frac{\bar{\xi}_1}{n(1-\rho)}, \quad \bar{y}_2 = \frac{\bar{\xi}_2}{n(1-\rho)},$$

$$C_{y1}^2 = \frac{n\rho}{(1-\rho)} \cdot (C_\tau^2 + C_x^2) \cdot \frac{\bar{x}}{\bar{\xi}_1}, \quad C_{y2}^2 = \frac{n\rho}{(1-\rho)} \cdot (C_\tau^2 + C_x^2) \cdot \frac{\bar{x}}{\bar{\xi}_2},$$

$$\bar{\xi}_1 = 60 \cdot (T_1 - 4,333), \quad \bar{\xi}_2 = 60 \cdot (23 - T_1), \quad (24)$$

где $\rho, \bar{x}, C_\tau, C_x$ – известные числовые характеристики приоритетного потока заявок, n – число процессоров. Путем численного решения уравнений (24) легко определяется искомое $T_1 = 12,58$ ч и соответствующие вероятности P1 и P2 несвоевременного обслуживания первого и второго пакета: $P1 = P2 = 1,1 \cdot 10^{-8}$.

Наряду с оптимальным временем T_1 начала ежесуточного ЧП решается ряд других вопросов, в частности, определяются сроки наращивания числа процессо-

ров. Полученные решения (часть которых иллюстрируется рисунком б), проверяются и подтверждаются непосредственным ИМ. На рис. 6 б) маркерами × и + отмечены результаты *точечных проверок*, полученные непосредственным ИМ.

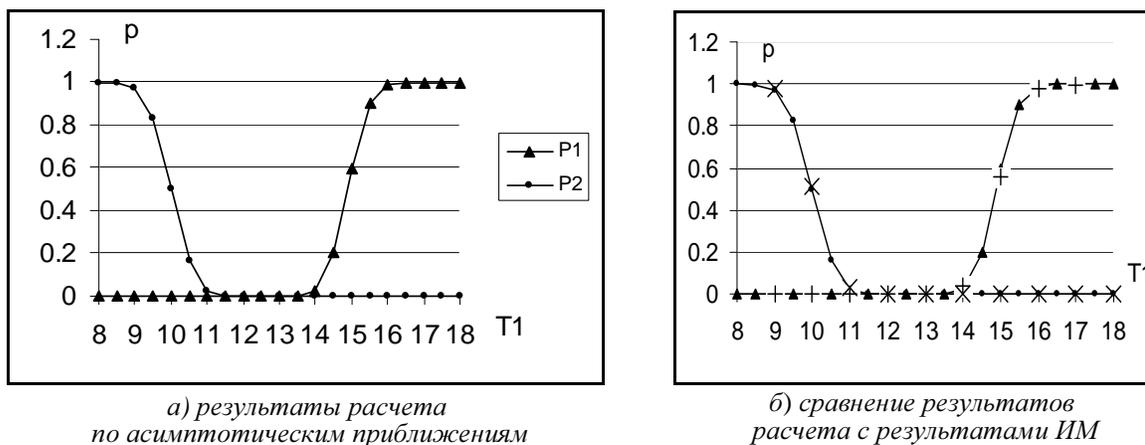


Рис. 6. Зависимость вероятностей P1, P2 от начала T1 часа профилактики

Применение здесь формул глав 2–4 с учетом времени их точечных проверок ускорило решение по сравнению с непосредственным ИМ примерно в 10^4 раз.

3. С применением ИМ выполнено тестирование точности и быстродействия метода НГ. В испытаниях выполнялась оптимизация пяти однородных немарковских СеМО (рис. 7, 8) с одно-

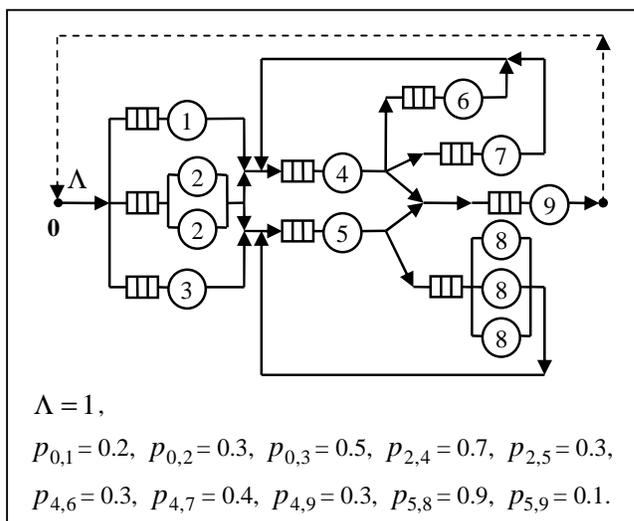


Рис. 7. Тестовый пример СеМО-1. Штриховая дуга соответствует замкнутой версии сети

и многоканальными узлами, с набором п.в. времени обслуживания в разных узлах, характеризуемым большим разбросом к.в. Число узлов в тестовых сетях менялось от 9 до 100, а общее число каналов – от 12 до 260. Результаты сравнивались с результатами ИМ, полученными применением: а) базового «классического» метода приращений и б) программы OptQuest, занимающей лидирующие позиции на рынке универсальных оптимизаторов.

При сравнении установлены принципиальные преимущества метода НГ как в скорости, так и в точности решения задачи. Так, заметно выигрывая у базового метода в глубине оптимизации, метод НГ превосходит его и в быстродействии. Например, решение, определяемое для СеМО-1 методом НГ за 7-11 итераций, обеспечивает среднее время ответа $E \approx 7,49 \dots 7,51$, а базовый метод за 100...140 итераций (т.е. за 1000...1400 прогонов модели) дает более слабый результат $E \approx 9,3 \dots 9,5$. Это преимущество достигается за счет резкого снижения числа итераций и за счет того, что для вычисления градиента на каждой итерации метод НГ использует один прогон имитационной модели, а не $(n + 1)$ прогонов, как базовый метод. При оптимиза-

пии СеМО-1 с помощью программы OptQuest за 750...1000 прогонов модели (т.е. при стократном превышении времени, затраченного методом НГ) достигается среднее время ответа $E \approx 7.91$, также заметно проигрывающее результату метода НГ.

Фазовая погрешность $\delta = 100\% \times |\bar{\mu}_{opt} - \bar{\mu}^*| / |\bar{\mu}_{opt}|$ метода НГ в тестовых задачах (табл. 3) не выходила за пределы 1%, и лишь в специально сконструированных «жестких» случаях – при перепаде значений к.в. времени обслуживания в разных узлах от 0 до 5 – достигала 3% (погрешность δ метода НГ оценивалась сверху на втором этапе двухэтапного метода). Здесь $\bar{\mu}_{opt}$ – оптимальное решение, $\bar{\mu}^*$ – решение методом НГ. Упущенный эффект $\varepsilon = 100\% \times [E(\bar{\mu}^*) - E(\bar{\mu}_{opt})] / E(\bar{\mu}_{opt})$ всегда составлял доли процента. Испытания метода на замкнутых сетях, функция $E(\bar{\mu})$ в которых «сильно не сепарабельна», показали, что при линейном ограничении (13) значение δ может достигать 3%, а при нелинейном, с разбросом коэффициентов β_i от 0,5 до 1,5, δ составляет (6-8)%. Но упущенный эффект остается достаточно малым ($\varepsilon < 1\%$).

Как видно из табл. 3, при числе узлов и каналов, достигающем сотен (СеМО-5), когда высокие затраты времени на ИМ существенно ограничивают возможность увеличения числа итераций, метод НГ позволяет достигать значительного сокращения времени E при числе итераций N , меньшем размерности n факторного пространства.

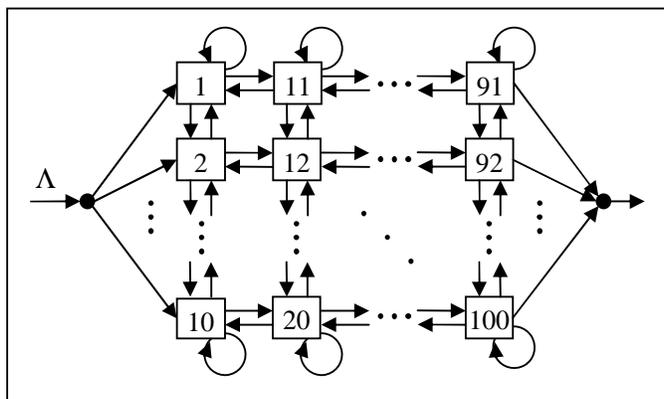


Рис. 8. Структура межузловых переходов в СеМО-5

Таблица 3

Результаты испытаний метода НГ на тестовых сетях СеМО-2, ..., СеМО-5

	СеМО-2		СеМО-3		СеМО-4		СеМО-5	
Число узлов n , общее число каналов K	$n=9, K=12$		$n=30, K=134$		$n=20, K=52$		$n=100, K=260$	
Число итераций $N \times$ время итерации T_M (мин)	20×2		60×10		40×5		50×94	75×94
Распределяемый ресурс M	24	32	96	128	39	52	437	582
Отклик \hat{E}_c в центре ОДР	13,98	8,043	49,92	27,88	78,08	38,19	278,1	163,3
Отклик $\hat{E}^* = \hat{E}(\bar{\mu}^*)$	10,83	6,225	41,53	22,97	55,12	29,94	249,6	143,8
Эффект оптимизации $(\hat{E}_c - \hat{E}^*) / \hat{E}_c \cdot 100\%$	22,5%	22,6%	17%	18%	29%	22%	10%	12%

4. Решаются задачи калибровки генераторов с НППС по эмпирическим РСС реальных больших сетей. Калиброванные генераторы выращивают графы с распределениями степени связности (рис. 9 – 11, см. сплошные линии), хорошо согласующимися с эмпирическими РСС (см. рис. 9 – 11, маркеры). Эти генераторы можно рассматривать и как результаты структурной идентификации реальных сетей. Рис. 9 отражает результат идентификации сети автономных систем Интернет по реальным данным, включающим параметры $N = 22\,961$ узлов и $R = 48\,436$ ребер (2006 г.). Параметры калиброванного генератора: $r_1 = 0,342$, $r_2 = 0,432$,

$r_3 = 0,096$, $r_4 = 0,13$, $m = r_1 + 2r_2 + 3r_3 + 4r_4 = 2,014$, $f_1, \dots, f_4 = 0,0017, 0,0245, 0,0999, 2,5303$ соответственно, и $f_k = 0,8603 \cdot k$ при $k \geq 5$.

Рис. 10 характеризует качество идентификации сети маршрутизаторов Интернет ($N = 124\ 651$, $R = 207\ 214$). Параметры генератора: $r_1 = 0,435$, $r_2 = 0,565$, $f_1 = 0,272$, $f_2 = 1,224$, $f_k = 0,512 \cdot \ln(k) + 0,372 \cdot k$, $k \geq 3$. Рис. 11 характеризует идентификацию сети участия актеров в общих фильмах ($N = 511\ 416$ – без изолированных узлов, $R = 1\ 463\ 331$). Два соответствующих актерам узла связаны ребром, если эти актеры снимались в одном фильме. Здесь $r_1, \dots, r_8 > 0$, численно фиксированы f_1, \dots, f_{10} , а при $k > 10$ общая формула весов имеет вид $f_k = 4,429 \cdot \ln(k)$.

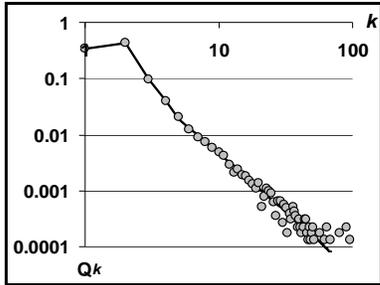


Рис. 9. РСС калиброванного графа и сети автономных систем Интернет

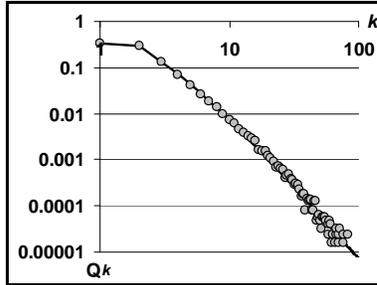


Рис. 10. РСС калиброванного графа и сети маршрутизаторов Интернет

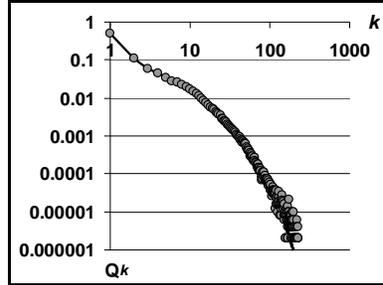


Рис. 11. РСС калиброванного графа и сети участия актеров в общих фильмах

5. Калиброванные графовые модели сетей можно использовать для анализа и оптимизации различных сетевых процессов. Интернет – «сеть сетей», представляет собой иерархическую структуру, модулями которой являются автономные системы (АС). Каждая АС – это система IP-сетей, имеющих свои уникальные IP-адреса, и маршрутизаторов (шлюзов), управляемая одним или несколькими операторами, которые осуществляют единую политику маршрутизации сообщений в пределах АС, независимую от остальной части Интернета. В состав АС может входить от единиц до нескольких тысяч компьютеров. Связь между АС осуществляют «граничные» шлюзы, прокладывающие маршруты между АС, рассматриваемыми как неделимые элементы и имеющими минимальный набор параметров, необходимых для маршрутизации. Этот набор включает уникальный номер АС, количество IP-сетей в АС, их адреса и внутренние расстояния до этих сетей от данного граничного шлюза. Граф сети автономных систем, выращиваемый калиброванным генератором, можно использовать для анализа ее связности при случайных отказах элементов (т.е. при потере связи с отдельными АС) и для расчета критической вероятности отказа, при достижении которой сеть распадается на несвязные компоненты. Администраторы АС, учитывая результаты такого анализа, могут уточнять свой выбор числа провайдеров и смежных АС для соединения с Интернетом. Уточнение выбора может привести к системному эффекту, несколько изменяющему ход развития Интернета. Этот эффект можно оценивать заранее, генерируя графы по первоначальным и по скорректированным правилам предпочтительного связывания, и используя в качестве затравки известное текущее состояние сети АС.

Калиброванный случайный граф, многократно реализуемый для определения средних значений искомым показателей, имеет ряд преимуществ по сравнению с графом известного мгновенного состояния сети. Поскольку большие сети (например, отдельные АС и Интернет в целом) быстро изменяются, сохраняя лишь свои существенные структурные свойства, то оптимизация решений под фиксированное мгновенное состояние сети может приводить при их реализации к результатам, далеким от ожидаемых. Отладку решений, тиражируемых для применения в разных вариантах системы, также лучше осуществлять на нескольких реализациях калиброванного случайного графа, моделирующего систему. Кроме того, калиброванные графы позволяют моделировать развитие сетей в будущем.

Графы социальных сетей, подобных сети участия актеров в общих фильмах, могут применяться в социальных исследованиях и проектах. Логарифмическая весовая функция в генераторе графа участия актеров в общих фильмах, найденная при калибровке (рис. 11), может быть объяснена типичной для человека логарифмической шкалой восприятия уровня сигналов (в данном случае – актерской славы). Важным приложением найденных методов в социальной сфере является исследование сетей распространения инфекций и разработка на калиброванных моделях этих сетей надежных методов борьбы с эпидемиями. В целом область приложения разработанных методов не менее широка, чем у графов БА.

В приложении Ж приводятся сведения о разработанных в ходе диссертационного исследования комплексах программ, реализующих методы АИМ систем с очередями и стохастических сетей, об их внедрении и государственной регистрации, копии свидетельств о регистрации, актов внедрения и актов приемки-сдачи.

В заключении сформулированы основные результаты работы, включающие:

- методы асимптотического анализа цикла обслуживания СМО с абсолютными приоритетами $GI_2|GI_2|n$ и $GI_2|GI_2|n|$, $n \geq 1$, приближенные формулы для очередей в этих СМО и ускоренный метод их имитационного моделирования;
- точное C -соотношение, связывающее первые два момента периодов занятости и незанятости в системе $GI|GI|1$ и его обобщение на системы $GI|GI|n$;
- математическую модель систем обслуживания с циклическим пакетированием неприоритетных заявок и результаты ее асимптотического анализа;
- эффективные двухуровневые аналитико-имитационные методы оптимизации однородных немарковских СеМО;
- расширенный метод редукции графовых моделей полумарковских процессов с вложенной в моменты смены состояний марковской цепью;
- методы и результаты точного анализа структурных свойств графа БА;
- основы общей теории случайных графов с НППС, включающие решение в общем виде задачи анализа и задачи синтеза генераторов графов, а также методику калибровки генераторов графов по эмпирическим РСС узлов в моделируемых больших стохастических сетях;
- численные методы и комплексы программ, реализующие разработанные математические модели и методы АИМ систем с очередями и стохастических сетей.

Основные результаты диссертации опубликованы в следующих работах

В монографиях

1. Задорожный В.Н. Аналитико-имитационные исследования систем и сетей массового обслуживания: *монография* / В.Н. Задорожный. – Омск: Изд-во ОмГТУ, 2010. – 280 с.
2. Задорожный В.Н. Аналитико-имитационные исследования Больших Сетевых Структур: *монография* / В.Н. Задорожный. – Омск: Изд-во ОмГТУ, 2011. – 208 с.

В журналах перечня ВАК

3. Задорожный В.Н. Асимптотический анализ систем с интенсивными прерываниями // Автоматика и телемеханика, 2008. – № 2 – С. 86-96.
4. Задорожный В.Н. Оптимизация однородных немарковских сетей массового обслуживания // Проблемы управления, 2009. – № 6. – С. 68-75.
(Zadorozhnyi, V.N. Optimizing Uniform Non-Markov Queueing Networks // Automation and Remote Control ISSN 0005-1179. – Vol. 71, No. 6, 2010. – P1158-1169).
5. Задорожный В.Н. Случайные графы с нелинейным правилом предпочтительного связывания // Проблемы управления, 2010. – №6. – С. 2-11.
6. Задорожный В.Н., Литунов С.Н., Штриплинг С.Л. К вопросу о проблеме муара при воспроизведении изображений печатными способами // Известия высших учебных заведений. Проблемы полиграфии и издательского дела, 2007. – № 1. – С. 30-39.
7. Задорожный В.Н. Общая статистическая структура простейших клеточных автоматов // Омский научный вестник, 2005. – № 2 (31) – С. 150-156.
8. Задорожный В.Н. Анализ систем с приоритетами методом декомпозиции // Омский научный вестник, 2005. – № 3 (32) – С. 126-132.
9. Задорожный В.Н., Пуртов А.М. Анализ чувствительности в имитационном моделировании сетей массового обслуживания // Омский научный вестник, 2005. № 4 (33). – С. 165-171.
10. Задорожный В.Н. Асимптотический анализ периодов повышенной нагрузки в приоритетных системах // Омский научный вестник, 2006. – № 3 (36). – С. 117-124.
11. Задорожный В.Н., Семенова И.И. Управление сложными техническими объектами и парадигмы имитационного моделирования // Омский научный вестник, № 6 (41), 2006. – С. 102-108.
12. Задорожный В.Н., Ершов Е.С., Канева О.Н. Двухуровневые градиентные методы для оптимизации сетей с очередями // Омский научный вестник, 2006. – № 7 (43). – С. 123-131.
13. Задорожный В.Н. Распределение календарного времени обслуживания неприоритетных заявок в системах с абсолютными приоритетами // Омский научный вестник, 2006. – № 8 (44). – С. 124-131.
14. Задорожный В.Н. Комбинированный метод моделирования циклических систем обслуживания // Омский научный вестник, 2006. – № 9 (46). – С. 156-163.
15. Задорожный В.Н. О качестве программных генераторов случайных чисел // Омский научный вестник, 2009. – № 2 (80). – С. 199-205.
16. Задорожный В.Н., Ершов Е.С. Градиентный метод и программа оптимизации сетей с очередями // Омский научный вестник, 2009. – № 2 (80). – С. 205-210.

17. Задорожный В.Н., Юдин Е.Б. Статистически однородные случайные графы: определение, генерация, применение // Омский научный вестник. – 2009. – № 3 (83). – С. 7-13.

18. Задорожный В.Н., Юдин Е.Б. Точная теория графа Барабаши-Альберт // Омский научный вестник. – 2009. – № 3 (83). – С. 13-18.

19. Задорожный В.Н. Распределение каналов в однородных немарковских сетях с очередями // Омский научный вестник, 2010. – № 1 (87). – С. 5-10.

20. Задорожный В.Н. Методы калибровки аддитивных генераторов стохастических графов // Омский научный вестник, 2010. – № 1 (87). – С. 171-176.

21. Задорожный В.Н. Предпосылки создания фрактальной теории массового обслуживания // Омский научный вестник, 2010. – № 2 (90) – С. 182-187.

22. Задорожный В.Н., Тулубаев Д.А. Мета модель систем оперативно-диспетчерского управления сложными крупномасштабными организационно-техническими объектами // Омский научный вестник, 2011. – № 1 (97) – С. 19-23.

В материалах международных и всероссийских научно-технических конференций

23. Задорожный В.Н. Методы ускоренной имитации процессов с интенсивными прерываниями. – Имитационное моделирование. Теория и практика: Материалы 2-й всерос. конф., Том 1. – СПб.: ФГУП ЦНИИТС, 2005. – С. 101-106.

24. Задорожный В.Н. Асимптотические приближения в имитационном моделировании систем с очередями // Имитационное моделирование. Теория и практика: Материалы 3-й всероссийской конференции. Том 1. – СПб.: ФГУП ЦНИИТС, 2007. – С. 117-122.

25. Задорожный В.Н. Методы двухуровневого моделирования систем с очередями // Труды VII Международной конференции «Идентификация систем и задачи управления» SICPRO '08. – Москва, 28-31 января 2008. – С. 1484-1563.

26. Задорожный В.Н. К дискуссии о качестве датчиков случайных чисел. Имитационное моделирование. Теория и практика (ИММОД-2009): Материалы 4-й всероссийской конференции. Том 1. – СПб.: ЦТ СС, 2009. – С. 128-134.

27. Задорожный В.Н. Оптимизация немарковских сетей с очередями // Динамика систем, механизмов и машин: Матер. VII междунар. науч.-техн. конф. – Омск, 2009. Кн. 1. – С. 288-292.

Свидетельства о регистрации комплексов программ

28. Задорожный В.Н., Юдин Е.Б. Система агентного моделирования «Simbigraph» (ОФЭРНиО № 16539) / М.: ИНИМ РАО, ОФЭР «Наука и образование», – 2011. – № 50201050316.

29. Ершов Е.С., Задорожный В.Н. Комплекс программ оптимизации сетей массового обслуживания «RedOpt» (ОФЭРНиО № 16589) / М.: ИНИМ РАО, ОФЭР «Наука и образование», – 2011. – № 50201150081.

30. Юдин Е.Б., Задорожный В.Н. Система агентного моделирования «Simbigraph» / Свидетельство № 2011612193 о государственной регистрации программы для ЭВМ / Федеральная служба по интеллектуальной собственности, патентам и товарным знакам. – М.: РОСПАТЕНТ, – 2011. – № 2011612193.