

На правах рукописи



АНТОНОВА ПОЛИНА ВАЛЕРЬЕВНА

**МЕТОДЫ МОДЕЛИРОВАНИЯ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ
С ОГРАНИЧЕННЫМ ВРЕМЕНЕМ ПРЕБЫВАНИЯ ЗАЯВКИ В ОЧЕРЕДИ И
ВРЕМЕНЕМ ОБСЛУЖИВАНИЯ, РАСПРЕДЕЛЕННЫМ ПО ЗАКОНУ
ЭРЛАНГА**

1.2.2. Математическое моделирование,
численные методы и комплексы программ

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Казань – 2025

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования «Казанский национальный исследовательский технологический университет»

Научный руководитель: доктор технических наук, доцент
Титовцев Антон Сергеевич

Официальные оппоненты: **Моисеева Светлана Петровна**
доктор физико-математических наук, профессор,
федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Томский
государственный университет», заведующий
кафедрой теории вероятностей и математической
статистики;

Тарасов Вениамин Николаевич

доктор технических наук, профессор, федеральное
государственное бюджетное образовательное
учреждение высшего образования «Поволжский
государственный университет телекоммуникаций и
информатики», заведующий кафедрой
программного обеспечения и управления в
технических системах.

Ведущая организация: федеральное государственное бюджетное
образовательное учреждение высшего образования
«Казанский национальный исследовательский
технический университет им. А.Н. Туполева–КАИ»,
г. Казань.

Защита состоится 27 июня 2025 года в 14.00 на заседании диссертационного совета 24.2.312.08, созданного на базе ФГБОУ ВО «Казанский национальный исследовательский технологический университет» (420015, г. Казань, ул. К. Маркса, д. 68, зал заседаний Ученого совета, А – 330).

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВО «Казанский национальный исследовательский технологический университет» и на сайте <https://www.kstu.ru/servlet/contentblob?id=547557>

Автореферат разослан «__» _____ 2025 года.

Ученый секретарь
диссертационного совета 24.2.312.08,
доктор экономических наук, доцент



Кудрявцева
Светлана
Сергеевна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Задачи исследования систем массового обслуживания (СМО) с ограничениями на управляющие параметры приобретают особую актуальность. Задачи такого рода играют важную роль в различных предметных областях, таких как теория телетрафика, логистика, теория коммуникаций и др. Согласно указу Президента Российской Федерации № 145 «О стратегии научно-технологического развития Российской Федерации», в соответствии с п. 21 а, е, и п. 22 стратегии, развитие современных интеллектуальных телекоммуникационных систем относится к одному из приоритетных направлений научно-технологического развития страны.

В условиях наступившей четвертой промышленной революции, предполагающей новый подход к производству, основанный на массовом внедрении информационных технологий в промышленность, масштабной автоматизации бизнес-процессов и распространении искусственного интеллекта, важнейшими становятся задачи, связанные с повышением показателей эффективности различного рода технологических процессов и технических систем. Исследования в направлениях теории массового обслуживания с ограничениями на управляющие параметры не только улучшают существующие технологии и процессы, но и открывают новые возможности для будущих исследований, способствуя развитию как научных, так и практических дисциплин.

В простейших случаях объекты, функционирующие по принципу СМО, описываются наиболее известными классическими моделями. Однако для более глубокого изучения различных систем требуются сложные модели, учитывающие их различные технические особенности. К таким моделям можно отнести СМО, которые имеют те или иные ограничения на параметры системы.

Среди задач с ограничениями на управляющие параметры особое внимание уделяется тем, в которых ограничения касаются времени пребывания заявки в системе. Иными словами, некоторые заявки в таких системах являются «нетерпеливыми». Для подобных систем массового обслуживания можно выделить два основных подхода к постановке задачи. В первом случае заявка может покинуть систему как из очереди, так и с обслуживания. Во втором случае «нетерпеливая» заявка покидает систему, находясь в очереди. Модели такого типа остаются наименее изученными среди всех типов систем массового обслуживания. Исследование этих моделей имеет большое значение при оптимизации процессов в подобных системах, что открывает новые возможности для улучшения их эффективности.

Одна из основных задач изучения подобных систем заключается в следующем. Для расчетов в простейших моделях используются суммы либо бесконечного, либо конечного числа слагаемых.

Для того, чтобы найти решения для таких задач целесообразно использование численных методов. В этом случае все численные показатели системы массового обслуживания определяются по отдельности путем суммирования нескольких первых членов соответствующего ряда, который может быть как конечный, так и бесконечный. Однако данный подход не позволяет получить замкнутое аналитическое решение. Несмотря на это подход с использованием приближенных численных методов позволяет оценить основные характеристики системы, что может быть достаточным для ряда прикладных задач. Детальное же изучение процессов в таких системах, как правило, остается затруднительным.

В данной работе впервые представлена одноканальная система массового обслуживания смешанного типа с временем обслуживания, представляющим собой сумму независимых экспоненциально распределенных случайных величин, с накопителем неограниченной емкости и решается задача комплексного исследования режимов функционирования СМО с эрланговски распределенным временем обслуживания, что подразумевает под собой разработку моделей, методов и комплекса программ, направленных на определение характеристик системы, обеспечивающих требуемое качество обслуживания. Подобного рода системы могут быть рассмотрены при разработке систем сопровождения, администрирования, и эксплуатации программно-технических средств и информационных ресурсов, а также для мониторинга и оптимизации производственных процессов, а значит, задача изучения СМО $M/E_r/1$ с ограничениями на время пребывания требования в очереди является актуальной.

Степень разработанности темы.

Теория массового обслуживания берет свои истоки в исследованиях, связанных с телефонными системами. Пионерами в этой области были такие ученые, как У. Феллер, М.Д. Кендалл и др. Феллер изучал вопросы, касающиеся концепции очередей. Кендалл ввел понятие вложенных цепей Маркова и разработал широко используемую классификацию СМО. Кларк первым нашел решение для уравнений процесса гибели и размножения с постоянными коэффициентами для переходного состояния.

Дальнейшие исследования были посвящены развитию методов решения задач теории массового обслуживания. С. Карлин и Мак-Грегор предложили метод решения с использованием коэффициентов, зависящих от числа требований в очереди. Т.Л. Саати разработал решение для переходного состояния марковской модели многоканальной СМО с использованием преобразования Лапласа. Эти ранние работы заложили основы в дальнейших исследованиях теории массового обслуживания (ТМО), которая нашла применение в различных предметных областях.

Вопросами, возникающими при изучении СМО в настоящее время в нашей стране, занимается ряд ученых: Т.И. Алиев, Ю.В. Гайдамака, А.Н. Дудин, В.Н. Задорожный, А.А. Назаров, А.Н. Моисеев, С.П. Моисеева, С.В. Пауль, А.В. Печинкин, В.А. Романенко, В.В. Рыков, К.Е. Самуйлов, С.П. Сущенко, В.Н. Тарасов, О.М. Тихоненко, М.П. Фархадов, М.А. Шнепс-Шнеппе и др. В большинстве случаев эти проблемы имеют частный характер и ограничиваются анализом первых моментов различных характеристик систем массового обслуживания.

В современных условиях в теории массового обслуживания отсутствие замкнутого аналитического решения является проблемой, которая рассматривается как прикладная область исследований. В области таких видов моделирования как имитационное и аналитическое существует немалое количество публикаций. Значительный вклад в развитие теории внесли такие ученые как А.А. Боровков, Б.В. Гнеденко, Д. Кендалл, Д. Кокс, Д. Литтл, Ф. Поллячек, В. Смит, Л. Такач, А.Я. Хинчин. Однако на практике аналитические расчеты до недавнего времени проводились с использованием формул, которые были предложены А. К. Эрлангом в начале прошлого века.

Классический метод Эрланга применим к системам с простейшим входящим потоком и показательным распределенным временем обслуживания. Однако если первое из названных допущений на основании ряда теоретических соображений можно признать реалистичным, то второе в большинстве приложений диктуется лишь

соображениями «считаемости». Расчёт систем с эрланговски распределенным временем обслуживания и простейшим входным потоком (согласно классификации Кендалла $M/E_r/1$) практически возможен только после «фазовой» аппроксимации распределения обслуживания с экспоненциально распределённой задержкой в каждой.

В системах массового обслуживания, где время обслуживания представляет собой сумму нескольких независимых случайных величин, распределенных по одному и тому же экспоненциальному закону, число заявок в системе в данный момент времени уже не будет представлять собой марковский процесс. Однако для таких систем существует возможность построить марковский процесс с непрерывным временем, применяя метод фиктивных фаз. Таким образом несмотря на то, что число заявок не является марковским процессом напрямую, можно использовать специальные методы для моделирования такой системы как марковского процесса с непрерывным временем.

Для эрланговского распределения времени обслуживания предполагается, что каждая заявка, прибывшая в систему, должна пройти несколько фаз обслуживания, длительность которых распределена по экспоненциальному закону. Аналогичный подход используется и в случае, когда времена обслуживания можно описать допускающими аналогичную вероятностную трактовку распределениями фазового типа.

Системы $M/E_r/1$ находят широкое применение в задачах организации трафика телекоммуникационных сетей и в задачах промышленной инженерии.

Используя матрично-геометрический подход, М. Ньютс исследовал системы массового обслуживания с эрланговским распределением времени обслуживания. Следует отметить, что этот метод является чрезвычайно трудным для практической реализации и зачастую страдает от вычислительной нестабильности из-за плохой обусловленности используемых матриц большой размерности.

Первая попытка применения нового метода изучения СМО с ограничениями на время нахождения заявки в очереди или в системе в целом в случае, когда все заявки можно назвать «нетерпеливыми» была предпринята в работах А.П. Кирпичникова. В основе данного подхода лежало применение функции Г. Миттаг-Леффлера 1-го порядка.

С точки зрения практического применения, более перспективным представляется расширенный вариант задачи, в котором каждая заявка ожидает не дольше случайного времени, распределенного по экспоненциальному закону. До сих пор не проводились исследования СМО смешанного типа с эрланговски распределенным временем обслуживания.

Объектом исследования являются технические системы с очередями, функционирующие в информационных системах, финансовых учреждениях, службах преодоления чрезвычайных ситуаций, системах мониторинга качества и пр. Проблемы, возникающие в данных областях, структурируются вокруг очередей, поэтому формальное описание современных систем логично формулировать в терминах теории массового обслуживания.

Предметом исследования являются вопросы изучения одноканальных СМО смешанного типа с временем обслуживания, распределенным по закону Эрланга.

Соответствие исследования паспорту научной специальности. Область диссертационного исследования соответствует следующим пунктам паспорта научной специальности 1.2.2. Математическое моделирование, численные методы и комплексы

программ (технические науки): п.6. Разработка систем компьютерного и имитационного моделирования, алгоритмов и методов имитационного моделирования на основе анализа математических моделей (технические науки); п.8. Комплексные исследования научных и технических проблем с применением современной технологии математического моделирования и вычислительного эксперимента; п.9. Постановка и проведение численных экспериментов, статистический анализ их результатов, в том числе с применением современных компьютерных технологий.

Цель и задачи исследования.

Целью исследования является разработка комплекса инструментов для анализа характеристик СМО с ограничениями на время пребывания заявки в очереди, где время обслуживания подчиняется закону Эрланга.

Задачи исследования:

1. Разработать математическую модель СМО смешанного типа с ограниченным временем ожидания заявки в очереди и временем обслуживания, распределенным по закону Эрланга, получить формулы для стационарных и нестационарных характеристик модели.

2. Разработать метод исследования СМО с эрланговски распределенным временем обслуживания, позволяющий найти требуемое количество фаз, обеспечивающее заданный уровень качества обслуживания.

3. Разработать имитационную модель СМО смешанного типа, а также метамодель, аппроксимирующую ключевые характеристики СМО на основе методов машинного обучения, для оценки показателей эффективности системы по заданным входным параметрам.

4. Разработать комплекс программных средств для решения задач по расчету ключевых характеристик СМО и оценки её эффективности.

5. Провести апробацию полученных результатов.

Научная новизна представленных результатов характеризуется развитием методов математического моделирования СМО, качественных и приближенных методов исследования СМО, методов и алгоритмов имитационного моделирования, и заключается в следующем:

1. *Разработана новая математическая модель* СМО смешанного типа, отличающаяся от известных моделей сочетанием свойств СМО с ограниченным временем ожидания заявки в очереди и СМО с временем обслуживания, распределенным по закону Эрланга. Строгими аналитическими методами *получены* общие формулы для характеристик модели в стационарном и нестационарном режимах, *позволяющие* задавать временные рамки нахождения заявок в очереди и количество фаз обслуживания (п. 8 паспорта специальности 1.2.2).

2. *Предложен новый метод* исследования СМО с эрланговски распределенным временем обслуживания, *отличающийся* возможностью нахождения требуемого количества фаз, обеспечивающего заданный уровень качества обслуживания, которое может быть оценено с помощью коэффициента вариации различных характеристик СМО. *Выявлено*, что граничные значения числа фаз обслуживания логарифмически связаны с изменением приведенной интенсивности входного потока заявок (п. 8 паспорта специальности 1.2.2).

3. *Разработана имитационная модель* СМО смешанного типа с использованием инструментальной среды AnyLogic PLE, а также *метамодель*, отличающаяся возможностью аппроксимации ключевых характеристик СМО на основе методов

машинного обучения. *Метамодель* позволяет оперативно оценивать показатели эффективности системы по заданным входным параметрам без проведения ресурсоёмких имитационных экспериментов, что значительно ускоряет процесс анализа (п. 6 паспорта специальности 1.2.2).

4. *Разработан комплекс специализированных программных средств, отличающийся* возможностью расчета ключевых характеристик СМО смешанного типа и оценки её эффективности (п. 9 паспорта специальности 1.2.2).

Теоретическая значимость диссертационной работы состоит в следующем:

– предложена новая разновидность моделей СМО с эрланговски распределенным временем обслуживания и ограниченным ожиданием в очереди, которые могут быть применены в различных областях;

– решена задача формализации стационарных и нестационарных характеристик СМО $M/E_r/1$ смешанного типа;

– разработан новый метод исследования СМО, основанный на вычислении предельных значений количества фаз обслуживания, при которых система поддерживает стабильный режим работы;

– впервые к исследованию СМО применены ансамблевые методы машинного обучения.

Практическая значимость диссертационной работы заключается в том, что полученные результаты могут найти применение в различных областях, таких как транспортные и телекоммуникационные системы, производство, логистика, сфера обслуживания и других, работающих по принципу СМО смешанного типа, что окажет влияние на эффективность функционирования данных объектов.

Методология и методы исследования включают применение теории вероятностей, теории функций комплексного переменного и теории случайных процессов и цепей Маркова. Для решения трансцендентных уравнений применялись соответствующие численные методы. Дополнительно была построена имитационная модель СМО смешанного типа с эрланговским распределением времени обслуживания в среде моделирования AnyLogic PLE. Для разработки метамодели СМО применены ансамблевые методы машинного обучения. Разработка и реализация программного комплекса осуществлялись с использованием Wolfram Mathematica.

Положения, выносимые на защиту:

1. Математическая модель СМО смешанного типа, сочетающая свойства СМО с ограниченным временем ожидания заявки в очереди и СМО с временем обслуживания, распределенным по закону Эрланга.

2. Новый метод исследования СМО с эрланговски распределенным временем обслуживания, полученные численные результаты: основной из них определяет логарифмический характер поведения граничных значений числа фаз обслуживания, соответствующих стабильному режиму по длине очереди с изменением приведенной интенсивности входного потока требований.

3. Имитационная модель СМО смешанного типа, а также метамодель, аппроксимирующая ключевые характеристики СМО на основе методов машинного обучения.

4. Комплекс специализированных программных средств для решения задач по расчету ключевых характеристик СМО и оценки её эффективности.

Степень достоверности результатов диссертации обеспечивается корректной постановкой задач, строгим выполнением математических выкладок и проверкой

полученных решений на соответствие известным частным случаям, описанным другими авторами.

Апробация. Разработанные методы и алгоритмы внедрены в деятельность следующих организаций, что подтверждается соответствующими справками:

1. ООО «Систематика Консалтинг», г. Москва при разработке системы технического сопровождения, администрирования, сервисной поддержки и эксплуатации программно-технических средств и информационных ресурсов Федеральной государственной информационной системы территориального планирования (ФГИС ТП).

2. ООО «НПО ЛАБС», г. Санкт-Петербург при разработке программно-аппаратного решения для мониторинга и оптимизации производственных процессов.

Полученные результаты исследований активно используются в учебном процессе ФГБОУ ВО «КНИТУ» для обучения студентов по направлениям подготовки 01.03.02 «Прикладная математика и информатика» и 02.03.03 «Математическое обеспечение и администрирование информационных систем»

Результаты диссертации докладывались на международной научной конференции «FarEastCon-2020» (г. Владивосток, 2020), IV международной школе-конференции «DCNAIR-2020» (г. Иннополис, 2020), 34-й международной научной конференции «Математические методы в технике и технологиях ММТТ-34» (г. Санкт-Петербург, 2021), 37-й международной научной конференции «Математические методы в технике и технологиях ММТТ-37» (г. Казань, 2024), международной научной конференции «Перспективные фундаментальные исследования и научные методы» (г. Санкт-Петербург, 2024), международной научной конференции «Перспективные исследования в технических и естественных науках» (г. Санкт-Петербург, 2024).

Личный вклад автора. Все представленные результаты получены автором лично. В зарегистрированном программном обеспечении автору принадлежит ведущая роль в разработке основных модулей для расчетов необходимых характеристик.

Публикации. По теме диссертации опубликовано самостоятельно и в соавторстве 10 печатных работ: 5 работ в рецензируемых научных изданиях, рекомендованных ВАК РФ; 1 – в изданиях, индексируемых в Scopus, 4 – в прочих изданиях. Зарегистрировано 2 электронных ресурса.

Структура и объём работы. Диссертация состоит из введения, четырех глав, заключения, списка литературы и приложения. Диссертация напечатана в 1,5 межстрочных интервала, полный объём составляет 126 страниц, включая 46 рисунков. Библиографический список включает 95 литературных источников.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** показана актуальность темы диссертационной работы, приведены её цель, задачи, научная новизна, теоретическая и практическая значимость, приводится краткий обзор глав.

В главе 1 приводится обзор и анализ литературных источников, результатов научных исследований, ранее выполненных по теме диссертации, дается характеристика предметной области и приводится постановка задачи.

В п. 1.1 приводится обзор исследований и теоретических результатов, посвящённых системам массового обслуживания. Анализируются основные подходы к изучению временных характеристик обслуживания, включая методы построения марковского процесса с «фиктивными фазами» и рассмотрение «нетерпеливых» заявок

(с ограничением на время ожидания). Показано, что ряд классических результатов по системам $M/M/1$ или $M/G/1$ не всегда корректно описывает ситуацию с многофазным (эрланговским) распределением времени обслуживания, характерным для более сложных сервисных процессов. Кроме того, обсуждаются отечественные и зарубежные работы, где представлены методы расчёта стационарных и нестационарных характеристик, а также подходы к моделированию систем, функционирующих в условиях высокой нагрузки или подверженных вариациям входящего потока.

В п. 1.2 рассматривается многоуровневая система технической поддержки ФГИС ТП, структура которой представлена в виде последовательности стадий обработки входящих вызовов. На первом этапе абонент совершает звонок на линию технической поддержки; при наличии свободных ресурсов вызов незамедлительно принимается оператором первого уровня поддержки, в противном случае запрос попадает в очередь ожидания. Если обращение не может быть решено на первом уровне (например, требует более глубокой компетенции или дополнительной проверки), оно эскалируется на следующий уровень поддержки. Аналогично, когда задачи выходят за рамки компетенций второго уровня, вызовы передаются на третью линию.

Архитектура системы поддержки включает:

Входной поток заявок (звонков), формируемый абонентами, обращающимися на линию. Этот поток может рассматриваться как простейший (пуассоновский) при больших объёмах обращений и отсутствии коррелированных поступлений.

Очередь ожидания, возникающая, если все операторы первого уровня поддержки заняты. Время пребывания в очереди ограничено (часть пользователей может прервать ожидание, не дождавшись ответа).

Первая линия приёма обращений — основной контакт-центр, решающий вопросы общего характера. При невозможности решения проблемы оператором первого уровня происходит эскалация вызова.

Вторая и третья линии технической поддержки — уровни, на которых работают специалисты узких направлений, способные решать более сложные и редкие проблемы. Процесс передачи заявок на эти линии также сопровождается ожиданием, однако при меньшем общем объёме обращений, чем на первом уровне.

С точки зрения теории массового обслуживания, предложенную схему можно описывать как систему с многофазным обслуживанием, где каждый последующий уровень поддержки выступает в роли отдельной «фазы». В диссертационном исследовании формулируется задача определения основных вероятностных и временных характеристик данной системы.

Таким образом, многоуровневая система технической поддержки рассматривается не как изолированный узел, а как часть инфраструктуры ФГИС ТП, обеспечивающая эффективное взаимодействие пользователей с сервисом и требующая анализа производительности для поддержания заданных качественных показателей обслуживания.

В п. 1.3 приводится математическая постановка задачи.

В качестве исходной математической модели рассматривается одноканальная система массового обслуживания вида $M/E_r/1$ с неограниченным объёмом накопителя. При поступлении заявок с интенсивностью λ время обслуживания каждой заявки распределено по закону Эрланга с параметрами r и μ .

Таким образом, заявка проходит r последовательных фаз, каждая из которых имеет экспоненциальное распределение с параметром μ . В результате суммарное время обслуживания удовлетворяет эрланговскому распределению.

Одним из методов работы с системами такого вида является построение марковского процесса, описывающий их функционирование, с использованием «фиктивных фаз». Пусть $\nu(t)$ - число заявок в системе в момент времени t . В случае эрланговского времени обслуживания можно считать, что каждая заявка проходит r фаз обслуживания, каждая из которых имеет экспоненциальное распределение с параметром μ . Пусть $\xi(t)$ - число фаз, оставшихся для обслуживания заявки в момент времени t . Случайный процесс, характеризующий число заявок в СМО в момент времени t уже не будет для данной СМО марковским. Определим процесс $\{\eta(t), t \geq 0\}$ следующим образом: если в системе в момент времени t нет заявок, то $\eta(t) = \nu(t)$. В данном случае марковское свойство обеспечивается пуассоновским характером входного потока в виду того, что обслуживания в данном случае не происходит. В противном случае, если в системе происходит обслуживание, то дополнительно указывается число оставшихся фаз в обслуживании, то есть $\eta(t) = (\nu(t), \xi(t))$. В силу экспоненциальности фаз обслуживания и пуассоновского характера входного потока заявок процесс $\eta(t)$ обладает марковским свойством.

Рассмотрение этой «классической» модели (без преждевременного ухода заявок) важно по двум причинам:

- Теоретическая база. Модель $M/E_r/1$ является базовой для исследования более сложных систем смешанного типа. На её основе выводятся фундаментальные уравнения Колмогорова и разрабатываются методы вычисления стационарных (а также нестационарных) характеристик: среднего числа заявок в системе, среднего времени ожидания и обслуживания, вероятности занятости и т.д.

- Практическая интерпретация. Для ряда сервисных процессов многофазное (эрланговское) распределение времени обслуживания описывает многошаговые процедуры (эскалация в рамках одной заявки, последовательность проверок, документооборота и т.п.).

Тем не менее, данная классическая схема предполагает, что все поступающие заявки остаются в системе до завершения обслуживания. В реальности, и особенно для служб технической поддержки, часть пользователей не готова ожидать сколь угодно долго: по прошествии некоторого (часто случайного) времени заявка покидает очередь.

В реальных условиях, включая многоуровневые контакт-центры, значительное число обращений «вылетает» из очереди, если время ожидания превысило допустимую для пользователя величину. Подобный уход заявок («нетерпеливое» поведение) приводит к следующим особенностям:

- нагрузочная характеристика системы меняется: часть входящего потока «самопроизвольно» отсекается из очереди, разгружая систему, но при этом ухудшает качество обслуживания (возникают «потерянные» звонки);

- оценка эффективности без учёта нетерпеливых заявок может давать нереалистично благоприятные результаты (заниженное среднее время ожидания и загруженности сервера) и не отражать реальных показателей;

- неоднородность потока: время «терпения» пользователей (то есть период, по истечении которого заявка прерывается) может иметь собственное вероятностное распределение, что усложняет расчёты и требует расширения классической модели.

Поэтому в дальнейшей постановке добавляется механизм ограничения на время пребывания заявки в очереди, моделирующий «вылет» (или досрочный уход) части заявок. Таким образом, модель расширяется до учёта нетерпеливых заявок, а для её анализа используется расширенная система уравнений Колмогорова. Подобная модификация позволяет точнее отразить специфику работы систем, где часть пользователей не готова ожидать ответа оператора сверх определённого времени.

В главе 2 содержатся аналитические результаты, полученные в процессе формализации основных стационарных и нестационарных характеристик СМО с ограниченным временем ожидания в очереди и временем обслуживания, распределенным по закону Эрланга. Получены аналитические выражения, описывающие основные числовые и вероятностные характеристики модели $M/E_r/1$. С помощью преобразования Лапласа и дальнейшего его обращения аналитически получены значения производящей функции, которая позволяет вычислить основные нестационарные характеристики для модели $M/E_r/1$.

В п. 2.1 вводится модель, позволяющая описать систему, где заявки могут прерывать ожидание и/или обслуживание. Эрланговское распределение времени обслуживания обеспечивает более гибкое и точное описание процессов обслуживания в реальных системах, которые часто состоят из нескольких этапов (фаз). Учет «нетерпеливых» заявок даёт возможность прогнозировать реальные показатели качества обслуживания (время ожидания, вероятность завершения обслуживания и т.д.) и принимать решения об оптимизации параметров системы. Приводится математическая формализация стационарных характеристик СМО с ограниченным временем ожидания в очереди и временем обслуживания, распределенным по закону Эрланга.

Как и ранее, пусть $v(t)$ - число заявок в системе в момент времени t . В случае эрланговского времени обслуживания будем считать, что каждая заявка проходит r фаз обслуживания, каждая из которых имеет экспоненциальное распределение с параметром μ , где $\xi(t)$ - число фаз, оставшихся для обслуживания заявки в момент времени t . Определим процесс $\{\eta(t), t \geq 0\}$ следующим образом: если в системе в момент времени t нет заявок, то $\eta(t) = v(t)$. В противном случае, если в системе происходит обслуживание, то дополнительно указывается число оставшихся фаз в обслуживании, то есть $\eta(t) = (v(t), \xi(t))$. Каждая прибывшая в систему заявка будет ожидать в очереди не более случайного времени, распределенного по экспоненциальному закону с параметром θ . Для описания вероятностных характеристик процесса $\eta(t) = (v(t), \xi(t))$ введём следующие обозначения:

$$p_{ij}(t) = P\{\eta(t) = (i, j)\},$$

где i – число заявок в системе в момент времени t , а j – число фаз, оставшихся в обслуживании. Здесь и далее $p_{ij}(t)$ -это вероятность того, что в момент времени t :

- в системе ровно i заявок (учитывая обслуживаемую, если она есть),
- заявке, которая сейчас обслуживается, осталось пройти ровно j фаз обслуживания.

$p_{ij}(t, t + \Delta t)$ будет обозначать вероятность перехода системы в состояние (i, j) за бесконечно малый промежуток времени $t + \Delta t$. Если рассматривать переходы в процессе $\eta(t) = (v(t), \xi(t))$, то они определяют, с какой вероятностью система за

бесконечно малый промежуток времени $t + \Delta t$ перейдет из текущего состояния в состояние (i,j) . Это сводится к двум основным типам событий:

- прибытие новой заявки, которое увеличивает общее число заявок в системе на единицу;
- завершение одной из фаз обслуживания, в результате чего уменьшается число оставшихся фаз обслуживания у текущей заявки на единицу (при достижении нуля заявка покидает систему).

При переходе к пределу $\Delta t \rightarrow 0$ вероятность возникновения более чем одного события одновременно (например, одномоментное прибытие нескольких заявок) является пренебрежимо малой и поэтому не учитывается в уравнениях Колмогорова. Эти вероятности используются при формулировании уравнений Колмогорова, отражающих динамику переходов между состояниями системы. При переходе к пределу $\Delta t \rightarrow 0$ получается система дифференциальных уравнений (уравнений Колмогорова), описывающая эволюцию вероятностей $p_{ij}(t)$ во времени.

Согласно графу (рис.1) имеем:

$$\begin{aligned}
 \frac{dp_0(t)}{dt} &= -\lambda p_0(t) + \mu p_{11}(t), \\
 \frac{dp_{1r}(t)}{dt} &= \lambda p_0(t) + \theta p_{2r}(t) - (\lambda + \mu)p_{1r}(t) + \mu p_{21}(t) \\
 \frac{dp_{ir}(t)}{dt} &= -(\lambda + \mu)p_{ir}(t) + \lambda p_{i-1,r}(t) + \mu p_{i+1,1}(t) + i\theta p_{i+1,r}(t) - \\
 &- (i-1)\theta p_{ir}(t), i > 1 \\
 \frac{dp_{ij}(t)}{dt} &= -(\lambda + \mu)p_{ij}(t) + \lambda p_{i-1,j}(t) + \mu p_{i,j+1}(t) + i\theta p_{i+1,j}(t) - \\
 &- (i-1)\theta p_{ij}(t), i > 1, 1 < j < r-1, \\
 \frac{dp_{1j}(t)}{dt} &= -(\lambda + \mu)p_{1j}(t) + \mu p_{1,j+1}(t) + \theta p_{2j}(t), 1 < j < r-1.
 \end{aligned}
 \tag{1}$$

Первым шагом в решении задачи является составление уравнений баланса для системы. В стационарном режиме "приток" в любое состояние системы должен равняться "оттоку" из него. Иными словами, для каждой возможной численности заявок в системе устанавливается равенство между суммарной вероятностью перехода в данное состояние и суммарной вероятностью перехода из него. На основании этой системы линейных уравнений можно получить рекуррентные соотношения для стационарных вероятностей состояний.

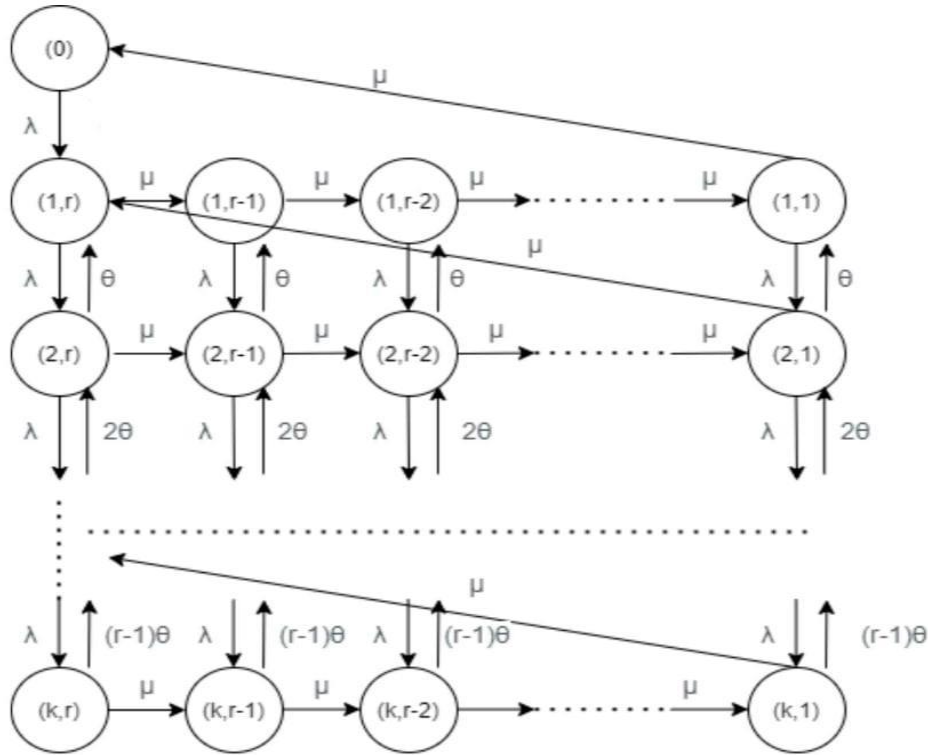


Рис. 1. Граф состояний и переходов СМО (разработан автором)

В стационарном режиме система уравнений баланса для одноканальной модели со значением числа фаз обслуживания и возможностью «вылета» заявок преобразуется в функциональное уравнение для производящей функции стационарного распределения. Производящая функция $P(z)$ задаётся суммой значений вероятностей состояний, умноженных на соответствующую степень z . Подстановка уравнений баланса в выражение для $P(z)$ даёт возможность свести бесконечную систему к одному функциональному уравнению, в котором интенсивности поступления и обслуживания (а также параметры ухода заявок) отражаются в виде коэффициентов. После определения конкретного решения с учётом условия нормировки ($P(1)=1$) производится извлечение искомым вероятностей из степенного разложения функции.

Для дальнейшего анализа используются тауберовы теоремы, устанавливающие связь между асимптотикой коэффициентов данного степенного ряда и аналитическими особенностями производящей функции. Для анализа стационарного распределения, а также с учётом выявленных аналитических свойств (через тауберовы теоремы), определяется конкретная функциональная форма решения и, в соответствии с условиями нормировки, приводится итоговое выражение для стационарных вероятностей. На этой основе формулируются получаемые в явном виде соотношения (формулы) для стационарных характеристик системы:

вероятность простоя СМО

$$p_0 = \frac{2(-1+\rho)^2(\mu-\lambda r-r\theta)}{2\mu-4\mu\rho+2\mu\rho^2-2r\theta+2r\rho\theta-\rho^2\theta-r\rho^2\theta}; \quad (2)$$

стационарные вероятности возможных состояний СМО

$$p_i = \sum_{j=1}^r p_{ij} = \frac{P^{(i)}(0)}{i!}; \quad (3)$$

вероятность обслуживания заявки

$$P_S = \frac{1}{\rho}(1-p_0) = \frac{1}{\rho}\left(1 - \frac{2(-1+\rho)^2(\mu-\lambda r-r\theta)}{2\mu-4\mu\rho+2\mu\rho^2-2r\theta+2r\rho\theta-\rho^2\theta-r\rho^2\theta}\right); \quad (4)$$

Умножение P_S на 100 % даёт долю (процент) заявок, которые система успешно обслуживает, а величина $1 - P_S$ (также умноженная на 100 %) определяет долю заявок, оставшихся без обслуживания.

Средняя длина очереди

$$Q = -1 + \frac{2(-1+\frac{\lambda r}{\theta})^2(\mu-\lambda r-r\theta)}{2\mu+6\lambda r^2+\frac{2\lambda^2\mu r^2}{\theta^2}-\frac{4\lambda\mu r}{\theta}+\frac{\lambda r^2}{\theta}-\frac{3\lambda^2 r^3}{\theta}-2r\theta} - \frac{(-2\mu+\lambda(-1+r))r(-1+\frac{\lambda r}{\theta})^2(\mu-\lambda r-r\theta)}{(\mu-\lambda r)^2(2\mu+6\lambda r^2+\frac{2\lambda^2\mu r^2}{\theta^2}-\frac{4\lambda\mu r}{\theta}+\frac{\lambda r^2}{\theta}-\frac{3\lambda^2 r^3}{\theta}-2r\theta)}; \quad (5)$$

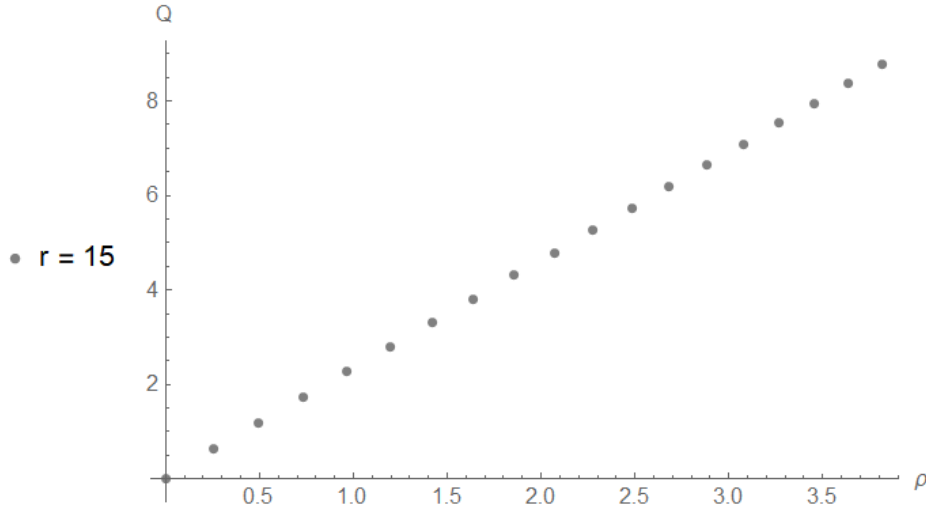


Рис.2. Зависимость средней длины очереди от приведенной интенсивности при значении числа фаз обслуживания $r=15$ (получена автором)

дисперсия числа заявок в очереди

$$\begin{aligned} \sigma_Q^2 &= \sum_k (k-1)^2 p_k - Q^2 = \sum_k k^2 p_k - 2 \sum_k k p_k + \sum_k p_k \\ &= P'(1) + P''(1) + 2P'(1) + (1-p_0) - Q^2 = \\ &= 1 - \frac{r^2(\mu\hat{\rho}(-2+(-1+r)\hat{\rho})+\lambda(2+r(-2+\hat{\rho})\hat{\rho}+r^2\hat{\rho}^2))^2}{(-2\mu(-1+\rho)^2+r(2+r(-2+\hat{\rho})\hat{\rho}+\rho^2)\theta)^2} - \\ &= \frac{\lambda^2 r(1+r)(6\mu^2-4\lambda\mu(-1+r)+\lambda^2(-1+r)r)(-1+\rho)^2(-\mu+r(\lambda+\theta))}{3\mu(\mu-\lambda r)^3(2\mu(-1+\rho)^2-r(2+r(-2+\hat{\rho})\hat{\rho}+\rho^2)\theta)} - \frac{2\rho(-2+(-1+r)\hat{\rho})(-\mu+r(\lambda+\theta))}{-2\mu(-1+\rho)^2+r(2+r(-2+\hat{\rho})\hat{\rho}+\rho^2)\theta} \\ &= \frac{2(-1+\rho)^2(-\mu+r(\lambda+\theta))}{-2\mu(-1+\rho)^2+r(2+r(-2+\hat{\rho})\hat{\rho}+\rho^2)\theta}, \quad (6) \end{aligned}$$

где $\hat{\rho} = \frac{\lambda}{\mu}$;

дисперсия числа заявок в системе

$$\begin{aligned} \sigma_N^2 &= \sum_k k^2 p_k - N^2 = P'(1) + P''(1) - N^2 = \frac{\rho^2(-2+(-1+r)\hat{\rho})^2(\mu-\lambda r-r\theta)^2}{(2\mu-4\mu\rho+2\mu\rho^2-2r\theta+2r\rho\theta-\rho^2\theta-r\rho^2\theta)^2} + \\ &+ \frac{\lambda^2 r(1+r)(6\mu^2-4\lambda\mu(-1+r)+\lambda^2(-1+r)r)(-1+\rho)^2(\mu-\lambda r-r\theta)}{3\mu(\mu-\lambda r)^3(2\mu-4\mu\rho+2\mu\rho^2-2r\theta+2r\rho\theta-\rho^2\theta-r\rho^2\theta)}; \quad (7) \end{aligned}$$

стационарное среднее время пребывания заявки в системе

$$v = -\varphi'(0) = \frac{(-2\mu+\lambda(-1+r))r(-\mu+r(\lambda+\theta))}{2\mu^3-\lambda^2 r^2(1+r)\theta+2\lambda\mu r^2(\lambda+\theta)-2\mu^2 r(2\lambda+\theta)} = \frac{N}{\lambda}. \quad (8)$$

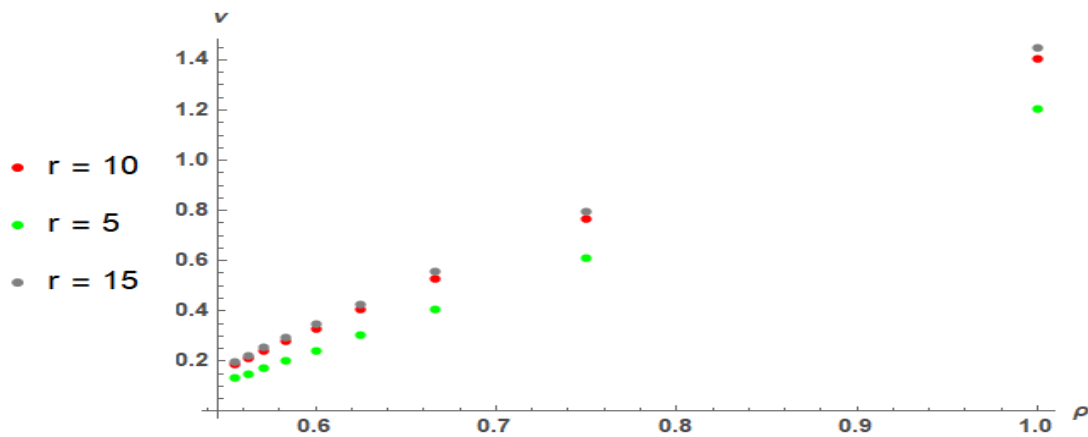


Рис.3. Зависимости стационарного среднего времени пребывания заявки в системе от приведенной интенсивности входного потока при различных значениях числа фаз обслуживания (получены автором)

В п. 2.2 приводится математическая формализация нестационарных характеристик СМО с ограниченным временем ожидания в очереди и временем обслуживания, распределенным по закону Эрланга.

В рамках проведённого исследования рассмотрены нестационарные характеристики системы массового обслуживания с эрланговским распределением времени обслуживания и возможностью ухода заявок. Для анализа временной динамики вероятностных состояний сформирована система дифференциальных уравнений Колмогорова, описывающая эволюцию системы во времени. Поскольку число состояний может быть бесконечным, использован метод производящих функций, позволяющий свести систему уравнений к более компактному виду.

Для решения полученных дифференциальных уравнений применено интегральное преобразование Лапласа, которое позволило упростить выражения, переведя дифференциальные уравнения в алгебраическую форму. В процессе обращения полученных выражений использованы теоремы разложения, в частности теорема Бореля, а также методы свёртки и интегралы Дюамеля, что позволило восстановить временные вероятностные характеристики в исходной области.

Анализ асимптотического поведения показал, что при выполнении определённых условий на интенсивности поступления заявок, обслуживания и ухода система сходится к стационарному распределению, соответствующему аналогичной системе без ограничений по времени ожидания. Определены параметры устойчивости, обеспечивающие существование установившегося режима.

На основе полученных выражений вычислены основные нестационарные характеристики:

вероятность простоя СМО

$$p_0(t) = 1 + \frac{2\theta\mu(\mu-\lambda r)t}{(-1 + e^{2\mu(\mu-\lambda r) + \theta(\mu-5\mu r + \lambda r(1+r))})\lambda r} \mu \quad (9)$$

После однозначного определения $p_0(t)$ с помощью производящей функции выводятся остальные нестационарные характеристики модели:

средняя длина очереди

$$Q(t) = \frac{p_0(t)\lambda}{\theta} + \frac{\mu - 2\mu r}{r(-\mu + \lambda r)} \quad (10)$$

среднее число заявок в системе

$$N(t) = \frac{p_0(t)\lambda}{\theta} + \frac{p_0(t)\lambda^2 r^3 - p_0(t)\lambda\mu r^2 + \mu^2(-1+2r)}{\mu r(\mu - \lambda r)}. \quad (11)$$

В п. 2.3 разработана имитационная модель открытой одноканальной СМО смешанного типа, где входящий поток заявок является пуассоновским, а время обслуживания распределено по эрланговскому закону.

В п. 2.4 разработана метамодель СМО с использованием ансамблевых методов машинного обучения, позволяющая по входным параметрам системы (таким как интенсивность входного потока заявок, число фаз обслуживания, ограничения на время ожидания) оперативно оценивать её производственные показатели (среднее время ожидания, длину очереди и т. д.). Ансамблевые методы машинного обучения комбинируют несколько базовых моделей для улучшения общей производительности и устойчивости к переобучению. В данном разделе рассматриваются три подхода: градиентный бустинг, AdaBoost и стэкинг-регрессор.

AdaBoost (Adaptive Boosting) – это ансамблевый метод, который комбинирует слабые модели (например, деревья решений) для создания сильной модели. В данном случае AdaBoost применяется для предсказания среднего времени ожидания (W).

StackingRegressor представляет собой ансамблевый метод, который объединяет несколько базовых моделей для создания более мощной финальной модели. В качестве финальной модели часто используется простая модель, такая как линейная регрессия, которая комбинирует предсказания базовых моделей.

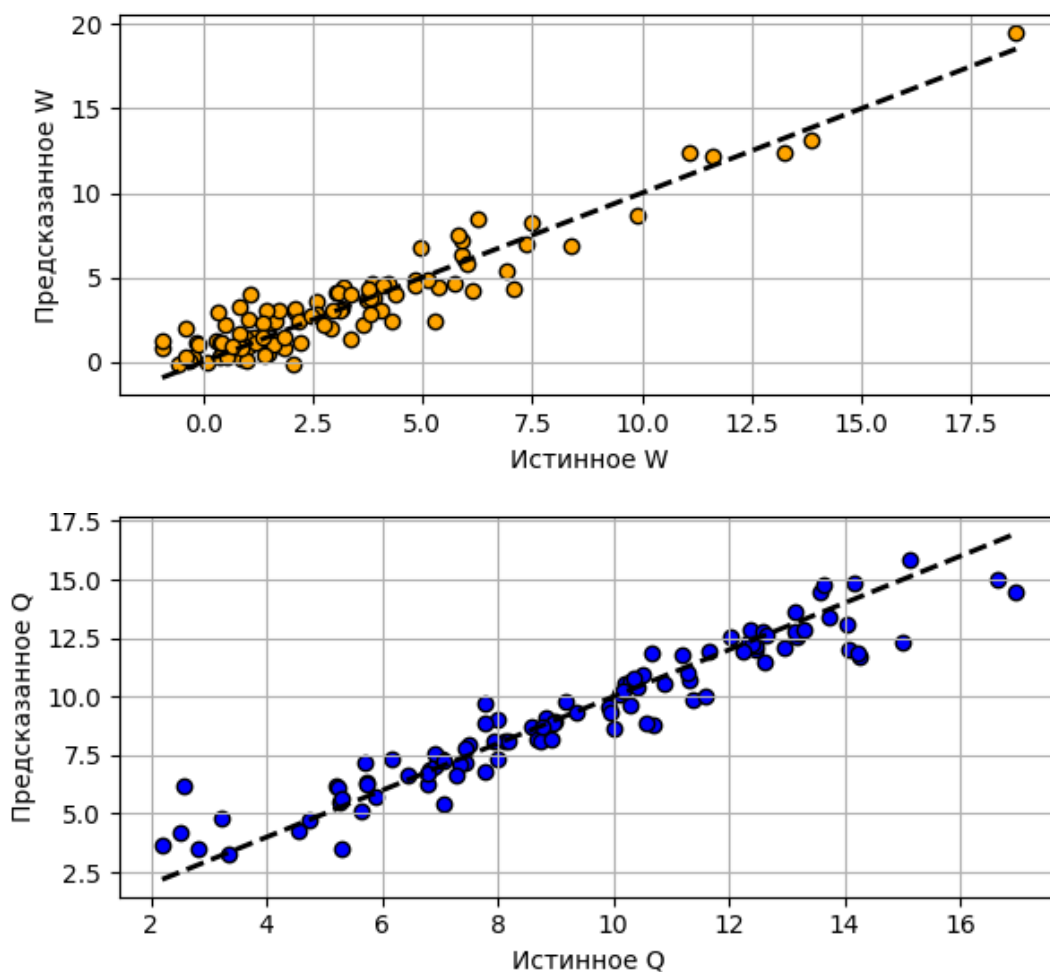


Рис. 4. Результаты предсказаний для характеристик СМО с использованием ансамблевых методов AdaBoost и StackingRegressor (получены автором)

В главе 3 проведено численное исследование стабильных режимов работы СМО по длине очереди и времени ожидания заявки в очереди.

Сначала исследуется характер поведения моментов длины очереди и времени ожидания заявки в очереди с изменением числа фаз обслуживания r . В результате выявляется, что существует некоторое граничное значение r , соответствующее точке пересечения моментов числа заявок в очереди, которое делит ось r на две части: область, в которой СМО длины очереди находится в пределах математического ожидания (соответствует стабильному режиму), и область, в которой разброс числа заявок в очереди превышает математическое ожидание (соответствует нестабильному режиму). Для исследования поведения r граничного при изменении нагрузки на систему со стороны входного потока с помощью разработанного программного комплекса был проведен цикл вычислительных экспериментов. По результатам вычислительных экспериментов установлена функциональная зависимость граничного числа фаз обслуживания, при котором обеспечивается стабильный режим работы системы, от приведенных интенсивностей потока, которая носит логарифмический характер:

$$r(\rho) = a - b * \ln(\rho). \quad (12)$$

В главе 4 описывается процесс разработки программного продукта с применением параллельной библиотеки платформы .NET.

В п. 4.1 приводится обоснование выбора платформы .NET в качестве инструментария разработки программных продуктов для реализации сложных систем массового обслуживания.

В п. 4.2 приводятся принципы разработки многоканальной СМО с ограниченной очередью с применением parallel framework. При добавлении заявок в очередь они обрабатываются параллельно каналами. Если очередь достигает своего предела, новые заявки отклоняются.

Использование Parallel Framework платформы .NET позволяет разрабатывать высокопроизводительные многоканальные СМО с ограниченной очередью. Средства параллельного программирования облегчают реализацию многопоточности, а применение техник балансировки нагрузки, динамического масштабирования и мониторинга позволяет оптимизировать систему для конкретных потребностей и условий работы.

В п. 4.3 описывается использование конкурентных коллекций вместе с Parallel Framework на платформе .NET, что позволяет эффективно реализовывать многоканальные СМО с ограниченной очередью. Конкурентные коллекции обеспечивают потокобезопасное управление данными, в то время как Parallel Framework позволяет легко распараллеливать обработку заявок по множеству каналов. Это не только улучшает производительность системы, но и обеспечивает более высокую точность моделирования в средах с высокой степенью параллелизма.

Другой важный аспект – это эффективное распределение ресурсов между каналами. Можно использовать Parallel Framework для динамического балансирования нагрузки между каналами обслуживания. Такой подход позволяет системе автоматически адаптироваться к изменениям в потоке заявок и распределять ресурсы между каналами наиболее эффективным образом.

В п. 4.4 было проведено сравнительное тестирование двух реализаций системы массового обслуживания 10 заявок по сокрытию файлов (объемом ~600 КБ) с

применением приложения для ассоциативной защиты файлов с использованием Parallel Framework и без его использования. Тестирование проводилось на процессоре Intel Core i5-9300H с базовой частотой 2,4 ГГц под управлением ОС Windows 11 Pro. Исходя из результатов тестирования, среднее время обработки одной заявки в реализации с использованием Parallel Framework составило 69 678 миллисекунд, в то время как в реализации без использования Parallel Framework – 240 345 миллисекунд. Это указывает на улучшение производительности в 3,45 раз при использовании Parallel Framework.

В заключении приводятся основные результаты, полученные в данной работе, а также рекомендации и перспективы дальнейшей разработки темы.

В приложении приводятся акты о внедрении результатов диссертационной работы, а также свидетельства о регистрации электронного ресурса.

ЗАКЛЮЧЕНИЕ

Основные результаты, полученные в данной работе, заключаются в следующем.

1. Разработана новая математическая модель СМО смешанного типа, отличающаяся от известных моделей сочетанием свойств СМО с ограниченным временем ожидания заявки в очереди и СМО с временем обслуживания, распределенным по закону Эрланга. Аналитическими методами получены общие формулы для характеристик модели в стационарном и нестационарном режимах, позволяющие задавать временные рамки нахождения заявок в очереди и количество фаз обслуживания.

2. Предложен новый метод исследования СМО с эрланговски распределенным временем обслуживания, позволяющий найти требуемое количество фаз, обеспечивающее заданный уровень качества обслуживания, которое может быть оценено с помощью коэффициента вариации различных характеристик СМО. Выявлено, что граничные значения числа фаз обслуживания логарифмически связаны с изменением приведенной интенсивности входного потока заявок.

3. Разработана имитационная модель СМО смешанного типа с использованием инструментальной среды AnyLogic PLE, а также метамодель, аппроксимирующая ключевые характеристики СМО на основе методов машинного обучения. Метамодель позволяет оперативно оценивать показатели эффективности системы по заданным входным параметрам без проведения ресурсоёмких имитационных экспериментов, что значительно ускоряет процесс анализа.

4. Разработан комплекс специализированных программных средств для решения задач по расчету ключевых характеристик СМО и оценки её эффективности. Проведено сравнительное тестирование двух реализаций СМО 10 заявок по сокрытию файлов (объемом ~600 КБ) с применением приложения для ассоциативной защиты файлов с использованием библиотеки Parallel Framework и без ее использования. Показано, что при использовании Parallel Framework производительность системы повышается в 3,45 раз.

Рекомендуется при проектировании различного рода технических систем, которые могут быть описаны математическими моделями в терминах теории массового обслуживания, учитывать выводы, представленные в настоящей работе.

Перспективы дальнейшей разработки темы должны охватывать построение аналитических выражений, описывающих другие немарковские модели СМО, а также и их эксплуатационные характеристики при различном количестве обслуживающих

устройств. Полученные результаты открывают ряд перспективных возможностей для управления сложными производственными системами. Использование этих результатов позволит более точно оценивать характеристики очередей и временные задержки в таких системах.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в журналах, входящих в перечень ведущих рецензируемых научных журналов и изданий, рекомендованных ВАК Министерства науки и высшего образования Российской Федерации

1. **Антонова П.В.**, Гиздатуллин Р.М., Титовцев А.С., Казаков М.В. Разработка метамодели системы массового обслуживания с использованием ансамблевых методов машинного обучения // Научно-технический вестник Поволжья. 2025. № 2. С. 18-21.

2. **Антонова П.В.** Разработка многоканальной системы массового обслуживания с ограниченной очередью с применением параллельной библиотеки платформы .NET // Computational Nanotechnology. 2023. Т. 10. № 3. С. 44-50.

3. Титовцев А.С., **Антонова П.В.** Характеристики системы массового обслуживания с ограниченным временем пребывания заявки в очереди и временем обслуживания, распределенным по закону Эрланга // Научно-технический вестник Поволжья. 2021. № 8. С. 79-82.

4. Титовцев А.С., **Антонова П.В.** Численное исследование стабильных режимов работы систем массового обслуживания с ограниченным средним временем пребывания заявки в системе и временем обслуживания, распределенным по закону Эрланга // Научно-технический вестник Поволжья. 2021. № 8. С. 83-85.

5. **Антонова П.В.**, Титовцев А.С. Характеристики нестационарного режима систем массового обслуживания смешанного типа с ограниченным временем пребывания заявки в очереди // Научно-технический вестник Поволжья. 2021. № 12. С. 171-173.

Публикация в сборнике, индексируемом в МБД Scopus:

6. **Antonova P.**, Titovtsev A., On Real Queue Length In A Queueing System With Erlang-R Service Time // Dynamics of Complex Networks and their Application in Intellectual Robotic, collection of materials of the IV International School-Conference of Young Scientists. Moscow, 2020. P. 29-33.

Прочие публикации:

7. **Антонова П.В.**, Гиздатуллин Р.М., Казаков М.В. Свидетельство о регистрации электронного ресурса №25456 «Ансамблевая метамодель для оценки параметров СМО». М.: ОФЭРНиО, 2025.

8. **Антонова П.В.**, Казаков М.В., Гиздатуллин Р.М. Разработка метамодели системы массового обслуживания с использованием ансамблевых методов машинного обучения // Перспективные фундаментальные исследования и научные методы: сборник статей международной научной конференции. – СПб.: МИПИ им. Ломоносова, 2024. С. 29-31.

9. **Антонова П.В.** Численное исследование стабильных режимов работы СМО // Перспективные исследования в технических и естественных науках: сборник статей международной научной конференции. – СПб.: МИПИ им. Ломоносова, 2024. С. 17-18.

10. **Антонова П.В.** Принципы разработки систем массового обслуживания с ограниченной очередью на платформе .NET // Программные системы и вычислительные методы. 2023. № 2. С. 15-28.

11. **Антонова П.В.** Исследование систем массового обслуживания смешанного типа // Навигатор в мире науки и образования. 2021. № 4 (53). С. 381-385.

12. **Антонова П.В.** Свидетельство о регистрации электронного ресурса №24942 «Исследование систем массового обслуживания смешанного типа». М.: ОФЭРНиО, 2021.