**Гиндин Сергей Игоревич. Модели и алгоритмы оценивания оперативности распределенной обработки данных в информационных. системах с учётом затрат на актуализацию контекста: диссертация ... кандидата технических наук: 05.13.18 / Гиндин Сергей Игоревич;[Место защиты: Петербургский государственный университет путей сообщения].- Санкт-Петербугр, 2014.- 116 с.**

**ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО**

**ТРАНСПОРТА**

**Федеральное государственное бюджетное образовательное учреждение**

**высшего профессионального образования**

**«Петербургский государственный университет путей сообщения**

**Императора Александра I» (ФГБОУ ВПО ПГУПС)**

На правах рукописи



**Гиндин Сергей Игоревич**

**МОДЕЛИ И АЛГОРИТМЫ ОЦЕНИВАНИЯ ОПЕРАТИВНОСТИ**

**РАСПРЕДЕЛЕННОЙ ОБРАБОТКИ ДАННЫХ В**

**ИНФОРМАЦИОННЫХ СИСТЕМАХ С УЧЕТОМ ЗАТРАТ НА**

**АКТУАЛИЗАЦИЮ КОНТЕКСТА**

**Специальность 05.13.18 – «Математическое моделирование, численные**

**методы и комплексы программ» Диссертация на соискание ученой степени кандидата технических наук**

**Научный руководитель: д.т.н., профессор А.Д. Хомоненко**

**Санкт-Петербург 2014**

2

**Содержание**

Введение 4

1. Объект, цель и задачи исследования 15

1. Описание объекта исследований 15
2. Требования к оперативности функционирования ИС 18
3. Обоснование показателей и критериев оперативности функционирования ИС 20
4. Анализ моделей ИС для расчета показателей оперативности 24
5. Анализ возможностей исследования показателей оперативности ИС методами теории массового обслуживания 28
6. Постановка задачи исследования 33

Выводы по разделу 35

2. Комплекс моделей многоканальных немарковских СМО с «разогревом»37

1. Классификация СМО 37
2. Современные методы анализа СМО 40
3. Учет влияния «разогрева» на показатели оперативности 42
4. Аппроксимация немарковских распределений с помощью распределений фазового типа 45
5. Модели многоканальных немарковских СМО с «разогревом» на основе гиперэкспоненциального распределения 52
6. Расчет вероятностно-временных характеристик оперативности для СМО с «разогревом» 66

Выводы по разделу 72

3. Программная реализация моделей СМО «с разогревом» для расчета
показателей оперативности и ее тестирование 75

1. Описание программной реализации 75
2. Задание исходных данных для тестирования комплекса 77
3. Подход к тестированию программной реализации комплекса 79
4. Оценка погрешностей 84

3

Выводы по разделу 87

4. Практическое применение результатов исследований 89

1. Моделирование системы управления финансами 89
2. Моделирование системы управления программой лояльности 94
3. Методические рекомендации по внедрению 98

Выводы по разделу 99

Заключение 101

Список использованных источников 103

Приложение 1. Исходные тексты основных функций комплекса программ
расчета характеристик оперативности ИС 112

4

**Введение**

Отличительной чертой современных информационных систем (ИС) является устойчивая тенденция перехода к системам, основанным на принципах распределенной обработки данных, когда процесс обработки выполняется на взаимодействующих между собой, но независимых компьютерах, в рамках совместного решения общей задачи.

Появление и широкое распространение распределенных ИС объясняется, в первую очередь, заметным развитием производства аппаратных средств, снижением себестоимости производства микроэлектронных компонентов, повышением их вычислительных характеристик при уменьшении размеров, а также повышением требований к оперативности и эффективности вычислительных систем, диктуемых решаемыми практическими задачами.

Распределенные ИС позволяют централизовать управление вычислительными ресурсами, абстрагировать их предоставление и оптимизировать их потребление по сравнению с традиционными схемами построения информационных систем, тем самым снизить конечную стоимость содержания и обслуживания ИС для ее потребителя. Значительную часть затрат на эксплуатацию распределенных систем обработки данных составляет потребление электроэнергии, которое напрямую связано с количеством эксплуатируемых вычислительных ресурсов.

При применении распределенных ИС в вычислениях на первый план выступают такие аспекты, как оперативность функционирования распределенной среды обработки данных, ограничения пропускной способности сети и защита данных, передаваемых между компонентами распределенной системы. Эти факторы способны серьезно повлиять на эффективное использование вычислительных ресурсов, являющееся на

5

сегодняшний день наиболее значительным преимуществом распределенных ИС.

Оценивание оперативности функционирования системы позволяет определять количество ресурсов, которое потребуется для того или иного вычислительного процесса. Наличие таких данных, в свою очередь, позволяет повысить точность прогнозирования стоимости работы системы, связанной с распределенными вычислениями.

Задача оценивания оперативности функционирования распределенных ИС является актуальной и важной не только для существующих продуктов, но и в процессе производства и проектирования ИС, при выборе архитектурных решений и используемых технологий.

За счет эффективного планирования вычислительных ресурсов при разработке распределенных систем достигается оптимизация сроков на разработку и, соответственно, снижение себестоимости, значительное повышение эффективности разработки в целом.

Для реализации распределенной обработки информации характерны следующие особенности:

· независимая и параллельная обработка данных на нескольких
компьютерах;

· регулярные взаимодействия между компонентами внутри структуры
распределенных ИС;

· изменение структуры компонентов ИС, участвующих в
вычислительном процессе.

Таким образом, в системах распределенной обработки данных необходимо уметь оценивать быстродействие обслуживания в условиях масштабирования инфраструктуры. Для корректной оценки оперативности важно правильно понимать характер происходящих в ИС процессов и изменений. Существенной особенностью современных распределенных

6

систем является происходящее в силу перераспределения общих ресурсов между узлами ИС периодическое обновление вычислительного контекста.

Под вычислительным контекстом понимается набор информации, необходимый системе для выполнения текущей задачи. Как правило, ИС поддерживает в своей области оперативной памяти одновременно контексты многих задач, что позволяет системе переключаться между различными задачами, приостанавливать и возобновлять выполнение задач с прерванного места.

В случае распределенных ИС вычислительные контексты хранятся независимо в различных частях системы, копии одних и тех же данных многократно растиражированы в локальных контекстах узлов. В ходе вычислительного процесса используемые данные могут изменяться на выделенном узле или части распределенной ИС. Для обеспечения слаженной работы всей распределенной системы в данных условиях требуется в случае изменений в одной копии данных обеспечить актуализацию контекста во всей системе — соответствующее изменение контекстов остальных компонентов и связанных систем. Данная операция как правило, требует дополнительных временных затрат при каждой актуализации контекста.

Таким образом, для распределенных систем существует **проблемная ситуация**: необходимость организации управления вычислительными ресурсами ИС при несовершенстве современных методов учета влияния временных затрат по актуализации контекста на оперативность вычислительных процессов.

Для решения указанных задач могут быть использованы средства математического моделирования функционирования распределенных ИС. В связи с увеличением применения распределенных систем в инженерной практике, этой теме уделяется большое внимание в научных публикациях и литературе. В работах Кудашева Е.Б., Мулюхи В.А., Федотова А.М., Филонова А.Н., Швецова А.Н., Шокина Ю.И., Яковлева В.В., Яковлева С.А.

7

выполнен анализ принципов построения распределенных ИС. Работы Бубнова В.П., Вишневского В.М., Грассмана В., Дудина А.Н., Жожикашвили В.А., Колахи С., Крейнина А.Я., Латуша Г., Рыжикова Ю.И., Сафонова В.И., Таранцева А.А., Тырвы А.В., Хомоненко А.Д., Эбата Дж. посвящены подходам к моделированию, исследованию и расчету характеристик для моделей, применимых для распределенных ИС. Получены некоторые характеристики оперативности на основе построенных моделей. Тем не менее, как показано в первом разделе, существующие модели обладают рядом недостатков, ограничивающих их использование для решения названных задач:

· при моделировании распределенных систем не учитываются
дополнительные временные затраты на актуализацию контекста,
которые на практике значительно влияют на показатели
оперативности;

· модели не обладают общностью, поскольку строятся на основе
определенных реализаций систем распределенных вычислений;

· результаты моделирования трудно использовать при практическом
прогнозировании характеристик оперативности, поскольку результаты
сильно зависят от типичных наборов задач, характерных для
рассматриваемой распределенной системы;

Указанные обстоятельства определяют **актуальность** исследуемой в диссертации **научной задачи**: оценивания оперативности обработки данных в узлах распределенных информационных систем с учетом дополнительных временных затрат на актуализацию контекста.

**Объект исследования**: узлы систем распределенной обработки данных.

8

**Предмет исследования**: модели и алгоритмы оценивания оперативности функционирования систем распределенной обработки данных.

**Цель исследования**: повышение точности расчета характеристик оперативности распределенной обработки данных в узлах информационных систем с учетом дополнительных временных затрат на актуализацию контекста («разогрев»).

**Задачи исследования включают**:

· проведение анализа современного состояния исследований в области
оценивания оперативности функционирования систем распределенной
обработки данных;

· разработка моделей многоканальных немарковских систем массового
обслуживания с «разогревом», обеспечивающих возможность учета
дополнительных временных затрат при оценивании оперативности
функционирования узлов в информационных системах распределенной
обработки данных;

· разработка численного метода расчета вероятностно-временных
характеристик многоканальных немарковских систем массового
обслуживания с «разогревом»;

· разработка программного комплекса расчета характеристик
оперативности для разработанного класса многоканальных моделей
систем массового обслуживания;

· разработка методики оценки вероятностно-временных характеристик
оперативности функционирования узлов систем распределенной
обработки данных;

· анализ характеристик оперативности функционирования узлов систем
распределенной обработки данных на примере распределенной
автоматизированной системы управления финансами; оценка
эффективности решения.

9

Основным **методом исследования** является математическое моделирование с использованием теории массового обслуживания, теории графов, математической статистики, теории вероятностей, теории дифференциальных уравнений.

**Научная новизна**

1. Разработанные математические модели оценивания оперативности обработки данных в узлах информационных систем на основе многоканальных немарковских систем массового обслуживания с «разогревом» обеспечивают возможность учета дополнительных временных затрат на актуализацию контекста.
2. Разработанная модификация общего итерационного метода Такахаши-Таками позволяет рассчитывать вероятностно-временные характеристики многоканальных немарковских систем массового обслуживания с «разогревом» на основе двухфазной гиперэкспоненциальной аппроксимации.
3. Разработанный комплекс программ позволяет на основе применения универсального объектно-ориентированного подхода рассчитывать вероятностно-временные характеристики оперативности многоканальных немарковских систем массового обслуживания с «разогревом» произвольными аппроксимирующими распределениями фазового типа.

**Практическая значимость работы**

1. Разработанная методика оценивания вероятностно-временных характеристик оперативности функционирования узлов систем распределенной обработки данных позволяет повысить точность расчета названных характеристик на основе учета дополнительных временных затрат на актуализацию контекста.
2. Разработанный комплекс моделей доведен до инженерных методик и программ, готовых к непосредственному применению при оценивании

10

вероятностно-временных характеристик оперативности

функционирования узлов систем распределенной обработки данных. 3. Полученные результаты позволяют снизить трудозатраты при проектировании новых и модификации существующих узлов систем распределенной обработки данных.

**Достоверность и обоснованность результатов** диссертации обеспечивается строгими математическими выкладками, корректным применением математических методов исследования и подтверждается положительными результатами функционального тестирования, а также близостью характеристик оперативности, полученных в компьютерных экспериментах по моделированию, к характеристикам моделируемой системы, апробацией на научно-технических конференциях и публикациями в научных изданиях.

**Положения, выносимые на защиту:**

1. Математические модели оценивания оперативности обработки данных в узлах информационных систем на основе многоканальных немарковских систем массового обслуживания с «разогревом».
2. Численный метод расчета вероятностно-временных характеристик многоканальных немарковских систем массового обслуживания с «разогревом» на основе гиперэкспоненциальной аппроксимации.
3. Комплекс программ по расчету вероятностно-временных характеристик оперативности для разработанного класса моделей многоканальных немарковских систем массового обслуживания с «разогревом» и аппроксимирующими распределениями фазового типа.

**Внедрение результатов исследования.** Теоретические положения и практические рекомендации диссертации были внедрены и использованы:

11

· на предприятии ООО «Опенвэй Сервис» при оценке и обосновании
принятия проектных решений по расчету характеристик оперативности
распределенной автоматизированной системы управления финансами;

· на предприятии ООО «Ф-Лайн Софтвер» («First Line Software») при
оценке характеристик оперативности и оптимизации проектных
ресурсов при построении высоконагруженных распределенных веб-
приложений для транспортной компании;

· в ФГБОУ ВПО ПГУПС в учебном процессе кафедры «ИВС» при
проведении лекционных занятий и практических работ по дисциплине
«Корпоративные информационные системы».

**Практическое использование** результатов диссертационной работы подтверждено соответствующими актами о внедрении.

**Публикации.** Основные научные результаты по теме диссертации представлены в 7 публикациях, в число которых входят: 4 статьи в журналах, входящих в Перечень рецензируемых научных изданий, рекомендуемых ВАК; 1 свидетельство на программы для ЭВМ, 2 работы в материалах международных конференций.

**Апробация работы.** Основные научные положения и результаты диссертационной работы докладывались, обсуждались и были одобрены на Восемнадцатой международной конференции по информационным технологиям на железнодорожном транспорте «Инфотранс» (Санкт-Петербург, Бизнес-Диалог, 2013), на Четвертой и Третьей международных научно-практических конференциях «Интеллектуальные системы на транспорте – ИнтеллектТранс» (Санкт-Петербург, ФГБОУ ВПО ПГУПС, 2014, 2013), на Всероссийской научно-технической интернет-конференции с международным участием «Информационные и управляющие системы на транспорте и в промышленности» (Омск, ОмГУПС, 2014), на Семьдесят четвертой, Семьдесят третьей, Семьдесят второй всероссийских научно-технических конференциях студентов, аспирантов и молодых ученых

12

«Транспорт: проблемы, идеи, перспективы. Неделя науки» (Санкт-Петербург, ФГБОУ ВПО ПГУПС, 2014, 2013, 2012).

**Структура и объем работы.** Диссертация состоит из введения, 4 разделов, заключения, приложений и списка литературы, включающего 86 наименований, всего 116 страниц машинописного текста.

**Краткая характеристика работы.**

**Во введении** определена цель диссертационного исследования, описано содержание диссертации по разделам, обоснована проблемная ситуация и актуальность темы исследования.

**В первом разделе** перечислены объект и предмет исследования, требования, предъявляемые к оперативности распределенных ИС при разработке и эксплуатации, обзор и обоснование используемых показателей и критериев оперативности, анализ существующих моделей ИС для расчета показателей оперативности, обзор методов теории массового обслуживания для исследования показателей оперативности. Делается вывод о недостатках существующих подходов к моделированию функционирования распределенных ИС. В заключении раздела формулируется постановка задачи исследования.

**Во втором разделе** приведена классификация систем массового обслуживания и современных методов их анализа, показано влияние актуализации контекста на показатели оперативности системы, приведен математический аппарат аппроксимации произвольных вероятностных распределений с использованием распределений фазового типа, описаны разработанные модели многоканальных немарковских СМО «с разогревом» на основе гиперэкспоненциального распределения. Для каждой модели дано описание и указаны ее особенности, приведены диаграммы и матрицы переходов между состояниями, описана система линейных алгебраических уравнений баланса. Также приведено описание модификации общего итерационного метода Такахаши-Таками, которая позволяет для

13

распределенных ИС рассчитывать вероятностно-временные характеристики оперативности многоканальных немарковских систем массового обслуживания с «разогревом» на основе двухфазной гиперэкспоненциальной аппроксимации. Выполнены расчеты, которые иллюстрируют применение моделей и получаемые распределения временных характеристик оперативности распределенных ИС. В заключении раздела указано, что разработанные модели являются обобщением существующих подходов к моделированию оперативности ИС за счет использования общих предположений о характере немарковских вероятностных процессов в ИС.

**В третьем разделе** описана программная реализация моделей систем массового обслуживания с «разогревом» для расчета показателей оперативности. Кратко описаны функции программного комплекса. Приведен метод задания исходных данных и подход к тестированию разработанного программного комплекса, оценка вычислительной точности получаемых результатов. В заключении раздела указано, что разработанный комплекс программ позволяет на основе применения универсального объектно-ориентированного подхода рассчитывать вероятностно-временные характеристики оперативности многоканальных немарковских систем массового обслуживания с «разогревом» аппроксимирующими распределениями фазового типа.

**Четвертый раздел** посвящен вопросам практического применения результатов исследования и оценке получаемого эффекта. Приведено описание объектов внедрения результатов исследования. Указаны результаты практического использования полученных результатов на предприятиях ООО «Ф-Лайн Софтвер», ООО «Опенвэй Сервис» при оценке характеристик оперативности распределенных приложений.

Делается вывод об эффекте от использования результатов исследования для повышения точности оценки оперативности функционирования ИС. В частности, показано, что использование

14

результатов исследования повышает точность расчетов характеристик оперативности в среднем на 10%.

**В заключении** приведены основные результаты исследований.

**В приложение 1** включены исходные тексты основных функций комплекса программ расчета характеристик оперативности ИС.

**Заключение**

В рамках решения поставленной в работе задачи повышения точности

расчета характеристик оперативности распределенной обработки данных в узлах информационных систем с учетом дополнительных временных затрат на актуализацию контекста («разогрев») были получены следующие научные и практические результаты:

· разработан комплекс моделей многоканальных немарковских систем
массового обслуживания с «разогревом» на основе распределений
фазового типа, обеспечивающих возможность учета дополнительных
временных затрат при оценивании оперативности функционирования
узлов в информационных системах распределенной обработки данных;

· разработан численный метод расчета вероятностно-временных
характеристик многоканальных немарковских систем массового
обслуживания с «разогревом» на основе гиперэкспоненциального
распределения;

· разработан программный комплекс расчета характеристик
оперативности для указанного класса многоканальных моделей систем
массового обслуживания;

· разработана методика оценки вероятностно-временных характеристик
оперативности функционирования узлов систем распределенной
обработки данных;

· проведен анализ характеристик оперативности функционирования
узлов систем распределенной обработки данных на примере
распределенной автоматизированной системы управления финансами
предприятия ООО «Опенвэй Сервис»; выполнена оценка
эффективности решения.

Результаты исследований имеют следующий практический эффект, подтвержденный соответствующими актами о внедрении:

102

· повышение точности оценивания характеристик оперативности
функционирования ИС в среднем на 10% (за счет использования
разработанных более точных и обобщенных моделей);

· повышение точности планирования необходимых вычислительных
ресурсов ПС в пределах 10%.

Направления дальнейших исследований включают:

· добавление возможностей по работе с моделями, где
аппроксимирующие распределения фазового типа имеют
комплексно-сопряженные параметры,

· возможность интеграции со средами имитационного
моделирования,

· разработку аналитических методов расчета вероятностно-
временных характеристик оперативности для распределенных ИС.