

ОРДЕНА ТРУДОВОГО КРАСНОГО ЗНАМЕНИ ФЕДЕРАЛЬНОЕ
ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «МОСКОВСКИЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ СВЯЗИ И ИНФОРМАТИКИ»

На правах рукописи

Канищева Маргарита Геннадьевна

Разработка модели и процедур эффективного
распределения ресурса мультисервисных узлов доступа
при обслуживании гетерогенного трафика современных
коммуникационных приложений

Специальность 2.2.15 — Системы, сети и устройства
телекоммуникаций

Диссертация на соискание ученой степени
кандидата технических наук

Научный руководитель:
доктор технических наук, профессор
Степанов Сергей Николаевич

Москва — 2025 г.

Оглавление

Введение	4
1 Распределение ресурса передачи информации в мультисервисном узле доступа	9
1.1 Введение к разделу 1	9
1.2 Актуальность проблематики	9
1.3 Мобильные сети четвертого и пятого поколений	11
1.3.1 Мобильные сети 4-го поколения	11
1.3.2 Мобильные сети 5-го поколения	14
1.3.3 Управление ресурсами в мобильных сетях на примере 4G	16
1.3.4 Пропускная способность узла доступа	21
1.4 Концепции и бизнес-модели использования узлов доступа	29
1.4.1 Network Slicing	30
1.4.2 Интернет вещей	33
1.4.3 Сервисы систем наблюдений	37
1.5 Обзор публикаций по тематике исследования	40
1.6 Постановка задачи диссертационного исследования	42
1.7 Выводы по результатам первого раздела	42
2 Построение и анализ обобщенной модели обслуживания гетерогенного трафика в мультисервисных узлах доступа	44
2.1 Введение к разделу 2	44
2.2 Математическое описание узла доступа	45
2.2.1 Ресурс передачи информации	45
2.2.2 Сложности пересылки гетерогенного трафика	45
2.2.3 Функциональная модель	47
2.2.4 Марковский процесс	48
2.2.5 Характеристики обслуживания заявок	50
2.3 Система уравнений равновесия	53
2.3.1 Рекурсивное формирование СУР	53
2.3.2 Анализ событий, меняющих состояние модели	54
2.3.3 Запись СУР одним соотношением	57
2.4 Частные случаи обобщенной модели узла доступа	58
2.5 Выводы по результатам второго раздела	59
3 Алгоритмы оценки характеристик обобщенной модели узла	62
3.1 Введение к разделу 3	62
3.2 Точная оценка характеристик узла методом Гаусса-Зейделя	62

3.2.1	Формулировка метода	62
3.2.2	Повышение эффективности расчетной процедуры	64
3.3	Оценка характеристик узла в частных случаях	65
3.3.1	Обслуживание файлов в режиме сервисов реального времени	65
3.3.2	Анализ эффективности дисциплины Processor Sharing при обслуживании эластичного трафика	68
3.3.3	Обслуживание файлов в режиме сервисов реального времени без ожидания	72
3.3.4	Анализ характеристик трафика реального времени	77
3.3.5	Оценка характеристик эластичного трафика	78
3.4	Приближенная оценка характеристик передачи эластичного трафика	82
3.5	Выводы по результатам третьего раздела	84
4	Оценка пропускной способности мультисервисного узла доступа в условиях учета особенностей формирования нагрузки	86
4.1	Введение к разделу 4	86
4.2	Анализ влияния неоднородности входных потоков на оценку ресурса	87
4.2.1	Общие положения	87
4.2.2	Обслуживание файлов по правилам сервисов реального времени	88
4.2.3	Обслуживание файлов по правилам эластичного трафика	92
4.3	Анализ сценариев дифференцированного обслуживания гетерогенного трафика	96
4.3.1	Использование архитектуры Network Slicing	96
4.3.2	Использование ожидания при передаче файлов	101
4.3.3	Использование выгрузки файлов	103
4.4	Особенности планирования пропускной способности узлов доступа	104
4.4.1	Оценка влияния группового поступления поступления файлов	104
4.4.2	Точные и приближенные алгоритмы оценки ресурса передачи информации	107
4.5	Выводы по результатам четвертого раздела	109
	Заключение	112
	Список литературы	114
	Приложение	126
	Приложение 1	127
	Приложение 2	128
	Приложение 3	129

Введение

Актуальность темы исследования. Развитие современных телекоммуникаций направлено на создание условий по качественному обслуживанию информационной нагрузки, создаваемой разнообразными пользователями услуг связи, большая часть которых — это технические устройства, предназначенные для сбора и передачи данных наблюдений. Информационные потоки, порожденные действующими и перспективными коммуникационными приложениями, характеризуются разными требованиями к скорости передачи, импульсным характером поступления сообщений, разной степенью терпимости к задержке, наличием или отсутствием свойства эластичности и т.д. Перечисленные особенности формирования и обслуживания поступающих заявок необходимо учитывать в процессе реализации новых телекоммуникационных технологий и концепций. В противном случае, как показали результаты статистических измерений и моделирования сетей связи, возникают проблемы с качеством предоставления услуг.

Отмеченные свойства создания и пересылки информационной нагрузки исследуются с помощью мультисервисных моделей. Они представляют большой интерес для практики и активно изучаются специалистами в области связи. Для обоснованного решения задачи планирования ресурса передачи информации узлов доступа и анализа условий дифференцированного обслуживания гетерогенной нагрузки необходимо построить новые математические модели, учитывающие рассмотренные выше особенности генерации информационных потоков, и провести их математический анализ с целью разработки эффективных точных и приближенных методов вычисления характеристик качества обслуживания поступающих запросов. Данные вопросы исследуются в работе Канищевой М.Г., что подтверждает актуальность темы, а также теоретическую и практическую ценность результатов.

Степень разработанности темы. Задачам оценки характеристик качества обслуживания поступающих запросов в разнообразных системах связи посвящены работы российских и зарубежных авторов. В их числе: Г.П. Башарин, В.М. Вишневский, Ю.В. Гайдамака, В.Г. Карташевский, И.А. Кочеткова, А.Е. Кучерявый, Е.А. Кучерявый, В.А. Наумов, А.П. Пшеничников, А.В. Росляков, К.Е. Самуйлов, С.Н. Степанов, М.С. Степанов, И.И. Цитович, и др., а также — T.Bonald, F.P. Kelly, V.B. Iversen, K.W. Ross, J. Virtamo и др. В ряде диссертационных исследований, таких как работы С.Д. Андреева, В.О. Бегешева, Е.А. Кучерявого, К.А. Агеева, Ж. Ндайикунда и других, рассматривались отдельные аспекты моделирования и анализа процессов генерации и обработки заявок в беспроводных узлах доступа. Предыдущие научные изыскания по данной тематике либо имеют чересчур завышенную теоретическую направленность, что затрудняет использование полученных результатов на практике, либо являются инженерными разработками, базирующимися на

эмпирических результатах, что делает невозможным их применение при изменении параметров решаемой задачи. В рамках данного исследования предлагается использовать многомерные марковские процессы для построения модели распределения ресурса и технику декомпозиции модели на отдельные модули для построения приближенных алгоритмов оценки характеристик обслуживания поступающих потоков трафика.

Цель работы. Построение обобщенной модели обслуживания в беспроводном узле доступа мультисервисного трафика современных коммуникационных приложений с учетом гетерогенных свойств поступающей нагрузки и импульсного характера формирования трафика данных.

Научная задача исследования заключается в обосновании возможности решения задачи планирования требуемой пропускной способности мультисервисного узла доступа, а также в разработке процедур дифференцированного обслуживания поступающих информационных потоков с учетом перечисленных выше особенностей их формирования и распределения пропускной способности узла доступа.

Для реализации поставленной цели и решения научной задачи требуется выполнить ряд взаимосвязанных частных исследований:

- разработать и проанализировать модель совместной обработки в беспроводном узле доступа приоритетного трафика онлайн-услуг и файловых данных с переменной скоростью передачи, учитывающую импульсный характер поступления файлов, а также возможность задержек и снижения актуальности передаваемой информации;
- на основе предложенной модели оценить ключевые параметры качества обслуживания поступающих запросов;
- разработать точные и приближенные методы вычисления полученных характеристик;
- сформировать практические рекомендации по эффективному распределению пропускной способности узла доступа между разнородными потоками трафика.

Объект исследования — беспроводной узел доступа, работающий в сети LTE¹.

Предмет исследования — алгоритмы распределения пропускной способности узла между обслуживаемыми информационными потоками.

Научная новизна.

1. Разработана и исследована комплексная модель совместного обслуживания разнородного трафика в беспроводном узле доступа. В отличие от существующих аналогов, данная модель позволяет учитывать ключевые особенности формирования и передачи данных, характерные для современных и перспективных коммуникационных сервисов. К таким особенностям относятся: зависимость скорости передачи от типа запроса, приоритетность нагрузки онлайн-услуг, возможность поступления файлов группами случайного размера, наличие или отсутствие у файлов эластичных свойств, присутствие у файлов возможности ожидания начала передачи и утраты ценности передаваемой информации с течением времени.

¹ Данное допущение не носит принципиального характера. Также могут рассматриваться узлы доступа стандартов 5G и 6G.

2. С помощью построенной математической модели узла получены выражения для оценки основных характеристик совместной передачи поступающих информационных потоков. Это доля запросов, получивших отказ, средний объем использованного ресурса передачи информации, среднее время передачи файла, среднее число файлов, ожидающих начала передачи. В отличие от более ранних исследований полученные выражения позволяют изучить действие учтенных в модели особенностей формирования потоков заявок и их обслуживания на решение задачи планирования ресурса и создание условий по их дифференцированному обслуживанию.
3. Стационарные вероятности модели и связанные с ними характеристики можно определить, решая систему уравнений баланса. В отличие от стандартных методов решения этой задачи, разработанный в диссертации алгоритм на основе метода последовательных замещений Гаусса-Зейделя позволяет рассчитывать характеристики моделей, включающих до нескольких миллионов состояний.
4. Для приближенной оценки характеристик узла предложена процедура декомпозиции модели на две составляющие. В каждой из них рассматривается по отдельности процесс обслуживания трафика сервисов реального времени и пересылка эластичного трафика файлов. В отличие от предыдущих исследований, расчет параметров обслуживания входящих запросов можно проводить с использованием несложных рекуррентных формул, применимых при любых значениях структурных параметров. Алгоритм ориентирован на реализацию в области малых потерь, где решается задача планирования требуемой пропускной способности узла доступа.

Теоретическая и практическая значимость работы. Теоретическая ценность исследования состоит в разработке и анализе комплексной модели обработки мультисервисного трафика современных приложений в беспроводном узле доступа с учетом гетерогенных свойств поступающей нагрузки и импульсного характера формирования трафика данных, а также в разработке алгоритмов расчета характеристик обслуживания заявок в подобных моделях. Практическая значимость работы заключается в обосновании возможности решения задачи планирования требуемой пропускной способности мультисервисного узла доступа, а также в разработке процедур дифференцированного обслуживания поступающих информационных потоков с учетом перечисленных выше особенностей их формирования и распределения ресурса передачи информации. Основные результаты диссертации получены при выполнении ряда НИР, включенных в результате конкурсного отбора в проектную часть тематического плана научных исследований МТУСИ в 2022–2025 гг, где автор был исполнителем. Результаты диссертации использованы в учебном процессе на кафедре «Сети связи и системы коммутации» МТУСИ и внедрены в компаниях ООО «ЕПСКОМ», ЗАО «Испытательный центр МирТелеТест». Реализация результатов работы подтверждена соответствующими актами.

Методология и методы исследования. Для решения поставленной задачи применялись методы теории вероятностей, вычислительной математики и математического моделирования.

Основные положения, выносимые на защиту:

1. Построенная комплексная модель мультисервисного беспроводного узла доступа дает возможность решить задачу планирования требуемой пропускной способности узла с учетом гетерогенных свойств поступающей нагрузки и импульсного характера поступления файлов.
2. Разработанная алгоритмическая версия итерационного метода Гаусса-Зейделя позволяет расширить возможности оценки значений стационарных вероятностей состояний построенной модели узла доступа, а с ними и точных величин характеристик обслуживания поступающих заявок для моделей с числом состояний до нескольких миллионов.
3. Для приближенного решения задачи планирования ресурса узла доступа предлагается использовать процедуру декомпозиции модели на две составляющие. В каждой из них рассматривается по отдельности процесс обслуживания трафика сервисов реального времени и пересылка эластичного трафика файлов. В каждом конкретном случае вычисление характеристик возможно на основе несложных рекурсивных формул при любых заданных структурных параметрах модели.
4. Расчеты с использованием модели показали, что групповой характер поступления файлов увеличивает требование к ресурсу передачи необходимому для обслуживания поступающих потоков заявок с заданным качеством. Игнорирование импульсных свойств потока данных может внести ошибку в оценку ресурса, достигающую трети итогового значения пропускной способности узла.
5. Показано, что в ситуации перегрузки условия по дифференцированному обслуживанию эластичного трафика можно создать на общем ресурсе без увеличения пропускной способности узла доступа, если воспользоваться возможностью ожидания файлов или их выгрузки на другие узлы. В зависимости от значений входных параметров экономия в требуемой величине пропускной способности достигает 10%–20% по сравнению с использованием традиционной модели слайсинга.

Соответствие диссертации паспорту научной специальности. Специальность диссертационной работы — 2.2.15 Системы, сети и устройства телекоммуникаций. Работа соответствует пунктам паспорта 1, 2 и 6 указанной специальности.

Степень достоверности и апробация результатов. Достоверность результатов диссертации обеспечивается корректностью используемого математического аппарата теории телетрафика и подтверждена численными экспериментами. Результаты работы обсуждались и докладывались на Международной научной конференции «Systems of Signals Generating and Processing in the Field of on Board Communications» (Москва, 2022-2024 гг.), на международной научной конференции «Distributed Computer and Communication Networks: Control, Computation, Communications» (Москва, 2024 г.), на отраслевой научно-технической конференции «Телекоммуникационные и вычислительные системы» (Москва, 2022 - 2023 гг.); на научных семинарах кафедры «Сети связи и системы коммутации». Основные результаты диссертации изложены в 17 опубликованных работах. В их числе 6

работ в изданиях, входящих международную базу цитирования WoS и SCOPUS, 3 работы в изданиях, включенных в список ВАК РФ, 3 работы в изданиях, входящих в базу РИНЦ, и 4 свидетельства о государственной регистрации программ для ЭВМ (Роспатент).

Основное содержание работы. Диссертация состоит из введения, четырех разделов, заключения, списка литературы и приложения. Основная часть (без приложения) изложена на 125 страницах машинописного текста, содержит 43 рисунка и 10 таблиц; список литературы состоит из 128 наименований.

Раздел 1

Распределение ресурса передачи информации в мультисервисном узле доступа

1.1 Введение к разделу 1

В разделе 1 рассматриваются основные характеристики мобильных сетей четвертого и пятого поколений, общие принципы управления ресурсами, построена общая модель исследуемого объекта. В подразделе 1.2 описывается актуальность проблематики. В подразделе 1.3 рассматриваются мобильные сети четвертого и пятого поколений, управление ресурсами в мобильных сетях стандарта LTE, дано определение и расчеты пропускной способности. Подраздел 1.4 содержит в себе концепции и модели использования узлов доступа, отдельно рассмотрены следующие пункты: network slicing, интернет вещей и сервисы систем видеонаблюдений. В подразделе 1.5 приведен обзор по тематике исследования. В подразделе 1.6 сформулированы задачи диссертационного исследования.

1.2 Актуальность проблематики

Одной из важнейших современных тенденций развития телекоммуникационных технологий является создание инфраструктурных решений, удовлетворяющих запросы различных групп пользователей, значительную долю среди которых составляют интеллектуальные устройства, генерирующие и обменивающиеся большими объемами данных. Это требует трансформации сетей связи из обычной среды передачи сигналов в интеллектуальную платформу, обеспечивающую высококачественный доступ к широкому спектру сервисов. Ключевым решением здесь выступает технология сетевого сегментирования (Network Slicing), оптимально распределяющая ресурсы между различными видами трафика с соблюдением необходимых уровней качества обслуживания. Учитывая ограниченность этих ресурсов, задача их эффективного управления приобретает особую значимость. Цель настоящего исследования состоит в разработке и анализе подходов к рациональному распределению ресурсов с использованием технологии Network Slicing для разнотипного трафика. Актуальность данной проблемы обусловлена необходимостью её решения для успешного функционирования современных коммуникационных систем.

Предполагается, что рассматриваемый узел связи работает автономно в качестве отдельной ячейки сети стандарта LTE или выше, при этом часть его ресурсов предоставляется оператором мониторинга для доставки информационных потоков, создаваемых видеокамерами и сенсорами телеметрической аппаратуры. На рис. 1.1 показана архитектурная

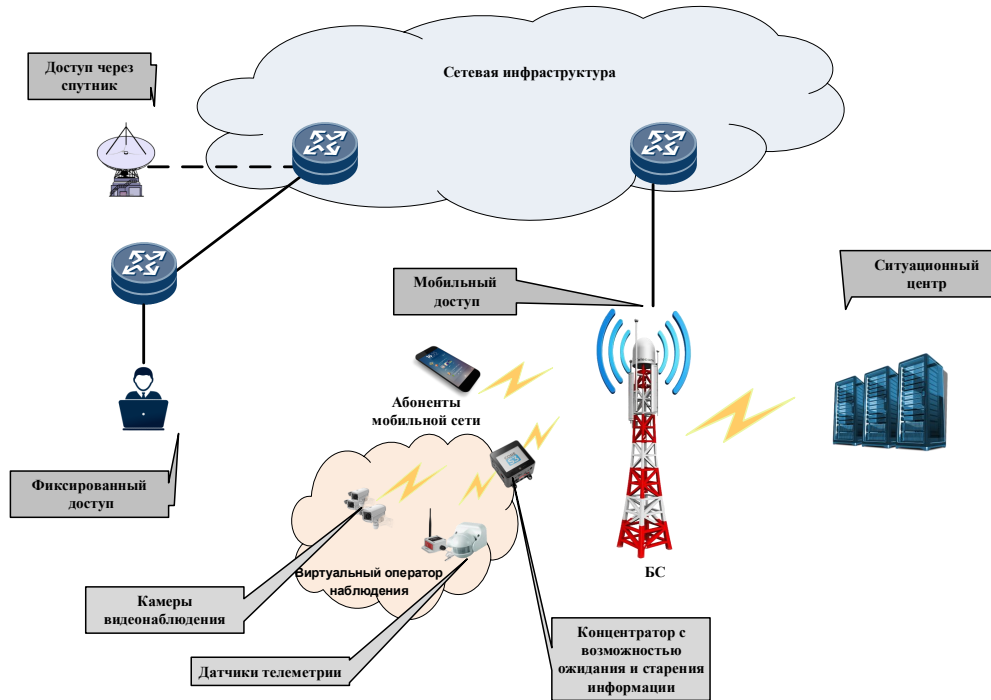


Рисунок 1.1 — Архитектура сети виртуального оператора систем наблюдения

структура. Распределение необходимого объема ресурсов осуществляется с помощью специальных механизмов, одним из которых является технология сегментации сети (Network Slicing). В телекоммуникациях основным показателем доступности ресурса выступает понятие «пропускная способность», которое характеризует техническую возможность узла поддерживать требуемый уровень сервиса для пользователей при информационном взаимодействии. Структуру функциональных возможностей узла иллюстрирует рис. 1.2.

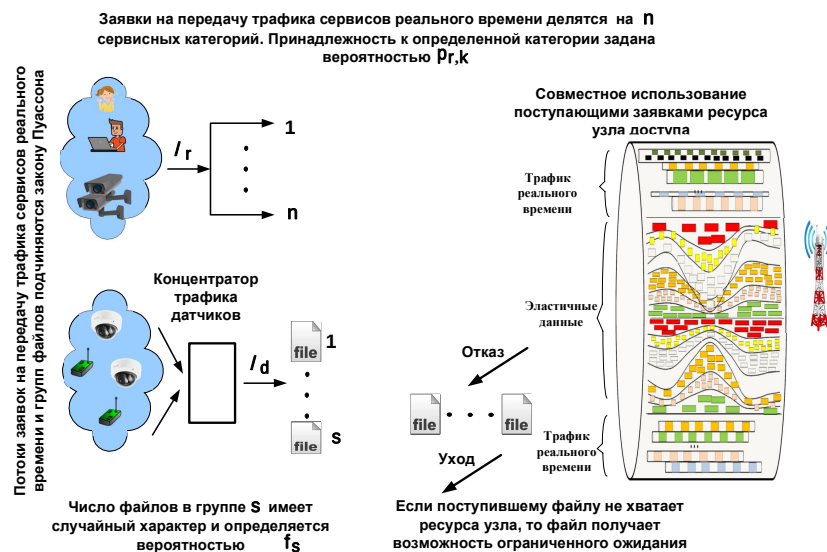


Рисунок 1.2 — Функциональная модель узла доступа

1.3 Мобильные сети четвертого и пятого поколений

1.3.1 Мобильные сети 4-го поколения

Технология LTE (Long-Term Evolution) представляет собой стандарт беспроводной широкополосной связи, позволяющий передавать большие объемы данных на высокой скорости для мобильных телефонов и других устройств. Стандарт был разработан 3GPP и определен в серии документов Release 8 и Release 9. Внедрение стандарта началось с конца 2009 года. В разных странах используются различные частоты и полосы для LTE, что требует использования многодиапазонных телефонов для подключения к сетям четвертого поколения по всему миру. Архитектура сети LTE построена на взаимодействии трех основных компонентов: пользовательское оборудование (UE, User Equipment), ядро сети (EPC, Evolved Packet Core), сеть радиодоступа (E-UTRAN, Evolved Universal Terrestrial Radio Access Network). Пользовательское оборудование (User Equipment, UE) осуществляет связь с базовой станцией и обеспечивает доступ к услугам сети. Радиосеть доступа (Evolved Universal Terrestrial Radio Access Network, E-UTRAN) представлена единственным элементом — базовой станцией eNodeB (eNB), выполняющей управление радиоресурсами и координацию взаимодействия с ядром сети. Наконец, ядро сети (Evolved Packet Core, EPC) организует коммутацию пакетов и поддерживает необходимые функциональные модули, включая обслуживающий шлюз (Serving Gateway, S-GW), шлюз пакетных данных (Packet Data Network Gateway, P-GW) и систему управления мобильностью (Mobility Management Entity, MME), обеспечивающую идентификацию и авторизацию пользователей, а также координирующую процесс переключения между различными зонами покрытия [9]. Архитектура сети мобильной связи стандарта LTE показана на рисунке 1.3.

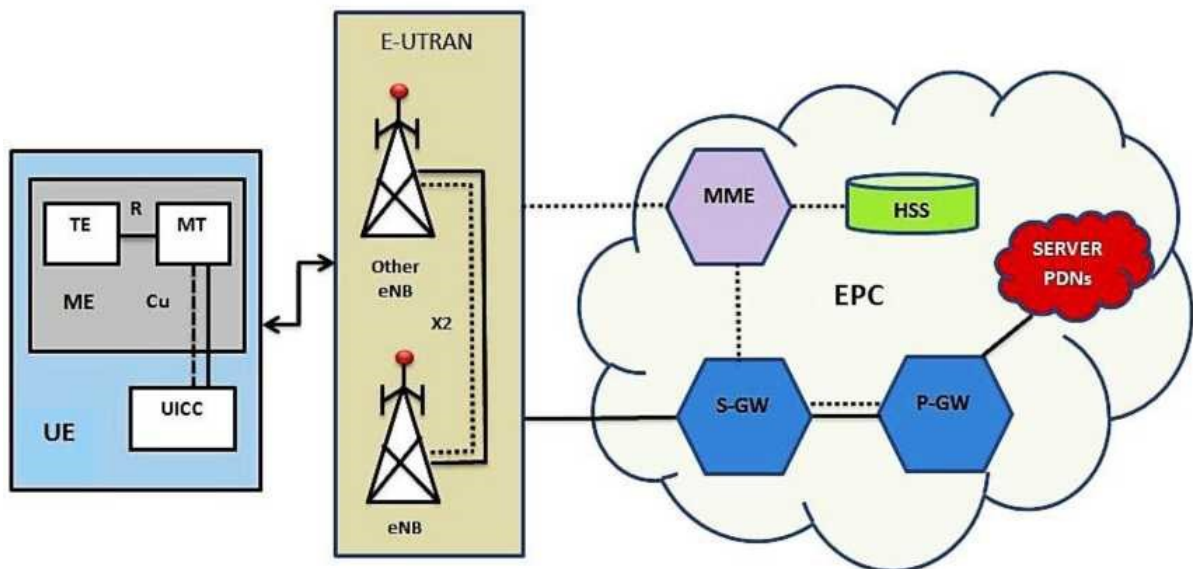


Рисунок 1.3 — Архитектура сети мобильной связи стандарта LTE [68,80]

В технологии LTE пропускная способность увеличена в 20 раз по сравнению с 3G. Средняя скорость загрузки в 3G составляет около 1–2 Мбит/с, а в 4G средняя скорость

может варьироваться от 10 до 50 Мбит/с и выше. Пропускная способность увеличена благодаря нескольким ключевым улучшениям: широкая полоса частот, использование ММО (Multiple Input Multiple Output), улучшенная модуляция, агрегация несущих, более высокая спектральная эффективность [16,97].

Полосы частот для LTE. Данная технология использует расширенную полосу частот для передачи данных, что позволяет передавать большее количество информации за фиксированный промежуток времени, что приводит к повышению скорости передачи. LTE поддерживает полосы пропускания частот от 1,4 МГц до 20 МГц и поддерживает как частотное разделение каналов (FDD), так и временное разделение (TDD). Полосы частот, используемые операторами связи, могут различаться [16,97].

Применение технологий ММО. Технология ММО представляет собой способ передачи данных в беспроводных коммуникациях, использующий множественные передающие и приёмные антенны для существенного повышения пропускной способности и качества канала связи. Принцип её функционирования основывается на пространственном мультиплексировании, позволяющем передавать сразу несколько независимых информационных потоков параллельно через разные антенны, что повышает эффективность использования спектра и снижает негативное воздействие эффектов многолучевого распространения сигналов [73].

Модуляция. Модуляция — это процесс преобразования цифровых данных в аналоговый сигнал, который затем передается по радиointерфейсу. Модуляция определяет способ представления цифровой информации в виде изменений амплитуды, фазы или частоты несущего сигнала. В LTE используются два основных типа модуляции: QPSK (Quadrature Phase Shift Keying) и QAM (Quadrature Amplitude Modulation). QPSK — это фазовая модуляция, которая изменяет фазу несущего сигнала для кодирования двух бит информации. В QPSK каждый символ несет две бита информации, так как возможны четыре состояния фазы: 0° , 90° , 180° и 270° . Преимущества QPSK включают низкое энергопотребление и устойчивость к шумам, что делает ее подходящей для ситуаций с низким уровнем сигнала. QAM — это комбинированная амплитудно-фазовая модуляция, которая изменяет как амплитуду, так и фазу несущего сигнала для кодирования большего количества бит информации. В технологии LTE используются различные уровни QAM, такие как 16-QAM (кодирует 4 бита информации на символ (16 возможных состояний)), 64-QAM (кодирует 6 бит информации на символ (64 возможных состояний)) и 256-QAM (кодирует 8 бит информации на символ (256 возможных состояний)). Чем выше уровень QAM, тем больше бит информации можно передать за один символ, однако это требует более высокого отношения сигнал-шум (SNR) для надежной передачи. Выбор конкретной схемы модуляции зависит от условий канала и требований к пропускной способности. Например, в условиях низкого SNR предпочтительнее использовать QPSK или низкий уровень QAM, поскольку они менее подвержены ошибкам. В условиях хорошего SNR можно использовать высокий уровень QAM для достижения большей пропускной способности. Также можно использовать режим адаптивной модуляции и кодирования, который динамически выбирает наилучшую схему в зависимости от текущих условий канала. Это позволяет максимально эффективно использовать доступные ресурсы и обеспечивать стабильную

связь [35,36].

Спектральная эффективность. Радиointерфейс LTE использует для нисходящего канала (от базовой станции к устройству) технологию OFDMA (Orthogonal Frequency Division Multiple Access) и SC-FDMA (Single Carrier FDMA) для восходящего канала (от устройства к базовой станции). Эти методы позволяют эффективно использовать спектр частот и обеспечивают высокую устойчивость к помехам [36].

Технология OFDMA — это метод множественного доступа, который используется в сетях LTE для передачи данных между базовой станцией (eNodeB) и пользовательскими устройствами (UE). При использовании технологии OFDMA общий диапазон частот разбивается на ортогональные поднесущие, число которых зависит от ширины канала и составляет 72, 180, 300, 600, 900 либо 1200. Каждая поднесущая допускает применение различных методов модуляции, таких как QPSK, 16-QAM и 64-QAM. Многопользовательская передача реализуется путем выделения определённых групп поднесущих разным пользователям внутри одного кадра. Ключевое достоинство OFDMA — способность успешно компенсировать негативные последствия многолучевого распространения сигнала при приёме. Тем не менее, эта технология имеет ряд недостатков. Один из основных — высокая чувствительность к частотной синхронизации. Кроме того, сигнал OFDMA характеризуется большим коэффициентом пиковой мощности к средней (PAPR), что вынуждает усилители работать в нелинейных режимах, снижая их эффективность. Это особенно критично для мобильных устройств с ограниченными энергетическими ресурсами. По этой причине в восходящем канале LTE применяется альтернативная технология множественного доступа, известная как SC-FDMA (Single Carrier Frequency Division Multiple Access)[3,67,96]. На рисунке 1.4 изображено представление OFDM и OFDMA-сигналов.

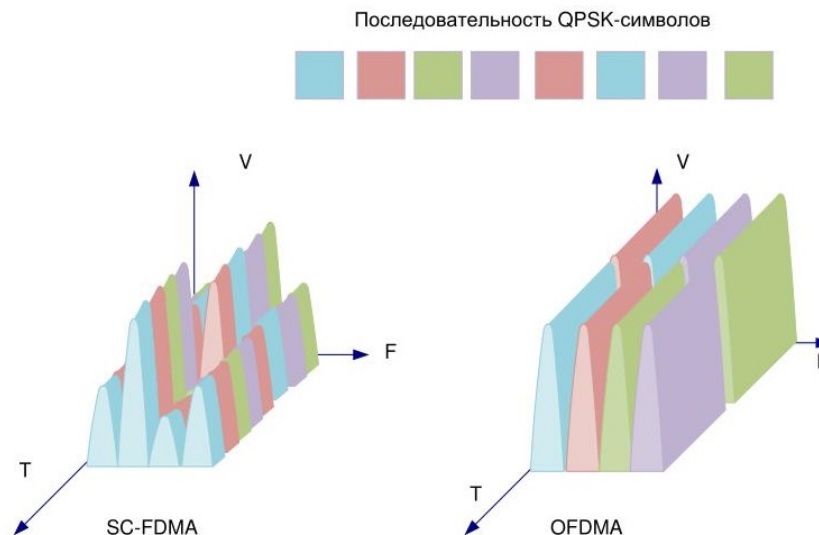


Рисунок 1.4 — Представление OFDM и OFDMA-сигналов [3,67]

Технология SC-FDMA представляет собой метод множественного доступа, используемый в сетях LTE для передачи данных в восходящем канале, то есть от абонентского

устройства к базовой станции. Эта технология представляет собой модификацию метода OFDMA, предназначенную для снижения пик-фактора мощности передаваемого сигнала. Метод SC-FDMA объединяет преимущества одночастотной передачи и многопользовательского доступа через разделение частот. Процесс передачи данных в SC-FDMA включает следующие этапы: преобразование исходного цифрового потока данных в символы, дискретное преобразование Фурье, распределение поднесущих, инверсное быстрое преобразование Фурье, передача сигнала, прием и обработка сигнала. Основное преимущество технологии SC-FDMA в сравнении с OFDMA заключается в уменьшенном пик-факторе мощности (PAPR). Благодаря применению одночастотной схемы, SC-FDMA значительно уменьшает пиковую мощность передаваемого сигнала, что имеет ключевое значение для мобильных устройств с ограниченной мощностью фккумулятора [3,67]. Недостатками SC-FDMA являются ограниченная поддержка различных скоростей передачи данных и высокие требования к ресурсам.

Агрегация несущих. При использовании агрегации несущих происходит объединение нескольких отдельных полос частот в единый широкий канал передачи данных. Это дает возможность операторам мобильным сетям адаптироваться под различные сценарии использования и увеличивать пропускную способность в условиях высокой нагрузки.

Совокупность этих технических решений позволила технологии LTE достичь значительных преимуществ перед предшествующими поколениями мобильной связи по скоростям передачи данных и стабильности соединений, обеспечивая качественное обслуживание в условиях современной высоконагруженной сети [96,97].

1.3.2 Мобильные сети 5-го поколения

Пятое поколение мобильной связи — это технология, следующая за стандартами 4G/ IMT-Advanced. Стандарт 5G определен международными организациями, такими как Международный союз электросвязи (МСЭ), который разработал план развития этой технологии и назвал её IMT-2020. Разработка и внедрение стандартов 5G продолжается, и многие страны, включая Россию, проводят тестирования и пилотные проекты для подготовки к массовому использованию этой технологии. Использование технологии 5G позволяет получить более высокую пропускную способность, более надежные системы коммуникации между устройствами и меньшее время задержки. Максимальные показатели стандарта 5G предусматривают достижение скоростей до 20 Гбит/с в направлении от сети к устройству (нисходящий канал) и до 10 Гбит/с в обратном направлении (восходящий канал), при этом задержка сигнала ограничена величиной менее 4 миллисекунд [14,18,19].

Стандарт 5G предусматривает три категории сценариев использования (см. рисунок 1.5): расширенная мобильная широкополосная связь (eMBB — enhanced Mobile Broadband), массовая машинная связь (mMTC — massive Machine Type Communication) и сверхнадежная связь с низкой задержкой (URLLC — Ultra Reliable Low Latency Communications). В первом случае речь идет о мультимедийных сервисах с новыми функциями, такими как виртуальная реальность, видеонаблюдение, улучшенные услуги мобильной широкополосной связи (MBB) в помещениях и на улице, облачные вычисления и т.д. Второй вариант

использования связан с передачей конфиденциальных данных умеренно малого объема без задержки между многочисленными подключенными устройствами (устройства должны быть недорогими и с длительным временем автономной работы), а также предназначен для предоставления услуг в различных областях, таких как интернет вещей, умные города, удаленный мониторинг и наблюдение, интеллектуальное сельское хозяйство и т.д. В то время как третий вариант использования предлагает критически важные услуги связи, к ним предъявляются очень строгие требования с точки зрения задержки, запаздывания, пропускной способности и т.д. Таким образом, его применение имеет решающее значение, и в качестве примеров можно привести систему мониторинга пациентов, беспилотные интеллектуальные транспортные средства, интеллектуальные сети и т.д. [68,86].

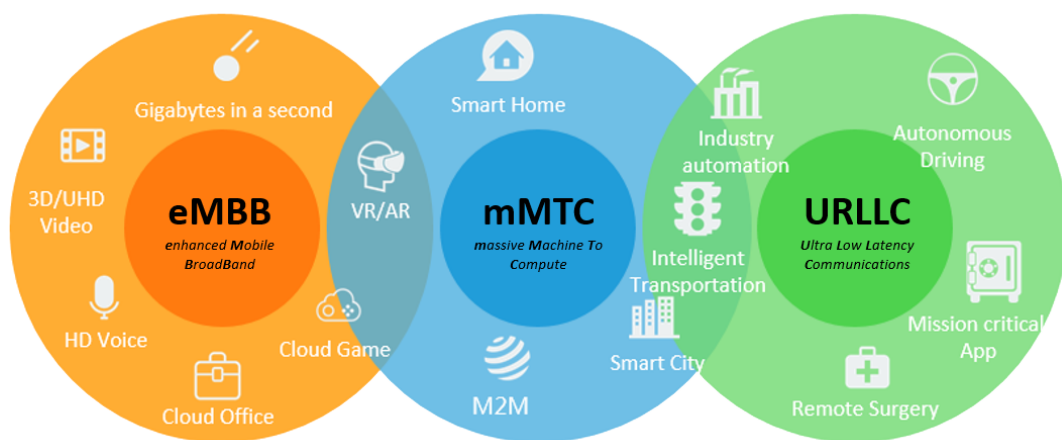


Рисунок 1.5 — Основные сценарии использования 5G

Диапазон радиочастот, применяемый для развертывания сотовых сетей пятого поколения, подразделяется на три большие группы. Первая группа охватывает частоты ниже 1 ГГц. Эти диапазоны характеризуются лучшими показателями распространения сигнала, что позволяет обеспечить широкое покрытие территории и глубокую проникаемость внутрь зданий. Это позволяет активно задействовать их для обеспечения большого числа интернет-вещей и расширения зоны покрытия мобильной широкополосной связи, что особенно актуально в городских агломерациях, пригородах и сельских районах. Вторая группа охватывает частоты от 1 до 6 ГГц (такая группа диапазонов имеет средние возможности проникновения и покрытия по сравнению с более ранней, в основном используется в текущих службах МВВ и может быть использована для первичного развертывания 5G). И третья группа охватывает частоты выше 6 ГГц (такой тип полос обладает значительной пропускной способностью и широкой полосой пропускания, которые могут быть использованы для расширенных приложений МВВ. Однако они имеют меньший охват и гораздо меньшие возможности проникновения) [85,88,94].

Разница между 4G и 5G. При сравнении технологий 5G и 4G основным преимуществом сетей пятого поколения является значительно увеличенная скорость передачи данных. В идеальных условиях сеть 5G обеспечивает максимальную скорость передачи данных до 20 Гбит/с, что в двадцать раз превосходит возможности 4G, чей максимум

ограничивается лишь 1 Гбит/с. Важнейшим преимуществом новой технологии является минимизация сетевой задержки до 1 мс, тогда как в 4G этот показатель варьируется от 10 до 50 мс. Дополнительно 5G отличается повышенной энергоэффективностью, устойчивостью к помехам и увеличенной общей емкостью сети [118].

1.3.3 Управление ресурсами в мобильных сетях на примере 4G

Управление ресурсами в сетях четвертого поколения играет ключевую роль в обеспечении эффективной передачи данных и оптимизации использования доступной полосы частот. Оно охватывает множество аспектов, включая распределение времени, мощности и частотных ресурсов между пользователями, а также управление качеством обслуживания. Давайте рассмотрим этот процесс подробнее. Основная цель управления ресурсами заключается в том, чтобы максимально эффективно использовать доступные физические каналы связи для удовлетворения требований всех активных пользователей. В технологии LTE это достигается благодаря нескольким механизмам: динамическое распределение ресурсов, планировщик, качество обслуживания, агрегация несущих).

Динамическое распределение ресурсов. В сетях четвертого поколения каждый кадр делится на временные слоты и частотные блоки (Resource Blocks, RBs). Эти ресурсы распределяются между пользователями в зависимости от их потребностей и текущей загрузки сети. Динамическое распределение ресурсов позволяет гибко изменять количество выделяемых каждому пользователю ресурсов в реальном времени, что помогает оптимизировать использование спектра и минимизировать задержки.

Планировщик является ключевым компонентом системы управления ресурсами. Его задача состоит в распределении временных слотов и частотных блоков между активными пользователями. На рисунках 1.6 и 1.7 представлен общий вид планирования радиоресурсов в восходящей и нисходящей линиях связи [68].

Планировщик производит распределение участков спектра среди пользователей для организации обмена информацией в направлениях приёма и передачи данных. Основная цель планировщика — эффективное удовлетворение запросов большого числа пользователей сети, учитывая такие важные аспекты, как качество радиосигнала, требования к уровню сервиса, приоритетность обслуживания и другие ключевые характеристики. Процесс распределения основан на анализе различных показателей для каждого ресурсного блока. Например, n -й ресурсный блок назначается пользователю с лучшими показателями. Процесс не стандартизирован и зависит от архитектуры eNodeB, что позволяет адаптировать алгоритмы под конкретные нужды [36,37].

Качество обслуживания (QoS, Quality of Service) играет важную роль для обеспечения эффективного распределения ресурсов между различными пользователями и приложениями. В стандарте LTE предусмотрены несколько ключевых параметров качества обслуживания (QoS), которые определяют приоритеты трафика и обеспечивают требуемый уровень сервиса для различных типов приложений. Вот основные параметры QoS:

1. QCI (QoS Class Identifier) — идентификатор класса качества обслуживания. Опреде-

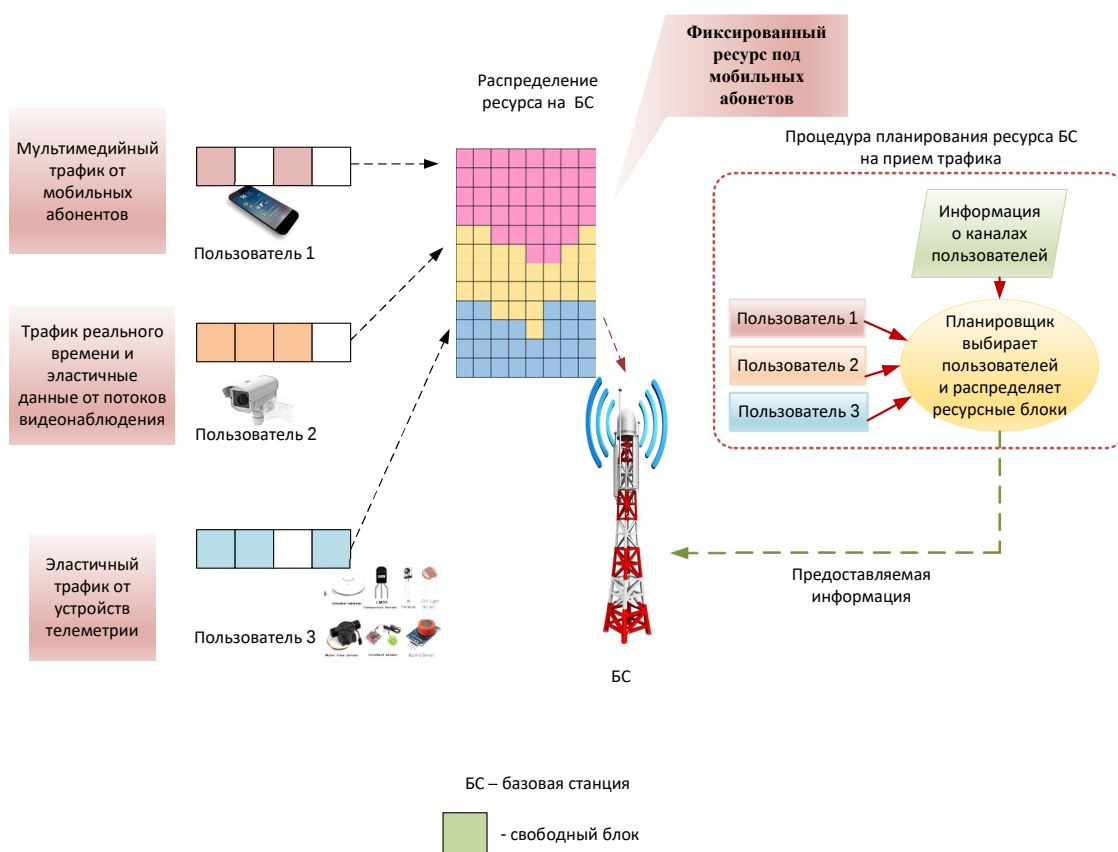


Рисунок 1.6 — Общая модель планировщика пакетов восходящей линии связи

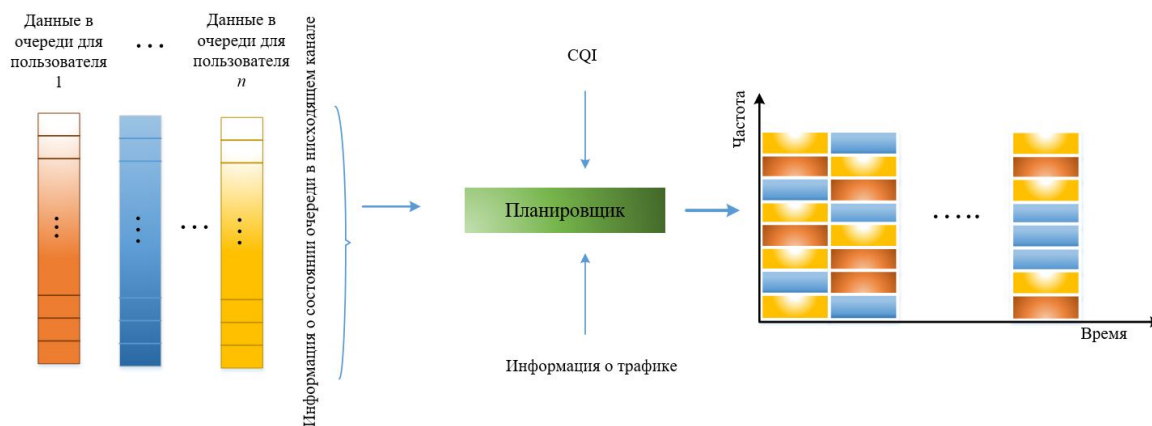


Рисунок 1.7 — Общая модель планировщика пакетов нисходящей линии связи

ляет приоритет трафика и соответствующие параметры задержки, потерь пакетов и пропускной способности. Стандартизованы девять классов QCI с разными уровня-

ми приоритета.

2. ARP (Allocation and Retention Priority) — приоритет выделения и удержания ресурса. Определяет важность трафика при выделении и удержании радиоресурсов.
3. MBR (Maximum Bit Rate) — максимальная битовая скорость. Ограничивает максимальную скорость передачи данных для конкретного сервиса.
4. GBR (Guaranteed Bit Rate) — гарантированная битовая скорость передачи данных. Она определяет минимальный порог пропускной способности, который должна гарантировать сеть для конкретного типа сервиса.
5. N-GBR (Non-Guaranteed Bit Rate) — негарантированная битовая скорость. Это значит, что для данного сервиса минимальная скорость передачи данных не гарантируется. Вместо этого трафик обслуживается на основе наилучших усилий (best-effort basis). Это значит, что трафик Non-GBR может конкурировать с другими потоками за доступные ресурсы сети, и его качество может варьироваться в зависимости от текущей загрузки сети.

Эти параметры позволяют операторам сети обеспечивать различный уровень качества обслуживания для разных типов трафика, таких как голосовые вызовы, видеоконференции, потоковое видео и другие приложения.

Идентификатор класса качества обслуживания характеризуется определенным уровнем приоритета, допустимой задержкой пакетов и приемлемым показателем потерь пакетов. Основная задача QCI — указать, какие требования предъявляются к передаче данных в зависимости от типа передаваемого трафика. Основные характеристики классов качества обслуживания охватывают следующие ключевые параметры: порядок приоритета обработки трафика, допустимый предел задержки, процент допустимых потерь пакетов и пропускная способность. Организация 3GPP в ходе разработки спецификаций стандарта LTE установила ряд категорий QoS-классов, каждому из которых соответствует определённый набор фиксированных параметров QCI, определяющих условия передачи данных и особенности обслуживания соответствующих видов трафика [2,90,93].

Таблица 1.1 демонстрирует стандартные классы качества обслуживания и их параметры. Исходя из требований к качеству обслуживания, однонаправленные радиоканалы подразделяются на две группы: каналы с гарантированной пропускной способностью, обеспечивающие стабильную минимальную скорость передачи данных для каждого вида трафика, и каналы с негарантированной пропускной способностью, где не предусмотрена обязательная поддержка определенной минимальной скорости передачи данных. Такие потоки чаще всего используются для сервисов, критичных ко времени доставки пакетов, например, для голосовых звонков (VoLTE), видеоконференций или потокового вещания. Например, оператор связи гарантирует, что даже при высокой загрузке сети передача данных будет идти с минимальной скоростью, достаточной для нормальной работы приложения. При использовании N-GBR сервисы получают ресурсы сети динамически, то есть битрейт не гарантируется, но сеть пытается обеспечить наилучшую возможную производительность в зависимости от текущей загрузки. Это характерно для большинства

Таблица 1.1 — Стандартизированные идентификаторы класса обслуживания QoS для LTE [88]

QCI	Тип канала	Приоритет	Допустимые задержки, мс	Допустимые потери	Пример приложений
1	GBR	2	100	10^{-2}	Телефония в реальном времени
2		4	150	10^{-3}	Видеотелефония, видео в реальном времени
3		5	300	10^{-6}	Низкокачественное видео (с буферной потоковой передачей)
4		3	50	10^{-3}	Игры в реальном времени
5		1	100	10^{-6}	IMS-сигнализация
6	Non-GBR	7	100	10^{-3}	Аудио, видео в реальном времени, интерактивные игры
7		6	300	10^{-6}	Видео с буферизацией
8		8	300	10^{-6}	TCP/IP, чат,
9		9	300	10^{-6}	FTP, P2P

обычных приложений, таких как веб-серфинг, просмотр видео, работа с социальными сетями и т.д.

Агрегация несущих позволяет объединить несколько физических каналов в один логический канал, увеличивая общую полосу пропускания и скорость передачи данных. Этот метод особенно полезен в условиях ограниченного спектра, когда оператор может комбинировать несколько частотных диапазонов для достижения лучших результатов. На рисунке 1.8 приведен пример агрегации несущих.

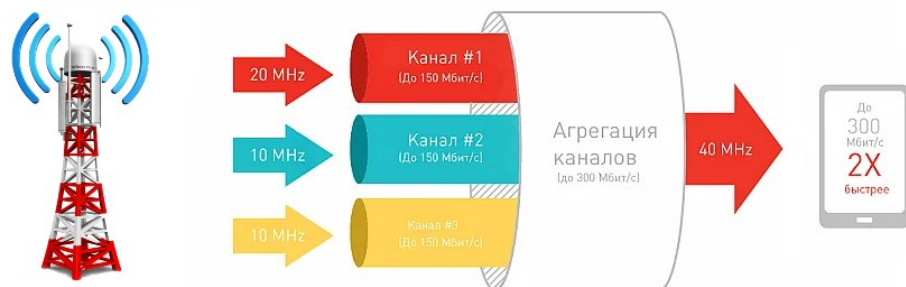


Рисунок 1.8 — Агрегация несущих

Процесс управления ресурсами в LTE является критически важным элементом, поз-

воляющим эффективно использовать доступные радиочастотные ресурсы и обеспечивать высокое качество обслуживания для множества пользователей. Управление ресурсами в LTE состоит из следующих шагов:

1. **Оценка состояния канала.** Прежде чем начать распределение ресурсов, eNodeB (базовая станция) должна оценить текущее состояние канала для каждого активного пользователя. Эта оценка основывается на нескольких параметрах:
 - CSI (Channel State Information): информация о состоянии канала, предоставляемая устройством пользователя. CSI включает такие показатели, как SINR (Signal-to-Interference-plus-Noise Ratio) и CQI (Channel Quality Indicator).
 - SINR: отношение сигнал-шум-плюс-интерференция, показывающее насколько сильно сигнал пользователя затенён шумом и другими сигналами.
 - CQI: индикатор качества канала, который показывает, какой уровень модуляции и кодирования может быть использован для передачи данных без ошибок.

На основе этих данных eNodeB определяет, сколько ресурсов потребуется каждому пользователю для обеспечения требуемого качества связи.

2. **Определение приоритетов и классов QoS.** После оценки состояния канала узел eNodeB начинает определять приоритеты для различных типов трафика. В LTE существует система классов QoS, которые определяют, какой уровень сервиса требуется для определенного типа трафика. Например: QCI 1 — голосовой трафик (например, VoLTE), требующий минимальной задержки и высокого приоритета; QCI 6 — потоковое видео, требующее стабильного потока данных, но допускающее небольшие задержки; QCI 9 — просмотр веб-страниц и электронной почты, для которых важна надежность доставки, но допустимы большие задержки. Каждый класс QoS имеет определенные параметры, такие как максимальная задержка, допустимый процент потерь пакетов и минимальная гарантированная пропускная способность.
3. **Работа планировщика.** Определение и работа планировщика описаны выше, само планирование происходит в eNodeB.
4. **Выделение ресурсов.** Когда планировщик принял решение о распределении ресурсов, eNodeB отправляет управляющие сообщения (DCI, Downlink Control Information) на физическом канале управления (PDCCH). Эти сообщения содержат информацию о том, какие временные слоты и частотные блоки выделены каждому пользователю. Пользовательские устройства получают эти сообщения и начинают передачу или приём данных в соответствии с выделенными ресурсами.
5. **Обработка обратной связи.** После завершения передачи данных пользователи отправляют отчёты о качестве связи обратно на eNodeB. Эти отчёты включают такие параметры, как успешность приёма данных, количество потерянных пакетов и текущее состояние канала. Эта обратная связь используется для коррекции распределения ресурсов в следующем цикле планирования.

6. **Повторение цикла.** Весь процесс управления ресурсами повторяется каждые несколько миллисекунд, позволяя системе постоянно адаптироваться к изменяющимся условиям сети

Как и любая технология, LTE имеет свои сильные и слабые стороны. Рассмотрим основные преимущества и недостатки этого стандарта. К преимуществам LTE можно отнести следующие параметры: высокая скорость передачи данных, низкая задержка, широкий охват и доступность, поддержка большого количества устройств, энергоэффективность, совместимость с предыдущими стандартами, масштабируемость. Из недостатков LTE можно выделить следующие проблемы: ограниченный спектральный ресурс, зависимость от инфраструктуры, энергопотребление базовых станций, проблемы с покрытием внутри зданий, не полное соответствие стандартам 4G, необходимость перехода к 5G, стоимость оборудования и лицензий, отсутствие полной глобальной совместимости, проблемы с безопасностью, ограниченное количество одновременных соединений, зависимость от погодных условий, устаревание по сравнению с новыми технологиями. Несмотря на наличие определенных ограничений, LTE остается важным и востребованным стандартом мобильной связи, обеспечивающим высокое быстродействие и минимальные задержки. По мере возрастания спроса пользователей и развития технологий, операторы и производители работают над совершенствованием данного стандарта и переходом к последующим поколениям мобильных коммуникаций.

1.3.4 Пропускная способность узла доступа

Пропускная способность — это количество данных, которое может передаваться через указанный путь за единицу времени. Обычно измеряется в битах в секунду (бит/с). Данные могут доставляться по физическим или логическим каналам связи или через сетевые узлы. Скорость передачи данных зависит как от характеристик физической среды передачи (медный кабель, оптическое волокно и т.д) так и от принятого способа передачи данных (технология Ethernet, FastEthernet и т.д). На скорость мобильной сети влияют следующие факторы: технологии, применяемые на базовых станциях, диапазоны и полосы частот, текущая нагрузка на сети, расстояние до базовой станции, помехи, качество транспортного канала до базовой станции, текущие настройки базовой станции, используемые абонентами мобильные устройства [8,17,18,21].

Существует несколько вариантов определения пропускной способности, рассмотрим некоторые из них [26,34]:

Максимальная пропускная способность представляет собой показатель, определяющий максимальный объем данных, который можно передать за единицу времени в идеальных условиях. Максимальная пропускная способность определяет верхний предел того, сколько данных система может обработать без потерь или задержек. Если нагрузка превышает этот порог, то возможны следующие последствия:

- потеря данных (например, потеря пакетов в сети);
- задержка передачи (увеличение времени ожидания ответа);

- перегрузка системы (снижение общей эффективности работы).

Пиковая пропускная способность — это характеристика, измеряемая на системе связи или на ее имитационной модели на малом интервале времени [34]. Эта характеристика обычно указывается производителем оборудования или провайдером интернет-услуг как максимальная теоретически возможная скорость передачи данных. Однако на практике она редко достигается из-за различных факторов, влияющих на реальную производительность.

Получаемая пропускная способность определяется через отношение объема передаваемого файла к времени передачи [34]. Этот показатель отличается от теоретической или пиковой пропускной способности, которая указывает на максимально возможное значение при идеальных условиях, т.к. это реальное количество данных, которое устройство или система получает за единицу времени в конкретных условиях работы.

Полезная пропускная способность определяется через интенсивность доставленной полезной информации, исключая служебную и повторно переданную [34]. Полезная пропускная способность исключает все служебные данные, такие как заголовки пакетов, контрольные суммы, подтверждения приема и другие виды накладных расходов, связанных с передачей данных.

Требуемая пропускная способность — это скорость передачи информации, которая необходима для пересылки заданных информационных потоков с заданным качеством [34]. Этот показатель важен для проектирования и планирования сетей, систем хранения данных и других инфраструктурных решений.

Также выделяют пропускную способность для восходящего (Uplink) и нисходящего (Downlink) каналов связи. Отличие пропускной способности между восходящим и нисходящим каналами в сетях LTE связано с тем, что эти каналы предназначены для разных типов передачи данных. Восходящий канал — это направление передачи данных от устройства пользователя к базовой станции сети оператора. В этом канале обычно передаются данные, генерируемые пользователем, такие как голосовые вызовы, сообщения, запросы веб-страниц и другие пользовательские активности. Пропускная способность этого канала ниже, чем у нисходящего, так как большинство пользователей чаще загружают информацию, чем отправляют её. Нисходящий канал — это направление передачи данных от базовой станции к устройству пользователя. В этом канале передаются данные, запрашиваемые пользователем, включая загрузку веб-страниц, видео, музыки и других медиафайлов. Пропускная способность этого канала выше, поскольку большая часть интернет-трафика приходится именно на него. В качестве примера можно привести LTE категории 4 — это категория LTE, которая поддерживает максимальную теоретическую скорость загрузки до 150 Мбит/с и скорость передачи до 50 Мбит/с. Она поддерживает агрегацию несущих до двух каналов по 20 МГц, MIMO с антеннами 2x2 и QAM до 64. Категория 4 обычно используется для смартфонов, планшетов и других мобильных устройств [87].

Категории LTE (LTE Cat) представляют собой группы устройств, классифицируемые по их возможностям обмена данными с базовыми станциями операторов. Эти категории определяются набором технических характеристик, включая количество потоков MIMO, типы модуляции и возможность агрегирования частотных полос. Устройство определен-

ной категории LTE способно поддерживать соответствующую ей максимальную скорость передачи данных. Ниже приведена сводная таблица категорий LTE (таблица 1.2).

Нижние категории LTE включают устройства с ограниченной пропускной способностью, которые чаще всего используются в устройствах Интернета вещей (IoT). Для таких устройств важнее не максимальная скорость интернета, а низкое потребление энергии, компактность модулей и надежность связи даже при слабом сигнале. IoT применяется для дистанционного управления умными домашними приборами и автоматизации производственных процессов. Спецификации 3GPP предлагают специализированные категории для IoT и межмашинного взаимодействия (M2M), такие как LTE Cat. 1, а также особые категории M1, M2, NB1 и NB2 [87].

Таблица 1.2 — Сводная таблица категорий LTE [87]

Кат. LTE	Макс. скорость загрузки Мбит/с	Макс. скорость передачи Мбит/с	Конфиг. MIMO	Тип модуляции (DL)	Тип модуляции (UL)
Cat. 0	1	1	1x1	64QAM	16QAM
Cat. 1	10,3	5,2	1x1	64QAM	16QAM
Cat. 2	51	25	2x2	64QAM	16QAM
Cat. 3	102	51	2x2	64QAM	16QAM
Cat. 4	150	51	2x2	64QAM	16QAM
Cat. 5	300	75	4x4	64QAM	64QAM
Cat. 6	301	51	2x2, 4x4	64QAM	16QAM
Cat. 7	301	102	2x2, 4x4	64QAM	16QAM
Cat. 8	2998	1497	8x8	64QAM	64QAM
Cat. 9	452	51	2x2, 4x4	64QAM	16QAM
Cat. 10	452	102	2x2, 4x4	64QAM	16QAM
Cat. 11	603	51	2x2, 4x4	64QAM, 256QAM	16QAM
Cat. 12	603	102	2x2, 4x4	64QAM, 256QAM	16QAM
Cat. 13	391	150	2x2, 4x4	256QAM	64QAM
Cat. 14	3916	9585	8x8	256QAM	64QAM
Cat. 15	749–798	226	2x2, 4x4	64QAM, 256QAM	64QAM
Cat. 16	979	105	2x2, 4x4	64QAM, 256QAM	64QAM, 256QAM
Cat. 17	25065	2119	8x8	256QAM	64QAM, 256QAM

Таблица 1.2 — Сводная таблица категорий LTE (продолжение) [87]

Кат. LTE	Макс. скорость загрузки Мбит/с	Макс. скорость передачи Мбит/с	Конфиг. MIMO	Тип модуляции (DL)	Тип модуляции (UL)
Cat. 18	1174–1206	211	2x2, 4x4, 8x8	64QAM, 256QAM	64QAM, 256QAM
Cat. 19	1566–1658	13653	2x2, 4x4, 8x8	64QAM, 256QAM	64QAM, 256QAM
Cat. 20	1948–2019	316	2x2, 4x4, 8x8	64QAM, 256QAM, 1024QAM	64QAM, 256QAM
Cat. 21	1348–1413	301	2x2, 4x4	64QAM, 256QAM	64QAM
Cat. 22	2349–2562	422	2x2, 4x4, 8x8	64QAM, 256QAM, 1024QAM	64QAM, 256QAM
Cat. 23	2695–2869	527	2x2, 4x4, 8x8	64QAM, 256QAM, 1024QAM	64QAM, 256QAM
Cat. 24	2936–3028	633	2x2, 4x4, 8x8	64QAM, 256QAM, 1024QAM	64QAM, 256QAM
Cat. 25	3132–3316	738	2x2, 4x4, 8x8	64QAM, 256QAM, 1024QAM	64QAM, 256QAM
Cat. 26	3422–3531	844	2x2, 4x4, 8x8	64QAM, 256QAM, 1024QAM	64QAM, 256QAM

Вычисление пропускной способности узла LTE. Для расчета теоретической пропускной способности базовой станции стандарта LTE в режиме дуплексирования с частотным разделением используется следующий алгоритм [1,17,18,21]:

1. Выбрать ширину канала (1.4, 3, 5, 10, 15 или 20 МГц). Более широкая полоса пропускания канала обычно обеспечивает более высокую пропускную способность, поскольку позволяет передавать больше данных одновременно.
2. На основе выбранной ширины канала рассчитываем количество ресурсных блоков. Ресурсный блок - единица ресурса передачи, состоящая из 12 поднесущих в частотной области и временного интервала (0,5 мс) во временной области. В таблице 1.3 сопоставлены ширина канала, количество ресурсных блоков и количество поднесущих.
3. Определяем номер модуляционно-кодовой схемы. Каждая модуляция имеет свою

Таблица 1.3 — Соответствие количества ресурсных блоков, ширины каналов и числа поднесущих[26]

Ширина канала МГц	Количество ресурсных блоков	Количество поднесущих
1,4	6	72
3	15	180
5	25	300
10	50	600
15	75	900
20	100	1200

пропускную способность в битах на символ. Один символ QPSK может передавать 2 бита, один символ 16QAM может передавать 4 бита, а символ 64 QAM может передавать 6 бит. Наряду с модуляцией существует термин, называемый скоростью кодирования. Скорость кодирования описывает эффективность конкретной схемы модуляции. Например, если мы говорим, что 16 QAM со скоростью кодирования 0,5, это означает, что эффективность этой модуляции составляет 50%, т.е., поскольку 16QAM может передавать 4 бита, но со скоростью кодирования 0,5, он может передавать 2 информационных бита, а остальные 2 бита используются для избыточности информации. В LTE используется другая скорость кодирования: QPSK, 16 QAM и 64QAM. Комбинация скорости модуляции и кодирования называется схемой модуляционного кодирования (MCS). На рисунке 1.9 показан индекс MCS, тип модуляции (2 для QPSK, 4 для 16QAM и 6 для 64 QAM) и размер транспортного блока (TBS). Стандарт LTE поддерживает от 0 до 28 MCS в нисходящей линии связи и от 0 до 22 MCS в восходящей линии связи.

Размер транспортного блока определяет какое количество данных (в битах) может быть передано за единицу времени (=1 мсек).

4. Рассчитываем количество ресурсных элементов. Количество ресурсных элементов в стандарте LTE определяется типом циклического префикса (Cyclic Prefix): нормальным или расширенным. Соответственно, один ресурсный блок содержит 84 ресурсных элемента при нормальном циклическом префиксе и 72 ресурса при расширенном. Рассмотрим случай с нормальным циклическим префиксом. Каждый ресурсный элемент способен нести различное количество бит в зависимости от метода модуляции: 2 бита при QPSK, 4 бита при 16QAM и 6 бит при 64QAM.

$$N_{OFDM} = N_{SFR} \cdot 2 \cdot N_{OFDMS} \cdot N_{SRB} \cdot N_{RB}, \quad (1.1)$$

где N_{SFR} – количество субкадров в FDD; N_{OFDMS} – количество OFDM символов в слоте; N_{SRB} – количество поднесущих в ресурсном блоке; N_{RB} – количество ресурсных блоков.

MCS Index	Модуляция	TBS Index
0	QPSK	0
1		1
2		2
3		3
4		4
5		5
6		6
7		7
8		8
9		9
10	16QAM	10
11		11
12		12
13		13
14		14
15		15
16		15

MCS Index	Модуляция	TBS Index
17	64QAM	15
18		16
19		17
20		18
21		19
22		20
23		21
24		22
25		23
26		24
27		25
28		26
29	QPSK	Reserved
30	16QAM	
31	64QAM	

Рисунок 1.9 — Таблица модуляции и индекса TBS [1]

Выполним расчет в соответствии с формулой (1.1):

$$N_{OFDM} = 10 \cdot 2 \cdot 7 \cdot 12 \cdot 100 = 168000 \text{ символов/сек} \quad (1.2)$$

Необходимо исключить из общего количества следующие ресурсы: - 1200 ресурсных элементов заняты первичным сигналом синхронизации; - 1200 отведены под вторичный сигнал синхронизации; - $1200 \times 4 = 4800$ элементов выделены для физического широкополосного канала; - $10 \times 100 \times 12 \times 3 = 36\,000$ ресурсов используется для целей управления.

Итого, из 168 000 ресурсных элементов 47 200 занимают служебные каналы, что соответствует доле в 28

$$V = 2 \cdot N_{OFDM} \cdot N_{SFR} \cdot N_{RB} \cdot \log(I_m), \quad (1.3)$$

где I_m - индекс модуляции.

5. Определяем количество антенн на прием и передачу (SISO - Single Input Single Output, MIMO). При переходе к MIMO 2x2 и MIMO 4x4 исходная скорость передачи умножается на 2 или на 4, соответственно.
6. Пропускная способность может быть определена следующим образом:

$$C = V \cdot N_{SRB} \cdot N_{RB}. \quad (1.4)$$

Выполним расчет в соответствии с приведенным выше алгоритмом. Выберем ширину канала 20 МГц, далее смотрим в таблицу 1.2 и находим, что для 20 МГц у нас передается 100 ресурсных блоков. При нормальном циклическом префиксе в 1 ресурсном блоке передается 168000 символов/мс. Вычтем из этого числа служебные символы, их число возьмем равным 25 % и получим 126000. Далее выбираем модуляцию 64QAM (один символ – 6 бит). Тогда пропускная способность будет составлять $12,6 \cdot 6 = 75,6$ Мбит/с для одного порта антенны. Если мы используем технологию MIMO 2x2, то скорость будет удвоена, т.е. $75,6 \cdot 2 = 151,2$ Мбит/с. В таблице 1.4 приведены расчеты скоростей для нисходящего канала.

Таблица 1.4 — Скорость передачи информации вниз, Мбит/с [1]

Тип модуляции	MIMO	1,4 МГц	3 МГц	5 МГц	10 МГц	15 МГц	20 МГц
QPSK	Single	0,94	2,34	4,01	7,99	11,83	15,84
16QAM	2x2	3,6	9,17	15,47	30,53	45,84	61,15
64QAM	2x2	8,78	22,13	36,67	73,39	110,11	151,2
64QAM	4x2	17,57	44,26	73,34	146,78	220,22	302,40

Что касается восходящего канала, то пропускная способность зависит от следующих факторов: выделенные блоки ресурсов, используемая схема модуляции, скорость кодирования и избыточность использования. В восходящей линии связи у нас есть только одна цепочка передачи на UE-конце, поэтому преимущества MIMO отсутствуют. Таким образом, при частоте 20 МГц мы можем получить максимальную скорость 100,8 Мбит/с, как показано в расчетах, приведенных выше, с одной антенной. После учета 25% резервирования полосы для служебных нужд (SRS, PUCCH, PRACH и др.) реальная скорость восходящего канала снижается до 75 Мбит/с. Тем не менее, такая производительность достигается только при условии полной совместимости клиентских устройств с технологией 64 QAM, поддерживаемой большинством базовых станций eNodeB. Хотя большинство базовых станций eNodeB способно поддерживать 64 QAM в восходящем канале, значительное число абонентских устройств ограничено поддержкой только 16 QAM в качестве максимальной модуляции, что заметно ограничивает практическую пропускную способность восходящего канала. В таблице 1.5 приведены расчеты скоростей для восходящего канала.

Достижение теоретической пропускной способности в реальных условиях практически невозможно, так как теоретические расчеты дают верхнюю границу возможной скорости.

Таблица 1.5 — Скорость передачи информации вверх, Мбит/с [1]

Тип модуляции	MIMO	1,4 МГц	3 МГц	5 МГц	10 МГц	15 МГц	20 МГц
QPSK	Single	0,94	2,34	4,01	7,99	11,83	15,84
16QAM	Single	1,8	4,58	7,74	15,26	22,92	30,58
64QAM	Single	4,39	11,06	18,34	36,7	55,06	75,38

Реальные показатели зависят от множества факторов, включая загрузку сети, уровень сигнала, помехи и другие условия. Для сравнения в таблице 1.6 приведены реальные скорости для нисходящего канала связи для стандарта LTE.

Таблица 1.6 — Скорость передачи информации вниз, Мбит/с [26]

Тип модуляции	MIMO	1,4 МГц	3 МГц	5 МГц	10 МГц	15 МГц	20 МГц
$QPSK^{1/2}$	Single	0,85	2,21	3,71	7,46	11,21	14,96
$16QAM^{1/2}$	2x2	3,35	8,53	14,29	28,69	43,09	57,49
$16QAM^{3/4}$	2x2	5,02	12,79	21,42	43,03	64,63	86,23
16QAM1	2x2	6,69	17,06	28,58	57,40	86,18	114,98
$64QAM^{1/2}$	2x2	5,02	12,79	21,43	43,03	64,63	86,23
$64QAM^{3/4}$	2x2	7,53	19,19	32,15	64,55	96,95	129,35
$64QAM^{9/10}$	2x2	9,03	23,03	38,58	77,46	116,34	155,22
64QAM1	2x2	10,04	25,59	42,87	86,07	129,27	172,47
64QAM1	4x2	19,09	48,47	81,11	162,71	244,31	325,91

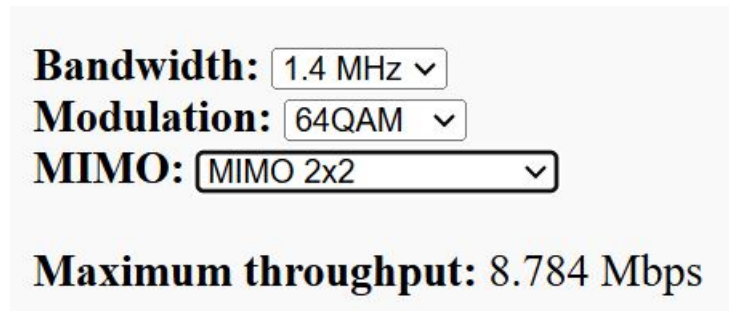
В таблице 1.7 приведены реальные скорости для восходящего канала связи для стандарта LTE.

Таблица 1.7 — Скорость передачи информации вверх, Мбит/с [26]

Тип модуляции	MIMO	1,4 МГц	3 МГц	5 МГц	10 МГц	15 МГц	20 МГц
$QPSK^{1/2}$	Single	0,85	2,21	3,71	7,46	11,21	14,96
$16QAM^{1/2}$	2x2	3,35	8,53	14,29	28,69	43,09	57,49
$16QAM^{3/4}$	2x2	5,02	12,79	21,42	43,03	64,63	86,23
16QAM1	2x2	6,69	17,06	28,58	57,40	86,18	114,98
$64QAM^{1/2}$	2x2	5,02	12,79	21,43	43,03	64,63	86,23
$64QAM^{3/4}$	2x2	7,53	19,19	32,15	64,55	96,95	129,35
$64QAM^{9/10}$	2x2	9,03	23,03	38,58	77,46	116,34	155,22
64QAM1	2x2	10,04	25,59	42,87	86,07	129,27	172,47
64QAM1	4x2	19,09	48,47	81,11	162,71	244,31	325,91

Калькуляторы для расчета пропускной способности

В сети можно найти огромное количество калькуляторов для расчета пропускной способности LTE и выше, в которых можно выполнить расчеты с разной степенью точности. Один из таких калькуляторов приведен на рисунке 1.10.



Bandwidth: 1.4 MHz ▾

Modulation: 64QAM ▾

MIMO: MIMO 2x2 ▾

Maximum throughput: 8.784 Mbps

Рисунок 1.10 — Пример калькулятора[1]

Приведенный пример калькулятора позволяет оценить максимальную пропускную способность сети LTE на основе ряда используемых параметров: ширина канала, тип модуляции, использование MIMO, потери на заголовки.

1.4 Концепции и бизнес-модели использования узлов доступа

Концепции и модели использования мультисервисных узлов доступа включают различные аспекты и возможности, предлагаемые этой технологией. Одной из ключевых особенностей является использование виртуализации сетевых функций и программно-определяемых сетей, что позволяет создавать гибкие и масштабируемые инфраструктуры. Ниже приведены некоторые из них.

Виртуализация сетевых функций (Network Function Virtualization, NFV) представляет собой технологию, которая использует виртуализацию для разделения аппаратного и программного обеспечения сетевых функций. Вместо того чтобы использовать специализированные физические устройства для каждой сетевой функции, NFV позволяет запускать эти функции в виде виртуальных машин на общих серверах [48].

Управление ресурсами в сетях пятого поколения включает в себя комплекс мероприятий по мониторингу и контролю использования сетевых ресурсов, таких как полоса пропускания, вычислительные мощности и память. Данный процесс направлен на эффективное распределение и использование имеющихся ресурсов, а также на предотвращение перегрузки сети. Управление ресурсами предполагает анализ данных о текущей нагрузке, прогнозирование будущих потребностей и автоматизированное регулирование параметров сети в соответствии с изменениями нагрузки. Это позволяет операторам мобильной связи поддерживать высокий уровень обслуживания и оптимизировать работу сети. Основными этапами процесса управления ресурсами являются сбор данных, их анализ, прогнозирование, автоматизация и отчетность. Важными преимуществами данного подхода являются

повышение эффективности использования ресурсов, стабильности и надежности сети, а также улучшение пользовательского опыта за счет более стабильных соединений и меньших задержек [52,56].

Гибридные облачные среды представляют собой сочетание публичного облака и частного облака или локальной инфраструктуры, что позволяет организациям воспользоваться преимуществами обеих моделей для повышения эффективности, гибкости и безопасности своих операций. Этот подход отличается от традиционного использования исключительно одного типа облачной среды, предоставляя бизнесу возможность выбирать оптимальную среду для размещения данных и приложений в зависимости от их специфических требований. Среди преимуществ гибридных облачных сред можно выделить максимальную гибкость, сокращение затрат, усиленную безопасность, резервирование и отказоустойчивость, соблюдение нормативных требований, а также легкую миграцию между различными облачными платформами. В контексте практического применения гибридные облака используются для хранения данных, размещения приложений, разработки и тестирования новых продуктов, а также для анализа больших объемов данных. Они способствуют увеличению роста и развития организаций за счет предоставления большего контроля над данными и приложениями [22].

Программно-определяемые сети (Software Defined Networking, SDN) представляют собой архитектуру, в которой управление сетью отделяется от её физического оборудования. Центральный логический контроллер управляет потоком трафика и политиками безопасности, основываясь на общих правилах, а не на конфигурации каждого отдельного сетевого элемента. Основные компоненты SDN включают контроллер, платформу управления, драйверы и абстракции, а также виртуальные сетевые функции (VNF). Преимущества SDN включают централизованное управление, гибкость и динамичность, прозрачность, снижение затрат и улучшенную безопасность. SDN широко применяется в облачных вычислениях, Интернете вещей, центрах обработки данных и у телеком-провайдеров, помогая им быстрее и эффективнее разворачивать новые сервисы и адаптироваться к изменениям спроса [48,56,60].

1.4.1 Network Slicing

Network slicing (нарезка сети) - это технология, позволяющая разделять сетевые ресурсы на логически изолированные сегменты, известные как слои, срезы или слайсы. Основным принципом данной технологии являются в создании независимых виртуализированных сетевых сегментов, каждый из которых функционирует как отдельная полноценная сеть. Эти сегменты работают параллельно друг другу, используя одни и те же физические ресурсы сети (инфраструктуру). Это делает возможным гибкое распределение ресурсов между различными приложениями и пользователями, обеспечивая оптимальное качество обслуживания для каждого конкретного случая. На практике Network Slicing реализуется через использование виртуализации сетевых функций и программного обеспечения для определения сетевого трафика. Виртуализации сетевых функций позволяет переносить функции сетевого оборудования в программное обеспечение, которое работает

на стандартных серверах, а SDN обеспечивает централизованное управление трафиком и настройкой сети. Это позволяет операторам связи предоставлять дифференцированный сервис и качество обслуживания для различных типов пользователей и приложений [5,41,50,73,82].

Основные характеристики Network Slicing включают в себя логическую изоляцию, управление слоями, виртуализацию ресурсов и оптимизацию использования ресурсов. Рассмотрим каждую из перечисленных характеристик подробнее [5,78,131].

Логическая изоляция. Каждый слайс представляет собой отдельную, изолированную среду, которая работает независимо от других слайсов. Это означает, что проблемы или сбои в одном слайсе не влияют на работу других. Изоляция обеспечивается за счёт использования виртуальных компонентов, которые отделены друг от друга на уровне операционной системы и гипервизора. Каждому слайсу выделяются собственные ресурсы (процессор, память, хранилище), что предотвращает конфликты и утечки данных между ними. Изоляция важна для повышения безопасности и устойчивости сети. Она защищает от распространения угроз, таких как DDoS-атаки или ошибки в коде, и гарантирует, что критически важные службы будут продолжать функционировать даже в случае проблем в других частях сети.

Управление слоями. Настройка слайсов выполняется через центральную систему управления, которая контролирует весь жизненный цикл слайса: от создания до удаления. Параметры каждого слайса задаются индивидуально, исходя из требований клиента или приложения. Каждому слайсу можно присвоить уникальные параметры и настройки, соответствующие конкретным требованиям приложений и сервисов. Это включает в себя уровень качества обслуживания, задержку, пропускную способность и другие ключевые показатели производительности.

Виртуализация ресурсов. Физическая сеть делится на несколько логических частей с использованием виртуальных машин или контейнеров. Эти виртуальные компоненты заменяют традиционные аппаратные устройства, такие как маршрутизаторы и коммутаторы. SDN управляет трафиком, распределяя его по различным слайсам, а NFV предоставляет необходимые сетевые функции в виде программного обеспечения. Виртуализация позволяет значительно снизить стоимость развертывания и эксплуатации сети, поскольку она устраняет необходимость в покупке дорогостоящего специализированного оборудования. Более того, такая архитектура обеспечивает большую гибкость и возможность быстрого реагирования на изменения в требованиях к услугам.

Оптимизация использования ресурсов. Слайсы создаются и конфигурируются таким образом, чтобы минимизировать потери и максимально эффективно использовать доступную инфраструктуру. Для этого система управления сетью непрерывно мониторит состояние каждого слайса и автоматически перераспределяет ресурсы (пропускную способность, вычислительную мощность, хранилища) в зависимости от текущей нагрузки. Например, если один слайс испытывает пиковый спрос, его ресурсы могут быть увеличены за счет других слайсов, которые в этот момент менее загружены. За счет оптимизации потребления ресурсов удастся сократить ненужные резервы и уменьшить энергозатраты. Время, необходимое для перераспределения ресурсов и изменения конфигурации слай-

сов в технологии Network Slicing, зависит от ряда факторов, таких как сложность сети, используемая инфраструктура и степень автоматизации процессов. Тем не менее, благодаря современным технологиям виртуализации и программно-конфигурируемых сетей, этот процесс стал значительно быстрее и эффективнее по сравнению с традиционными методами управления сетью. Оценочные временные рамки:

- Создание нового слайса: Обычно занимает от нескольких минут до часа, в зависимости от сложности конфигурации и масштаба сети.
- Изменение параметров существующего слайса: Может занимать от нескольких секунд до минуты, если изменения касаются лишь настроек качества обслуживания или перераспределения ресурсов.
- Полное перераспределение ресурсов между несколькими слайсами: Может занять от нескольких минут до нескольких часов, в зависимости от объема перераспределяемых ресурсов и сложности сети.

Важно отметить, что конкретные цифры зависят от конкретной реализации технологии и инфраструктуры, используемых оператором связи.

Архитектура технологии Network Slicing состоит из нескольких ключевых блоков, которые совместно обеспечивают эффективное управление и функционирование виртуализированных сетевых срезов. Основные блоки: блок реализации сетевого среза, блок управления и настройки сетевого среза. Первый блок отвечает непосредственно за выполнение всех операций, связанных с созданием и функционированием конкретного сетевого среза. Он содержит механизмы, которые позволяют реализовать и поддерживать работоспособность отдельных срезов. Второй блок играет роль центра управления всей системой сетевого слайсинга. Он контролирует и настраивает функции и процессы, необходимые для координации работы различных срезов. Основная задача блока — гарантировать эффективное сосуществование и координацию между несколькими сетевыми срезами. Эти два блока тесно связаны и обеспечивают целостное управление сетью, позволяя оптимизировать ресурсы и обеспечивать высокий уровень качества обслуживания для различных типов трафика и пользователей [92,95,131]. Архитектура нарезки сети представлена на рисунке 1.11

Преимущества и недостатки технологии Network Slicing. Технология Network Slicing имеет следующие преимущества: гибкая настройка под различные сценарии использования, оптимизированное использование ресурсов, повышенная надежность и безопасность, поддержка различных бизнес-моделей, быстрая адаптация к изменениям, улучшение качества обслуживания.

Недостатками технологии Network Slicing являются: высокие начальные инвестиции, сложность управления, зависимость от инфраструктуры и совместимости, проблемы с обеспечением QoS, требовательность к вычислительным ресурсам.

Бизнес-модели. С точки зрения бизнес-модели, каждый сетевой срез администрируется оператором мобильной виртуальной сети (MVNO, Mobile Virtual Network Operator). Поставщик инфраструктуры (владелец телекоммуникационной инфраструктуры) сдает в

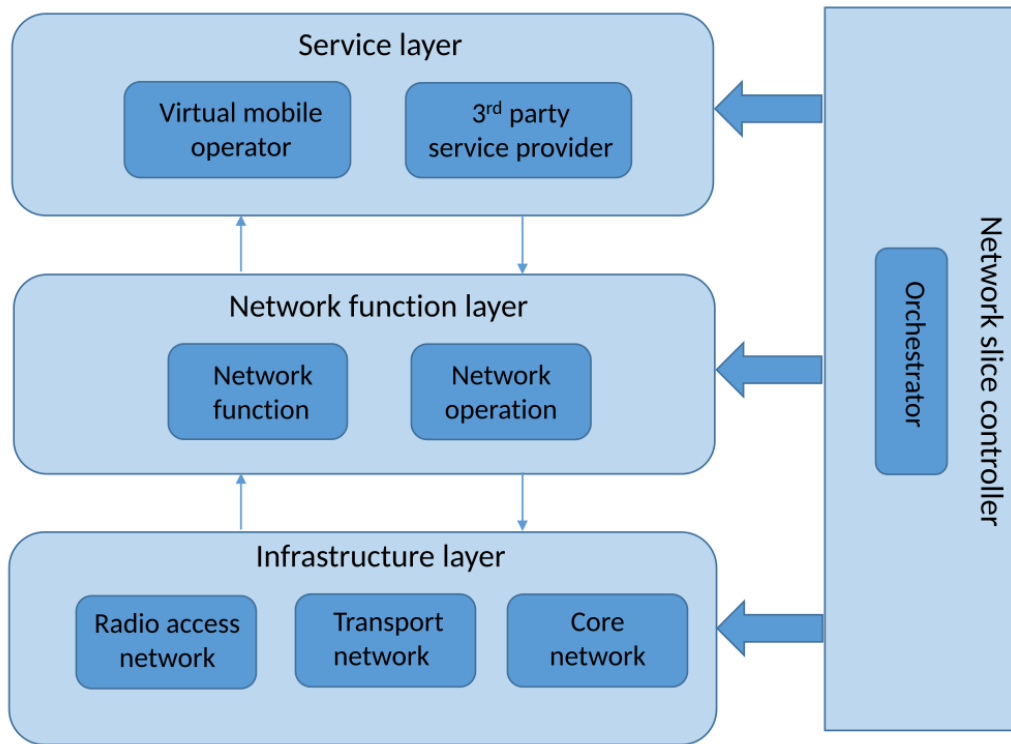


Рисунок 1.11 — Архитектура нарезки сети [92]

аренду свои физические ресурсы операторам мобильной виртуальной сети, которые совместно используют базовую физическую сеть. В зависимости от доступности назначенных ресурсов MVNO может автономно развертывать несколько сетевых срезов, которые настраиваются на различные приложения, предоставляемые его собственным пользователям [6,79]. Примером такой архитектуры служит схема сети представленная на рисунке 1.12, где оператор телекоммуникаций предоставляет свои радиоресурсы в аренду виртуальным операторам, не имеющим собственных физических активов беспроводной сети, позволяя совместно использовать единую инфраструктуру. Между виртуальным оператором и основным провайдером заключается договор об уровне обслуживания, определяющий пределы сегментации сети исходя из показателей скорости передачи данных.

1.4.2 Интернет вещей

Интернет вещей (IoT, Internet of Things) представляет собой концепцию распределенной вычислительной инфраструктуры, объединяющей физический мир с цифровым посредством встраивания сенсоров, микропроцессоров и коммуникационных модулей в различные устройства и предметы повседневного использования. Этот процесс создает сеть взаимосвязанных объектов, способных автономно собирать, анализировать и передавать данные, а также взаимодействовать друг с другом и внешними системами без необходимости прямого вмешательства человека. На рисунке 1.13 представлены типы устройств IoT и их взаимодействие.

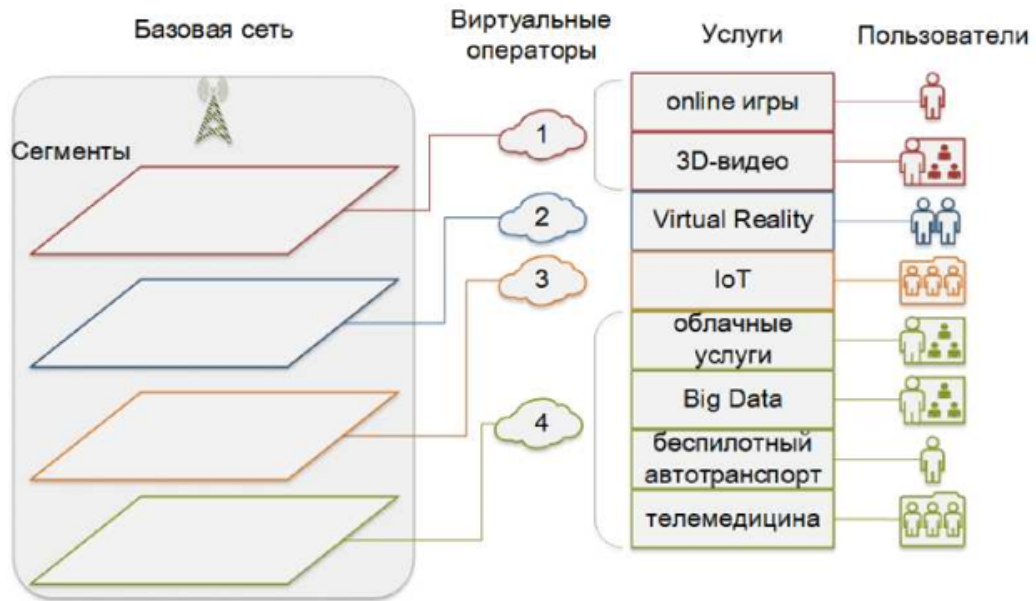


Рисунок 1.12 — Нарезка ресурсов по виртуальным операторам [6]

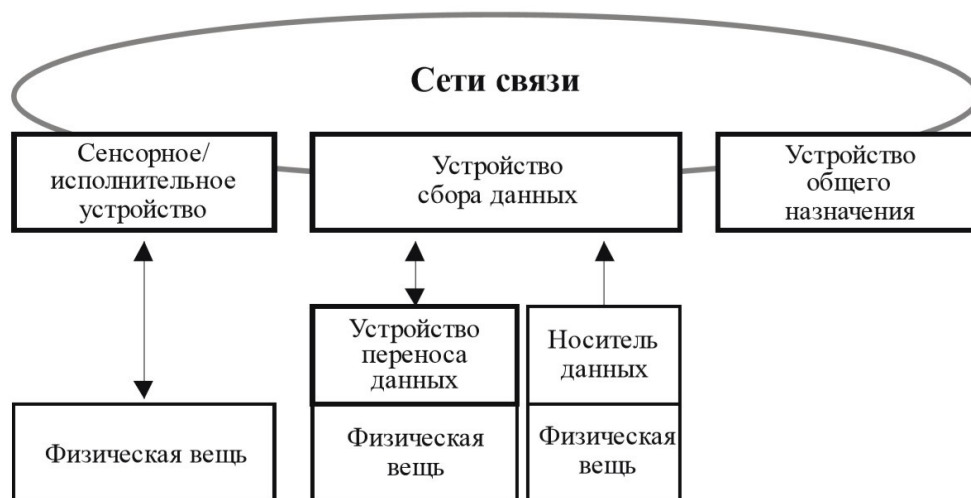


Рисунок 1.13 — Типы устройств IoT и их взаимодействие

Архитектура Интернета Вещей в международных стандартах описывается как многоуровневая система, включающая несколько ключевых компонентов и уровней взаимодействия. Стандартизация играет важную роль в обеспечении надежности и безопасности устройств и систем IoT. Рассмотрим основные подходы к описанию архитектуры IoT в стандартах. Рекомендации Международного союза электросвязи описывают архитектуру IoT как многослойную систему, состоящую из нескольких уровней:

1. Многоуровневая архитектура. Сюда входят: физический уровень, сетевой уровень,

уровень приложений. Физический уровень включает в себя датчики, исполнительные механизмы и другие физические устройства, которые взаимодействуют с окружающим миром. Сетевой обеспечивает связь между устройствами и системами, используя различные протоколы, такие как Wi-Fi, Bluetooth, LoRaWAN и др. Уровень приложений содержит приложения и сервисы, которые обрабатывают и анализируют данные, полученные от устройств.

2. Функциональные компоненты. Сюда входят: устройства, шлюзы, облачные платформы. Под устройствами понимают физические объекты, оснащенные датчиками и коммуникационными модулями. Шлюзы — это узлы, которые связывают устройства с внешними сетями и платформами. Облачные платформы представляют собой централизованные системы для хранения, обработки и анализа данных.

В рекомендации Y.2060 содержится эталонная модель IoT, которая служит основой для стандартизации. Эталонная архитектура даёт разработчикам понимание того, какие функции нужны в IoT и как они взаимодействуют [36]. На рисунке 1.14 представлена эталонная модель IoT.



Рисунок 1.14 — Эталонная модель IoT

Технология IoT может применяться в следующих областях:

- Умный дом: управление освещением, отоплением, безопасностью и другими функциями дома через мобильные устройства.
- Промышленность: мониторинг производственных процессов, автоматизация оборудования, прогнозирование неисправностей.

- **Здравоохранение:** дистанционный мониторинг состояния пациентов, носимые медицинские устройства, автоматическое оповещение врачей при критических состояниях.
- **Транспорт и логистика:** отслеживание грузов, оптимизация маршрутов, мониторинг состояния транспортных средств.
- **Городская инфраструктура:** «умные» города с автоматизированным управлением уличным освещением, парковками, транспортом и утилизацией отходов.

Преимущества и недостатки IoT. Преимущества Интернета вещей включают в себя возможность повышения эффективности и автоматизации процессов, улучшения качества жизни за счет удобства использования бытовых приборов и сервисов, оптимизации затрат благодаря возможности мониторинга и предсказания поломок оборудования, а также возможность сбора больших объемов данных для аналитики и принятия обоснованных решений. В контексте промышленного применения (IIoT), IoT позволяет производить и оптимизировать новые продукты, быстро реагировать на потребности в продуктах, осуществлять цифровое управление технологическими процессами, оптимизировать безопасность и охрану оборудования, управлять активами с помощью прогнозного обслуживания, обеспечивать максимальную надежность, интегрировать промышленные системы управления с интеллектуальными сетями для оптимизации энергопотребления, а также использоваться для процессов индустриализации строительства. В машиностроении интеграция IoT-технологий позволяет собирать данные о состоянии и работе оборудования в реальном времени, предотвращать отказы оборудования и улучшать его производительность. Недостатки Интернета вещей включают в себя: безопасность и защита данных, совместимость и стандартизация, масштабируемость, управление данными, стоимость внедрения и эксплуатации, надёжность и устойчивость.

Взаимодействие устройств IoT и LTE/5G. Устройства Интернета вещей могут взаимодействовать с сетью LTE различными способами. Один из них — использование технологии VoLTE (голос по LTE), которая позволяет передавать голосовые вызовы в сети LTE, ускоряя процесс осуществления голосового вызова. Другой способ — Circuit-switched fallback (CSFB), при котором LTE обеспечивает только услуги передачи данных, а для совершения или принятия голосового вызова устройство возвращается к сети с коммутацией каналов, такой как GSM или UMTS. Кроме того, современные устройства IoT могут работать на базе различных беспроводных технологий, включая стандарты сотовой связи, такие как 5G, LTE, 3G и 2G. Это позволяет им подключаться к интернету и обмениваться данными через мобильные сети [15,48,71,74,79,80]. Ниже приведены основные сценарии работы IoT с мобильными сетями:

- **EC-GSM-IoT:** расшифровывается как Extended Coverage GSM Internet of Things («Расширенное покрытие GSM для Интернета вещей»). Представляет собой модернизированную версию стандарта GSM, специально адаптированную для работы Интернета вещей. Среди изменений можно отметить улучшение зоны покрытия путём

доработки физического уровня, упрощение конструкции устройств за счёт оптимизации протоколов, повышение эффективности энергопотребления и увеличение долговечности аккумуляторов, а также усиление защиты данных.

- **LTE Category-M1:** также известен как LTE-Machine-Type-Communication (LTE-M). Это модификация LTE, предназначенная для работы в низкоэнергетических широкомасштабных сетях (LPWA), ориентированных на подключение устройств Интернета вещей. Категория M1 обладает рядом преимуществ по сравнению с классическим LTE: меньшая полоса пропускания, сниженные требования к мощности и энергопотреблению, а также оптимизация для отправки небольшого объёма данных с умеренными требованиями к скорости (обычно от 200 до 400 кбит/с).
- **NB-IoT:** это сокращение от Narrowband Internet of Things («Интернет вещей с узкой полосой»). Эта технология основана на стандарте LTE и предназначена для работы с большим количеством устройств с низкими требованиями к пропускной способности. Она обеспечивает ряд преимуществ, включая увеличенную проникающую способность сигнала внутри помещений, поддержку множества маломощных устройств, снижение стоимости оборудования, экономию энергии и оптимизацию сетевой инфраструктуры. Задержка передачи данных составляет примерно 10 секунд. Технология также отличается простотой реализации, поддержкой высокоуровневых протоколов и улучшенной системой сигнализации. NB-IoT поддерживает различные сценарии применения в области IoT, предлагая разнообразные решения и услуги. Важно отметить, что эта технология обратно совместима с предыдущими стандартами сотовой связи, такими как LTE и GSM, и работает на тех же частотах, что позволяет избежать необходимости в новой инфраструктуре вроде базовых станций.

1.4.3 Сервисы систем наблюдений

Системы видеонаблюдения (VSS, Video Surveillance System) широко используются в крупных системах для использования в стратегических местах, таких как домашняя безопасность, общественный транспорт, банки, центры банкоматов, городские центры, аэропорты и дороги общего пользования, и играют важную роль в защите критически важных инфраструктур. Традиционно системы видеонаблюдения строят двумя способами: с установкой серверов для хранения и анализа видеоизображений на самом предприятии (например, DVR/NVR) либо с использованием программного обеспечения для управления видео. В первом случае видео накапливается и анализируется на серверах, расположенных на предприятии, с возможностью просмотра нужных фрагментов через интернет уполномоченным персоналом. Во втором случае применяется облачная структура для хранения и управления. В этом случае сетевая инфраструктура предоставляется оператором либо виртуальным оператором, на самом предприятии устанавливаются только сами камеры видеонаблюдения. Таким образом, расходы на развёртывание системы видеонаблюдения значительно снижаются. Виртуальный оператор сотовой связи — это компания, которая предоставляет услуги мобильной связи, используя сетевые ресурсы, арендованные

у другого оператора (или нескольких операторов). Виртуальные операторы не обладают собственной физической инфраструктурой, но продают услуги под своим брендом. Они отличаются зависимостью от базового оператора, акцентом на маркетинговых стратегиях, технической поддержке, ценовой политике и функциональных особенностях. Примеры виртуальных операторов включают Tele2, TuGo и Mint Mobile. Виртуальные операторы способствуют расширению выбора для потребителей и стимулированию конкуренции на рынке мобильных услуг. На рисунке 1.15 представлена схема сети виртуального оператора, предоставляющего услуги видеонаблюдения.

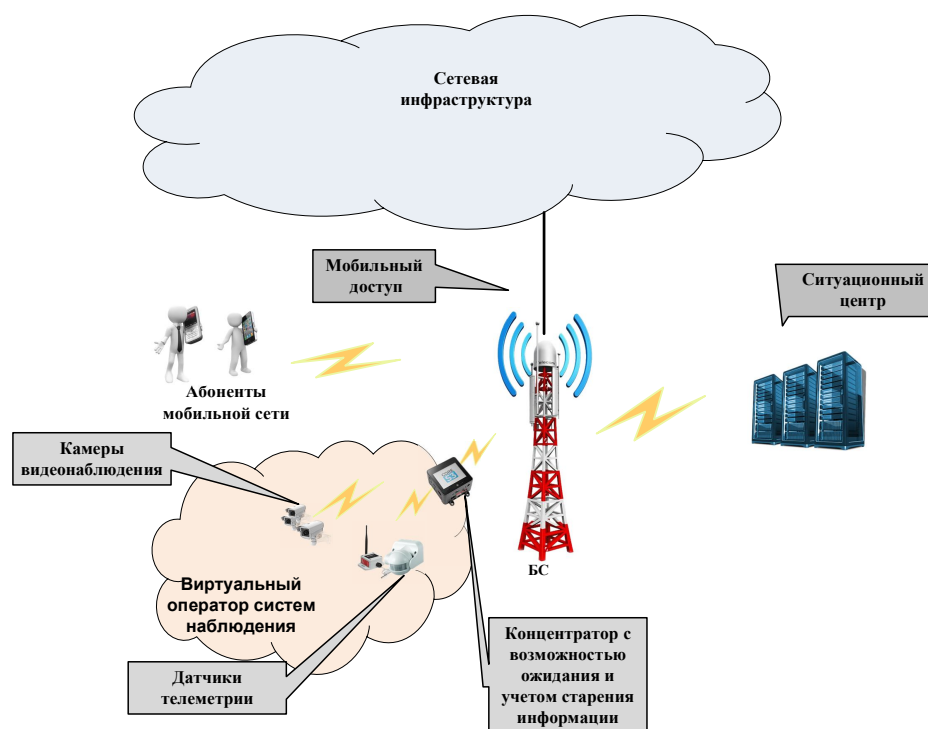


Рисунок 1.15 — Сеть оператора систем видеонаблюдения

Классификация систем видеонаблюдения осуществляется в зависимости от способа соединения камер наблюдения с аналитическими центрами и методами передачи видеосигнала. Выделяют три вида систем: аналоговые, цифровые и сетевые. В аналоговых системах видеоконтент передается на короткие расстояния по коаксиальному кабелю, а для удалённого размещения точек приема сигнала применяется волоконно-оптическая линия связи. В цифровых же системах видеонаблюдение построено на передаче данных через IP-сеть, что позволяет интегрироваться с существующей локальной вычислительной сетью предприятия и получать преимущества в обмене мультимедийными файлами и аналитической обработкой непосредственно в цифровом формате [114]. Использование цифровых систем видеонаблюдения даёт значительные преимущества по сравнению с аналоговым оборудованием: удобство управления камерами, гибкая масштабируемость, качественное воспроизведение сигнала и эффективные методы архивации данных. К особенностям цифрового режима относятся применение кодеков MPEG-4 и H.264 для видеопередачи, возможность удаленного наблюдения за видео и изображениями и другие функции. Третий режим

— сетевое видеонаблюдение (IP-VSS) — подразумевает применение сетевых камер, подключающихся непосредственно к IP-сети. Такой подход позволяет организовать полный цикл действий с видеозаписями: сбор, мониторинг, обмен и хранение визуального материала [115]. Далее развитие данной технологии идёт в сторону внедрения беспроводных сенсорных сетей, состоящих из датчиков, собирающих разнообразные сведения в разных природных условиях. Собранный информация направляется на специальные базовые станции, обладающие большими возможностями вычислений и коммуникации, откуда далее поступает конечному потребителю.

В беспроводных мультимедийных сенсорных сетях задействовано большое количество специализированных узлов, оборудованных видеокамерами, датчиками и передающими устройствами. Они собирают данные различного характера (визуальные, звуковые, числовые) и отправляют их на центральные сервера через промежуточные базовые станции. Подобные решения широко внедряются в различных областях жизни общества: аэропортах, больницах, офисных зданиях, общественных пространствах, военных объектах и других местах массового пребывания людей, повышая безопасность и защищённость [116].

Для реализации всех перечисленных сценариев необходимы специализированные камеры, выбор которых обусловлен характеристиками конкретной системы видеонаблюдения и особенностями применяемых сетей. Наиболее распространены следующие разновидности камер: поворотные PTZ-камеры, IP-камеры, обычные камеры видеонаблюдения, модели с ночным видением, датчики движения и др. Их главная задача — контролировать ситуацию в зоне ответственности и своевременно сообщать обо всех событиях для последующего анализа (распознавания лиц, отслеживания перемещения, выявления скоплений людей и т.п.) [117]. Информация, собранная такими камерами, проходит дальнейшую обработку: изначально полученные кадры подвергаются сжатию с применением общепринятых алгоритмов кодирования (JPEG, MPEG, H.264, H.265 и аналогичных). Особняком стоит алгоритм H.264, получивший наибольшее распространение в современных системах видеонаблюдения, так как он позволяет сохранять высокое качество изображения при низком объёме данных. Сравнение популярных алгоритмов кодирования приведено в таблице 1.8.

Таблица 1.8 — Кодеки для различных видеоформатов с точки зрения качества и битрейта [118]

Битрейт	Приложение	Качество	Кодек
400 Кбит/с	VoD (например rutube)	240p / SD	H.264
3,5 Мбит/с	SD-телевидение	480p / SD	MPEG-2
3,8 Мбит/с	VoD (например vk видео)	720p / HD	H.264
6,8 Мбит/с	VoD (например rutube)	1080 p / HD	H.264
8-15 Мбит/с	HDTV	1080 p / HD	MPEG-4 AVC
40 Мбит/с	Диск Blue-ray	1080 p / HD	MPEG-4 AVC VC-1

1.5 Обзор публикаций по тематике исследования

Модели мультисервисных узлов доступа являются предметом интенсивных исследований начиная с конца прошлого века. Возникший интерес обусловлен развитием телекоммуникационных технологий, которые обеспечили техническую возможность предоставления широкого спектра услуг связи [5,22,24,26,57,51,79]. Необходимость решения задач планирования ресурса передачи информации узлов доступа, требуемого по нагрузке и качеству обслуживания поступающих заявок привела к появлению мультисервисной модели Эрланга, которая была построена и исследована в работах [2,12,34,57]. Результатом изучения модели явился рекурсивный алгоритм оценки основных показателей качества обслуживания поступающей нагрузки: доли потерянных заявок и среднего объема, занятой пропускной способности узла доступа. Причем эти характеристики были рассчитаны для всех потоков. Дальнейшие исследования были направлены на уточнение структуры входного потока и процедур допуска поступающих заявок к занятию ресурса передачи информации. Были исследованы несколько моделей мультисервисных узлов с ограниченным доступом [2,12,34,57]. В их числе модель, где прием заявки основан на анализе данных о числе заявок рассматриваемого потока, находящихся на обслуживании. Модель позволяет создать условия для дифференцированного обслуживания абонентов и учесть ограничение по приёму заявок, обусловленное иерархическим принципом построения сетей доступа [34]. В моделях другого типа прием заявки основан на анализе занятости ресурса передачи информации. Процедура резервирования позволяет создать условия для дифференцированного обслуживания абонентов и обеспечивает пользователя услуг связи гарантированными показателями QoS [34].

Здесь также необходимо упомянуть модель мультисервисного узла доступа с ограниченным числом пользователей услуг связи [16,34]. В моделях этого типа интервалы времени между поступлением заявок имеют экспоненциальное распределение с параметром, зависящим от числа обслуживаемых заявок данного потока. Эту модель также называют мультисервисной моделью Энгсета [16,34]. Для перечисленных выше моделей характерно наличие эффективного алгоритма оценки основных показателей обслуживания заявок. Говоря об эффективности следует отметить, что наличие этого свойства предполагает возможность оценки основных характеристик обслуживания заявок с использованием стационарных вероятностей макросостояний модели, а не микросостояний, чьи вероятности рассчитываются из решения системы уравнений равновесия. Также необходимо отметить, что перечисленные модели предназначены для анализа процесса обслуживания трафика сервисов реального времени или трафика передачи данных, но на условиях обслуживания трафика сервисов реального времени, т.е. с предоставлением фиксированной на все время соединения скорости передачи информации.

Дальнейшее развитие телекоммуникаций привело к расширению перечня предоставляемых телекоммуникационных сервисов. Наряду с передачей голосовых сообщений появилась необходимость пересылки файлов, являющихся результатом реализации отдельных положений концепции Интернета Вещей [79,80]. Информационные потоки, порожденные действующими и перспективными коммуникационными приложениями, характеризуются

разными требованиями к скорости передачи, импульсным характером поступления сообщений, разной степенью терпимости к задержке, наличием или отсутствием свойства эластичности и т.д. Перечисленные особенности формирования и обслуживания поступающих заявок необходимо учитывать при решении задач планирования требуемой величины ресурса передачи информации и создания условий по их дифференцированному обслуживанию. В противном случае, как показали результаты измерений и моделирования телекоммуникационных сетей, возникают проблемы с качеством предоставления услуг.

Появился ряд исследований, где рассмотренные выше особенности формирования поступающих заявок и распределения ресурса между ними в процессе обслуживания были включены в математическое описание модели [4,31–35,43–46]. Однако за редким исключением [34,57], которое относится к оценке характеристик группового поступления файлов без предоставления возможности ожидания начала обслуживания в ситуации блокировки, где эта задача решается с использованием рекурсии, а также ряда исследований [16,34], где для оценки характеристик используются рекурсии специального вида, следующие из наличия свойства мультипликативности, вычисление показателей обслуживания заявок выполняется с помощью решения системы уравнений равновесия итерационным методом Гаусса—Зейделя [16,34,31–35,43–46] или матричным методом [4,65,84]. Реализация этих алгоритмов позволяет решать системы уравнений равновесия с числом состояний до нескольких миллионов. Однако этого мало, если рассматривать задачи оценки необходимого объема ресурса, которые решаются методом перебора вариантов. Все это затрудняет использование моделей узлов доступа в практических приложениях. Ситуация еще более усложняется, если ставится задача изучения совокупного влияния перечисленных выше факторов на решение задач планирования необходимого ресурса узлов доступа и разработку процедур дифференцированного обслуживания поступающих потоков гетерогенного трафика. Здесь получены отдельные результаты, которые можно использовать при рассмотрении теоретических вопросов реализации концепции Network Slicing [4,6,51,64,69,71,72], когда поступающие информационные потоки группируются по близости параметров их формирования и обслуживания. Если исключить использование имитационного моделирования для оценки характеристик обслуживания заявок по причинам большой трудоемкости, то единственным реальным способом расчета моделей узлов доступа остается использование приближенных методов. Часть из них была разработана и исследовалась в работах [34]. Отдельные задачи, относящиеся к построению и исследованию моделей формирования и обслуживания заявок в беспроводных узлах доступа рассматривались в диссертационных работах: С.Д. Андреева ¹, В.О. Бегишева ², Е.А. Кучерявого ³ и др. Ранее выполненные исследования по заявленной тематике либо имеют чересчур завышенную теоретическую

¹Андреев, С.Д. Разработка и исследование моделей множественного доступа и алгоритмов управления потоками трафика для гетерогенных беспроводных сетей: дис. ... доктора технических наук: 05.12.13 / Андреев Сергей Дмитриевич. — ГУАП. — Москва, 2018.

²Бегишев, В.О. Особенности построения вероятностных моделей взаимодействия устройств: дис. ... кандидата технических наук : 05.13.17 /Бегишев Вячеслав Олегович. — РУДН. — Москва, 2019.

³Кучерявый, Е.А. Разработка и исследование комплекса моделей и методов распределения ресурсов в беспроводных гетерогенных сетях связи: дис. ... доктора технических наук: 05.12.13 /Кучерявый Евгений Андреевич; — МТУСИ. — Москва, 2018.

направленность, что затрудняет использование полученных результатов на практике, либо являются инженерными разработками, базирующимися на эмпирических результатах, что делает невозможным их применение при изменении параметров решаемой задачи. В рамках данного исследования предлагается использовать многомерные марковские процессы для построения модели распределения ресурса и технику декомпозиции модели на отдельные модули для построения приближенных алгоритмов оценки характеристик обслуживания поступающих потоков трафика.

1.6 Постановка задачи диссертационного исследования

Для обоснованного решения задачи планирования пропускной способности беспроводных узлов доступа и анализа процедур дифференцированного обслуживания неоднородного трафика необходимо построить новые математические модели мультисервисных узлов доступа, в которых учитываются особенности формирования информационной нагрузки, и разработать на их основе точные и приближенные алгоритмы оценки характеристик качества обслуживания поступающих запросов.

Для достижения указанной цели необходимо решить следующие научные задачи:

- разработать и проанализировать модель совместной обработки в беспроводном узле доступа приоритетного трафика онлайн-услуг и файловых данных с переменной скоростью передачи, учитывающую импульсный характер поступления файлов, а также возможность задержек и снижения актуальности передаваемой информации;
- на основе предложенной модели оценить ключевые параметры качества обслуживания поступающих запросов;
- разработать точные и приближенные методы вычисления полученных характеристик;
- сформировать практические рекомендации по эффективному распределению пропускной способности узла доступа между разнородными потоками трафика.

1.7 Выводы по результатам первого раздела

1. Выполнено исследование основных структурных параметров сетей мобильной связи на базе стандарта LTE. Важнейшей характеристикой мультисервисного узла доступа является пропускная способность, играющая существенную роль при решении задачи обеспечения заданного качества обслуживания поступающих заявок. Рассмотрена зависимость этой характеристики от типа соединения (Uplink, Downlink), величины выделенной полосы передачи, используемых технологий доступа FDD и TDD, типа модуляции, алгоритма кодирования, использования антенных технологий MIMO, загрузки узла и типов предоставляемых сервисов и т.д. Приведены значения достигаемых скоростей передачи информации (максимальной и реальной). Рассмотрены алгоритмы работы калькуляторов для расчета пропускной способности узлов доступа.

2. Приведена характеристика новых концепций и бизнес-моделей, используемых в узлах доступа, для создания условий по качественному обслуживанию возникающих требований. Среди них: интернет вещей, сетевая нарезка, операторы систем наблюдения.
3. Показано, что успешная реализация новых концепций развертывания сетей и управления распределением пропускной способности между поступающими требованиями требует построения новых моделей мультисервисных узлов доступа, в которых учитывается совокупное влияние основных значимых факторов, влияющих на формирования входных потоков заявок и их обслуживание. Среди них: зависимость скорости передачи от типа запроса, приоритетность нагрузки онлайн-услуг, возможность поступления файлов группами случайного размера, наличие или отсутствие у файлов эластичных свойств, присутствие у файлов возможности ожидания начала передачи и старения передаваемой информации.

Раздел 2

Построение и анализ обобщенной модели обслуживания гетерогенного трафика в мультисервисных узлах доступа

2.1 Введение к разделу 2

Объектом исследования в диссертации выступает беспроводной узел доступа, работающий в стандарте LTE¹. Часть его пропускной способности оператор систем наблюдения использует для передачи данных, поступающих с видеокамер и датчиков телеметрии. Математическая модель узла доступа рассмотрена в подразделе 2.2. В сетях связи для характеристики ресурса передачи информации используется понятие пропускной способности. Более подробно это понятие обсуждается в подразделе 2.2.1. Совместная передача гетерогенного трафика приводит к проблеме неконтролируемого перераспределения пропускной способности узла в пользу так называемого легкого трафика, требующего малый объем ресурса для своего обслуживания. Возникающие сложности в обслуживании неоднородного трафика рассмотрены в подразделе 2.2.2. Функциональная модель узла доступа приведена в подразделе 2.2.3. На ее основе в подразделе 2.2.4 построен марковский процесс, описывающий изменение во времени состояний исследуемой модели узла доступа. В подразделе 2.2.5 свойства марковского процесса применяются для вычисления ключевых параметров совместного обслуживания поступающих запросов.

Данные параметры определяются на основе стационарных вероятностей состояний системы, полученных в результате решения уравнений статистического баланса. Рекурсивная форма представления уравнений баланса представлена в подразделе 2.3. В зависимости от конкретики решаемой задачи все перечисленные особенности формирования трафика могут присутствовать в модели распределения ресурса узла доступа. Понятно, что при необходимости можно ограничиться учетом меньшего числа характеристик потоков. В следующих двух подразделах рассмотрены частные случаи построенной комплексной модели узла доступа, которые будут использованы для построения приближенных алгоритмов расчета показателей обслуживания заявок, и сформулированы выводы по результатам второго раздела.

¹ Данное допущение не носит критического характера. Также могут рассматриваться узлы доступа стандартов 5G и 6G.

2.2 Математическое описание узла доступа

2.2.1 Ресурс передачи информации

Основной характеристикой узла доступа является используемый телекоммуникационный ресурс, предназначенный для передачи информации, генерируемой запросами абонентов при обращении к различным сервисам. Численной характеристикой ресурса является его объем, который оценивается на основе введения единицы ресурса (виртуального канала). Пропускная способность виртуального канала выбирается в каждом конкретном случае из условий возможности проведения дальнейшего численного анализа модели и возникающей при этом погрешности в оценке характеристик обслуживания заявок. Величина пропускной способности узла оценивается через максимально возможное число создаваемых виртуальных каналов.

В узлах доступа для численной характеристики скорости передачи информации используется понятие *пропускная способность (throughput)*. Оно имеет много толкований, представленных в технической литературе. В дальнейшем говоря об этой характеристике объекта исследования будем подразумевать, что речь идет либо о максимальной пропускной способности (maximum throughput), либо о требуемой пропускной способности (required throughput). Оба параметра могут быть рассчитаны с использованием соответствующих калькуляторов или моделирования (см. подраздел 1.3.4). Первая характеристика представляет из себя теоретический показатель, определяющий максимальный объем цифровой информации, который может быть пропущен системой связи в единицу времени в идеальных условиях. Её численная величина служит верхней оценкой для реальных значений формируемой скорости передачи узла доступа. Вторая характеристика задаёт скорость передачи информации требуемую для пересылки заданных информационных потоков с заданным качеством. Её значение находится в результате моделирования.

2.2.2 Сложности пересылки гетерогенного трафика

До недавних пор в основе планирования сетей связи лежал принцип объединения возникающих информационных потоков и использования общего ресурса для их передачи. Интуитивно понятно, а в ряде случаев это можно доказать и теоретически, что реализация этого принципа уменьшает потери поступающих заявок и увеличивает коэффициент использования единицы ресурса по сравнению с раздельным использованием ресурса передачи информации каждым потоком. Со временем выяснилось, что отмеченная эффективность достигается в основном в случае обслуживания моносервисного трафика, когда поступающие заявки требуют одинаковую порцию ресурса для своего обслуживания.

С увеличением числа сервисов меняются свойства информационных потоков, которые появляются в процессе их предоставления. Первое, на что необходимо обратить внимание — это появление свойства гетерогенности. Его наличие означает присутствие существенного разброса в требованиях к ресурсу передачи информации у поступающих заявок. Будем считать, что данное свойство выполняется, если пропускная способность, требуемая поступающими заявками, отличается более чем в десять раз. Трафик с низкими требованиями

к пропускной способности назовем легким, а с высокими — тяжелым. Известно, что при совместной обработке легкого и тяжелого трафика последний вытесняется из системы обслуживания.

В подразделе 4.2 проведено подробное численное исследование влияния свойства гетерогенности на характеристики совместного обслуживания информационных потоков, порожденных современными коммуникационными приложениями. Для проведения исследования использовалась комплексная математическая модель узла доступа, параметры и характеристики которой будут введены в подразделах 2.2.3–2.2.5, а алгоритм расчета характеристик построен и исследован в подразделе 3.2. Здесь приведем только один пример, иллюстрирующий отмеченные выше проблемы совместного обслуживания гетерогенного трафика на общем ресурсе.

Рассмотрим процесс обслуживания легкого и тяжелого трафика в узле доступа, обладающего пропускной способностью $C = 100$ Мбит/с. Для обслуживания легкого трафика (передача эластичных файлов) может быть использован любой свободный ресурс, обеспечивающий скорость не ниже 1 Мбит/с, а для передачи тяжелого трафика (передача видеoinформации) используется фиксированная скорость 20 Мбит/с.

В левой части рисунка 2.1 представлены графики, отражающие долю потерь заявок для трафика реального времени (кривая RT) и эластичных данных (кривая DPS) в зависимости от загрузки единицы ресурса ρ . В правой части рисунка показаны средние значения использования ресурса каждым типом трафика, причем кривые обозначены аналогичными символами. Из графиков видно, что при увеличении нагрузки легкий трафик практически полностью вытесняет тяжелый трафик из системы обслуживания.

Такое положение дел может серьезно нарушить принятое соглашение об обслуживании мультимедийного трафика. Требуется принимать меры для устранения возникающих проблем. Наиболее радикальное решение заключается в использовании положений концепции Network Slicing (см. подраздел 1.4.1). Согласно этому подходу каждому информационному потоку или группе потоков с примерно одинаковыми свойствами выделяется определенный объем передаточного в количестве достаточном, чтобы обеспечить заданные показатели обслуживания поступающей нагрузки. Тем самым у проектировщиков появляются возможности создавать условия по дифференцированному обслуживанию поступающих потоков мультимедийного трафика.

Однако у этого подхода имеется один существенный недостаток — неэффективное использование передаточных возможностей ресурса. Переконфигурирование ресурса, выделенного под слайсы, требует времени, а его может не хватить в ситуации частых изменений профиля поступающей нагрузки. В этой ситуации предлагается создавать условия для дифференцированного обслуживания трафика, используя средства контроля за доступом заявок в узел, зависящие от его загрузки. Анализ численных данных, представленных на рисунке 2.1 показывает, что в случае малой нагрузки все поступающие заявки не испытывают проблем в занятии ресурса. Проблемы, вызванные доминированием одного типа заявок над другими, появляются с ростом нагрузки на канал. В этой ситуации предлагается использовать сценарии дифференцированного обслуживания гетерогенного трафика. Анализ соответствующих алгоритмов доступа рассмотрен в подразделе 4.3.

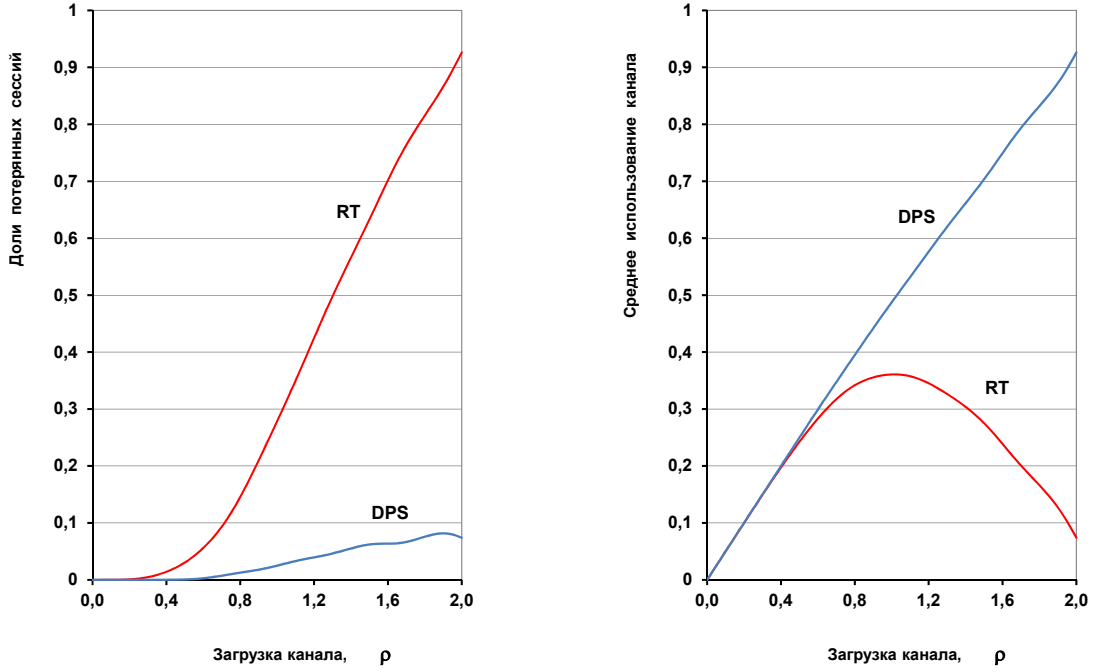


Рисунок 2.1 — Зависимость характеристик обслуживания гетерогенного трафика от ρ

2.2.3 Функциональная модель

Пусть C обозначает пропускную способность мультисервисного узла доступа, определяемую стандартом мобильной связи и измеряемую в бит/с. Для упрощения моделирования представим как саму пропускную способность узла, так и требования входящих запросов в терминах виртуальных каналов. Пропускную способность одного виртуального канала обозначим символом c (бит/с). Выбор численной величины c исследовался в подразделе 2.2.2. Он зависит от точности аппроксимации исходных данных, относящихся к параметрам входных потоков. Во многих случаях, в качестве величины c можно брать минимальную скорость передачи, необходимую для обслуживания поступившей заявки.

В узле доступа обслуживается простейший поток запросов на передачу нагрузки онлайн-услуг интенсивности λ_r . Запросы разделены на n категорий. Доля запросов k -й категории находится из соотношения $p_{r,k}$. Запросу k -й категории предоставляется ресурс c_k (бит/с) на случайное время, распределенное экспоненциально с параметром $\alpha_{r,k}$, где $k = 1, \dots, n$.

Чтобы упростить задачу моделирования величины C и c_k выражаются в виртуальных каналах (в.к.). Общее число каналов v определяется как $v = \lfloor \frac{C}{c} \rfloor$ в.к. Количество каналов b_k для обслуживания одной заявки k -й категории вычисляется по формуле $b_k = \lceil \frac{c_k}{c} \rceil$ в.к. Интенсивность потока заявок k -й категории равна $\lambda_{r,k} = \lambda_r p_{r,k}$ ($k = 1, \dots, n$).

Помимо обработки информационных потоков, связанных с сервисами реального времени, узел доступа также передает групповой поток эластичных данных в виде файлов. Поступление групп файлов описывается простейшим потоком с интенсивностью λ_d . Каждое поступление заявки в данном потоке означает, что требуется передать группу файлов.

Количество файлов в группе s определяется вероятностью f_s , где $s = 1, 2, \dots, m_g$, а m_g — максимально возможный размер группы. Среднее число файлов в поступающей группе вычисляется по формуле:

$$\bar{b} = \sum_{s=1}^{m_g} f_s \cdot s.$$

Файлам из группы, которым не хватило свободной пропускной способности были приняты на передачу сразу предоставляется возможность ожидания освобождения ресурсов. Время пребывания в состоянии ожидания ограничено случайной величиной, имеющей экспоненциальное распределение со средним значением $\frac{1}{\sigma}$. Если за это время передача так и не началась, файл считается утерянным и не возобновляется. Указанные функциональные характеристики комплексной модели узла доступа представлены на рисунке 2.2.

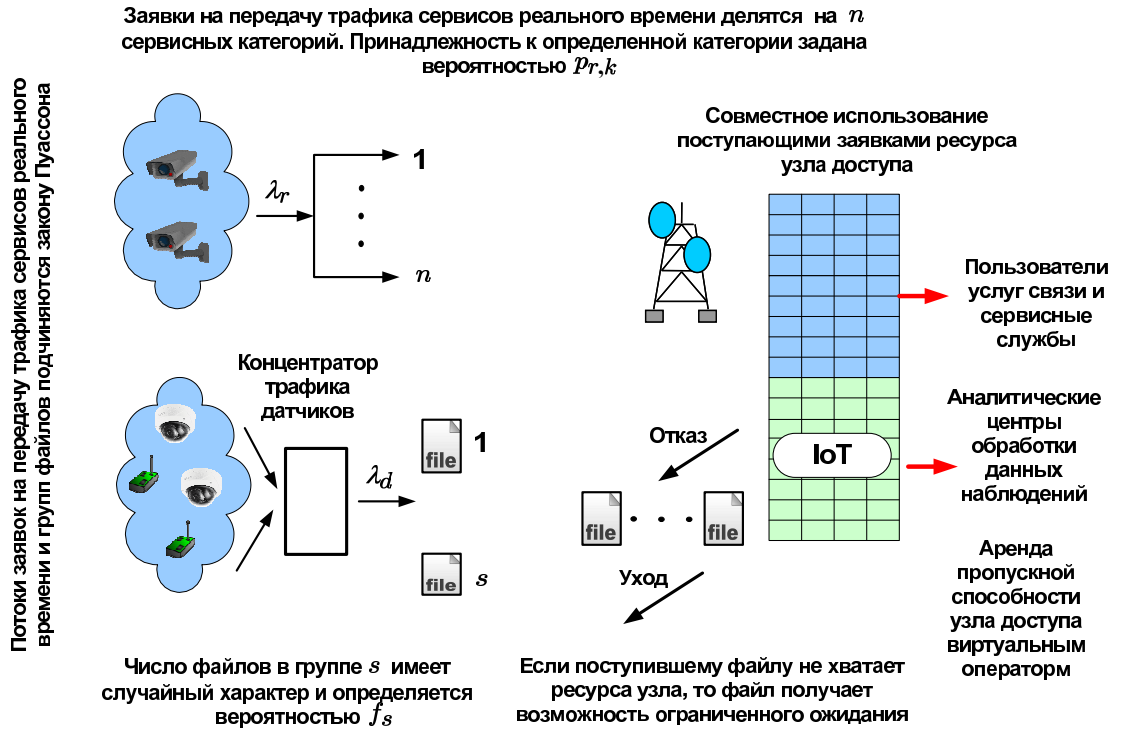


Рисунок 2.2 — Функциональная модель исследуемого узла доступа

2.2.4 Марковский процесс

Допустим, что объем файла имеет экспоненциальное распределение. Его средняя величина F бит. По предположению файл получает для своей передачи пропускную способность узла доступа не менее одной единицы ресурса передачи информации т.е. одного виртуального канала. По этой причине длительность передачи одного файла с использованием одного канала имеет экспоненциальное распределение с параметром α_d , величина которого определяется из выражения $\alpha_d = \frac{c}{F}$. Пусть в момент поступления запроса ℓ — число

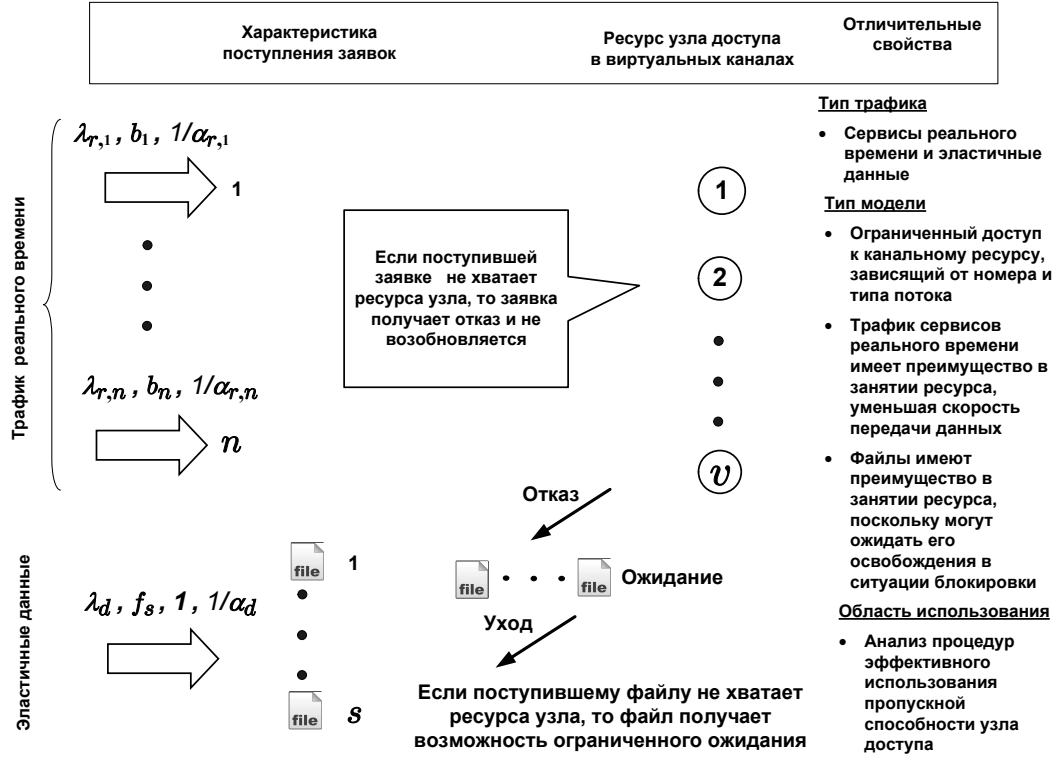


Рисунок 2.3 — Математическая модель узла доступа

2.2.5 Характеристики обслуживания заявок

Характеристики обслуживания заявок k -го потока нагрузки онлайн-услуг определяются следующими выражениями:

- вероятность блокировки поступивших заявок — $\pi_{r,k}$;
- математическое ожидание количества используемых единиц ресурса передачи информации — $m_{r,k}$;
- математическое ожидание количества заявок, находящихся на обслуживании: $y_{r,k} = m_{r,k}/b_k$;
- математическое ожидание количества используемой пропускной способности узла доступа $z_{r,k} = m_{r,k}c$.

Качество обслуживания запросов на пересылку файлов, допускающих переменную скорость передачи характеризуется следующими показателями:

- $\pi_{d,t}$ — вероятность блокировки из-за недостаточности пропускной способности узла доступа для приема поступившего файла;
- $\pi_{d,c}$ — вероятность блокировки из-за недостаточности пропускной способности узла доступа для приема поступившего файла и нехватки в этой ситуации позиций для постановки файла на ожидание;

- π_d — вероятность блокировки файлов по всем причинам анализируемым в построенной модели узла доступа;
- m_d — математическое ожидание количества единиц ресурса узла доступа, используемых на пересылку файлов;
- $z_d = m_d c$ — математическое ожидание величины пропускной способности узла доступа, занятой передачей файлов;
- h_d — математическое ожидание времени пребывания файла в узле доступа;
- b_f — математическое ожидание числа единиц ресурса узла доступа, используемых для передачи одного файла;
- $c_d = b_f c$ — математическое ожидание величины пропускной способности узла, используемой для передачи одного файла;
- L_q — математическое ожидание количества файлов, ожидающих начала обслуживания;
- L_s — математическое ожидание количества файлов, находящихся в состоянии передачи;
- $L = L_q + L_s$ — математическое ожидание количества файлов в системе (ожидание + обслуживание).

Значения введенных характеристик можно рассчитать, если известны величины предельных вероятностей $p(i_{r,1}, \dots, i_{r,n}, i_d)$ для всех состояний модели $(i_{r,1}, \dots, i_{r,n}, i_d) \in S$. Для состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$ обозначим через ℓ число единиц ресурса узла доступа, используемых на пересылку нагрузки онлайн-услуг

$$\ell = i_{r,1}b_1 + \dots + i_{r,n}b_n.$$

Предполагается, что максимальное число файлов m_g в группе не превышает $v + w$.

Формальные выражения для перечисленных характеристик можно получить через стационарные вероятности $p(i_{r,1}, \dots, i_{r,n}, i_d)$ марковского процесса $r(t)$ имеют вид:

$$\begin{aligned} \pi_{r,k} &= \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d + b_k > v\}} p(i_{r,1}, \dots, i_{r,n}, i_d); \\ m_{r,k} &= \sum_{(i_{r,1}, \dots, i_{r,n}, i_d) \in S} p(i_{r,1}, \dots, i_{r,n}, i_d) i_{r,k} b_k; \\ k &= 1, \dots, n. \end{aligned} \tag{2.2}$$

Введем характеристики качества обслуживания запросов на передачу эластичных данных. Величина $\pi_{d,t}$ рассчитывается суммированием вероятностей состояний, в которых поступивший файл либо становится на ожидание, либо теряется. Выражение для $\pi_{d,t}$ имеет

вид:

$$\pi_{d,t} = \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d \geq v\}} p(i_{r,1}, \dots, i_{r,n}, i_d).$$

Найдем выражение через вероятности стационарных состояний для $\pi_{d,c}$. Обозначим через Λ_b интенсивность потока файлов, которые будут потеряны из-за нахождения узла доступа в состояниях, когда $\ell + i_d = v + w$. Найдем среднее число файлов, которые будут потеряны в состоянии узла $(i_{r,1}, \dots, i_{r,n}, i_d - i)$, где $(i_{r,1}, \dots, i_{r,n}, i_d) \in S$, $\ell + i_d = v + w$ и $i = 0, 1, \dots, i_d$. Появление группы файлов, содержащих $i + 1, i + 2, \dots, v + w$ файлов выразится в потере соответственно $1, 2, \dots, v + w - i$ файлов. Среднее число файлов, получивших отказ в $(i_{r,1}, \dots, i_{r,n}, i_d - i)$, определяется из выражения

$$f_{i+1} \cdot 1 + f_{i+2} \cdot 2 + \dots + f_{v+w} \cdot (v + w - i) = \sum_{s=i+1}^{v+w} f_s (s - i).$$

Усредняя по всем $i = 0, 1, \dots, i_d$, получаем выражение для Λ_b

$$\Lambda_b = \lambda_d \sum_{i=0}^{i_d} \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d = v + w\}} p(i_{r,1}, \dots, i_{r,n}, i_d - i) \sum_{s=i+1}^{v+w} f_s (s - i).$$

Общая интенсивность поступающего потока файлов определяется из выражения $\lambda_d \bar{b}$. Окончательное выражение для $\pi_{d,c}$ имеет вид

$$\pi_{d,c} = \frac{1}{\bar{b}} \times \left(\sum_{i=0}^{i_d} \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d = v + w\}} p(i_{r,1}, \dots, i_{r,n}, i_d - i) \sum_{s=i+1}^{v+w} f_s (s - i) \right). \quad (2.3)$$

Для оценки доли всех потерянных файлов π_d необходимо поделить интенсивность файлов, потерянных по всем причинам, анализируемым в модели на интенсивность поступающих файлов. Получаем такой ответ

$$\pi_d = \frac{1}{\lambda_d \bar{b}} \times \left(\Lambda_b + \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d > v\}} p(i_{r,1}, \dots, i_{r,n}, i_d) (\ell + i_d - v) \sigma \right). \quad (2.4)$$

Оставшиеся характеристики качества обслуживания файлов определяются из следующих выражений

$$m_d = \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid i_d > 0\}} p(i_{r,1}, \dots, i_{r,n}, i_d) (v - \ell); \quad (2.5)$$

$$L_q = \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d > v\}} p(i_{r,1}, \dots, i_{r,n}, i_d) (\ell + i_d - v);$$

$$\begin{aligned}
L_s &= \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d \leq v\}} p(i_{r,1}, \dots, i_{r,n}, i_d) i_d + \\
&+ \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d > v\}} p(i_{r,1}, \dots, i_{r,n}, i_d) (v - \ell); \\
L &= L_q + L_s = \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S\}} p(i_{r,1}, \dots, i_{r,n}, i_d) i_d; \\
h_d &= \frac{L}{\lambda_d \bar{b} (1 - \pi_{d,c})}, \quad b_f = \frac{m_d}{L_s}.
\end{aligned}$$

Введенные характеристики обслуживания поступающих заявок связаны системой соотношений, которые для каждого потока следуют из равенства интенсивностей поступающих, потерянных и обслуженных потоков заявок. Число этих соотношений равно числу потоков $n + 1$. Они имеют вид:

$$\lambda_{r,k} b_k = \lambda_{r,k} \pi_{r,k} b_k + m_{r,k} \alpha_{r,k}; \quad k = 1, \dots, n; \quad (2.6)$$

$$\lambda_d \bar{b} = \lambda_d \pi_{d,c} \bar{b} + m_d \alpha_d + L_q \sigma.$$

Указанные соотношения являются следствием формулы Литтла и допускают также алгебраический вывод. Для этого требуется умножить каждое уравнение системы равновесия (2.8) на величины $i_{r,k}$ (где $k = 1, \dots, n$) и i_d , а затем выполнить суммирование полученных выражений по всем допустимым состояниям $(i_{r,1}, \dots, i_{r,n}, i_d) \in S$. Результат такого суммирования дает возможность получить альтернативные выражения для расчета рассматриваемых характеристик. В частности, при условии $w = 0$ величины h_d и b_f вычисляются по формулам:

$$h_d = \frac{L_s}{m_d \alpha_d}, \quad b_f = \frac{1}{h_d \alpha_d}.$$

Кроме того, соотношение (2.6) может служить для проверки правильности решения системы уравнений равновесия (СУР) (2.8) при использовании методов последовательных замещений.

2.3 Система уравнений равновесия

2.3.1 Рекурсивное формирование СУР

Расчёт вероятностных показателей модели основан на использовании предельных вероятностях её состояний $p(i_{r,1}, \dots, i_{r,n}, i_d)$. Для определения этих вероятностей необходимо составить и решить систему уравнений статистического равновесия, устанавливающих связи между вероятностями различных состояний. Система выводится по стандартной методике теории марковских процессов, исходя из принципа равенства суммарной интенсивности выхода процесса $r(t)$ из заданного состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$ и суммарной интенсивности входа в это состояние. Высокая размерность пространства состояний и отсутствие упрощающих структур (таких как возможность применения рекурсивных алгоритмов) вынуждают решать полученную систему классическими методами линейной алгебры. Эксперты

отмечают [16, 34, 32–35], что наибольшую эффективность здесь демонстрируют алгоритмы последовательных замещений, в частности, метод Гаусса-Зейделя. Ключевым условием для их успешного использования является запись системы уравнений баланса в форме единого соотношения, коэффициенты которого явно определяются компонентами состояния. Такое представление приводится далее.

В рассматриваемой модели выделяется пять типов событий, изменяющих её состояние:

1. Поступление заявки k -го потока нагрузки online-услуг. Интенсивность этого события для любого состояния модели равна $\lambda_{r,k}$.
2. Поступление группы заявок на пересылку файлов данных, допускающих переменную скорость передачи. Интенсивность события постоянна и равна λ_d .
3. Освобождение ресурса передачи информации узла доступа после завершения обслуживания заявки k -го потока нагрузки online-услуг. Если в системе находится $i_{r,k}$ заявок, интенсивность события рассчитывается из выражения $i_k \alpha_{r,k}$.
4. Освобождение ресурса передачи информации узла доступа после завершения пересылки файлов данных. Если обслуживаются i_d файлов, то интенсивность рассматриваемого события равна $(v - \ell) \alpha_d$, где ℓ — количество единиц ресурса передачи информации узла доступа занятых на пересылку нагрузки online-услуг.
5. Прекращение ожидания начала обслуживания для файлов в очереди. Если в ожидании находится i_d файлов, интенсивность события равна $i_d \sigma$.

2.3.2 Анализ событий, меняющих состояние модели

Изучим возможные сценарии, при которых модель покидает $(i_{r,1}, \dots, i_{r,n}, i_d)$, а также возвращается в него, после чего построим систему уравнений баланса для $r(t)$. Сначала запишем левую часть уравнения статистического равновесия, включающую ненормированную вероятность $P(i_{r,1}, \dots, i_{r,n}, i_d)^2$. Для этого необходимо определить условия, при которых происходят все указанные события, и выразить интенсивность выхода модели из состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$. Введем обозначение ℓ — общее количество занятых единиц ресурса в состоянии $(i_{r,1}, \dots, i_{r,n}, i_d)$:

$$\ell = i_{r,1} b_1 + \dots + i_{r,n} b_n.$$

Приход заявки k -го потока для передачи трафика сервисов реального времени (с интенсивностью $\lambda_{r,k}$) приводит к выходу модели из состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$, если заявка принимается на обслуживание с возможной адаптацией скорости передачи эластичных данных (т. е. для всех состояний, где $\ell + i_d + b_k \leq v$). В этом случае с интенсивностью $P(i_{r,1}, \dots, i_{r,n}, i_d) \lambda_{r,k}$ осуществляется переход из $(i_{r,1}, \dots, i_{r,n}, i_d)$ в $(i_{r,1}, \dots, i_{r,k} + 1, \dots, i_{r,n}, i_d)$.

²Здесь и далее заглавные буквы обозначают ненормированные вероятности состояний, а строчные — нормированные.

Поступление заявки на передачу группы файлов, допускающих переменную скорость передачи (с интенсивностью потока заявок λ_d) всегда выводит модель из состояния модели $(i_{r,1}, \dots, i_{r,n}, i_d)$, если из поступившей группы на передачу или ожидание принимается хотя бы один файл. Для этого требуется выполнение условия $\ell + i_d + 1 \leq v + w$. В этом случае с интенсивностью $P(i_{r,1}, \dots, i_{r,n}, i_d)\lambda_d$ происходит переход из $(i_{r,1}, \dots, i_{r,n}, i_d)$ в $(i_{r,1}, \dots, i_{r,n}, i_d + i)$, где i — число принятых файлов.

Завершение обслуживания заявки k -го потока трафика сервисов реального времени (с интенсивностью $i_{r,k}\alpha_{r,k}$) всегда изменяет состояние $(i_{r,1}, \dots, i_{r,n}, i_d)$, если в данный момент обслуживается хотя бы одна такая заявка (т. е. при $i_{r,k} > 0$). Соответственно, с интенсивностью $P(i_{r,1}, \dots, i_{r,n}, i_d)i_{r,k}\alpha_{r,k}$ система переходит из $(i_{r,1}, \dots, i_{r,n}, i_d)$ в $(i_{r,1}, \dots, i_{r,k} - 1, \dots, i_{r,n}, i_d)$.

Следующие два события связаны с завершением пребывания файла в узле доступа.

1. Если в системе есть файлы, находящиеся в процессе передачи (т.е. $i_d > 0$), то окончание передачи (с интенсивностью $(v - \ell)\alpha_d$) всегда приводит к изменению состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$. При этом с интенсивностью $P(i_{r,1}, \dots, i_{r,n}, i_d)(v - \ell)\alpha_d$ система переходит из состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$ в состояние $(i_{r,1}, \dots, i_{r,n}, i_d - 1)$.
2. Если в системе присутствуют файлы, ожидающие начала обслуживания (т.е. для всех состояний $(i_{r,1}, \dots, i_{r,n}, i_d)$, где $\ell + i_d > v$), то окончание периода ожидания (с интенсивностью $(\ell + i_d - v)\sigma$) гарантированно изменяет текущее состояние. В этом случае с интенсивностью $P(i_{r,1}, \dots, i_{r,n}, i_d)(\ell + i_d - v)\sigma$ происходит переход из $(i_{r,1}, \dots, i_{r,n}, i_d)$ в $(i_{r,1}, \dots, i_{r,n}, i_d - 1)$.

В таблице 2.1 приведено формирование коэффициентов для левой части системы уравнений равновесия.

В таблице 2.2 приведены компоненты, необходимые для формирования правой части СУР.

Определены отдельные компоненты, формирующие интенсивность выхода марковского процесса $r(t)$ из состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$. Для составления соответствующего выражения в правой части уравнения статистического равновесия (где в левой части стоит ненормированная вероятность $P(i_{r,1}, \dots, i_{r,n}, i_d)$) необходимо аналогичным образом рассмотреть события, приводящие к переходу $r(t)$ в состояние $(i_{r,1}, \dots, i_{r,n}, i_d)$.

Переход процесса $r(t)$ в состояние $(i_{r,1}, \dots, i_{r,n}, i_d)$ может быть вызван: поступлением и принятием на обслуживание заявки k -го потока нагрузки online-услуг. Интенсивность события $\lambda_{r,k}$. При этом возможно изменение скорости передачи файлов. Для осуществления этого перехода необходимо выполнение двух условий:

1. Состояние $(i_{r,1}, \dots, i_{r,k} - 1, \dots, i_{r,n}, i_d)$ должно находиться в используемом пространстве состояний модели S (т. е. $i_{r,k} > 0$).
2. После приёма заявки занятый ресурс не должен превышать v (т. е. $\ell + i_d \leq v$).

Если условия выполнены, переход из состояния $(i_{r,1}, \dots, i_{r,k} - 1, \dots, i_{r,n}, i_d)$ в состояние $(i_{r,1}, \dots, i_{r,k}, \dots, i_{r,n}, i_d)$ происходит с интенсивностью: $P(i_{r,1}, \dots, i_{r,k} - 1, \dots, i_{r,n}, i_d) \cdot \lambda_{r,k}$.

Таблица 2.1 — Формирование коэффициентов при $P(i_{r,1}, \dots, i_{r,n}, i_d)$ в левой части системы уравнений.

Событие	Переход в состояние	Интенсивность	Условие
Приём заявки k -го потока	$(i_{r,1}, \dots, i_{r,k} + 1, \dots, i_{r,n}, i_d)$	$\lambda_{r,k} P(i_{r,1}, \dots, i_{r,n}, i_d)$	$I(\ell + i_d + b_k \leq v)$
Окончание обл. заявки k -го потока	$(i_{r,1}, \dots, i_{r,k} - 1, \dots, i_{r,n}, i_d)$	$i_{r,k} \alpha_{r,k} P(i_{r,1}, \dots, i_{r,n}, i_d)$	$I(i_{r,k} > 0)$
Приём группы файлов	$(i_{r,1}, \dots, i_{r,n}, i_d + 1)$	$\lambda_d P(i_{r,1}, \dots, i_{r,n}, i_d)$	$I(\ell + i_d + 1 \leq v + w)$
Окончание процедуры пересылки файла	$(i_{r,1}, \dots, i_{r,n}, i_d - 1)$	$(v - \ell) \alpha_d P(i_{r,1}, \dots, i_{r,n}, i_d)$	$I(i_d > 0)$
Окончание процесса ожидания нач. обл.	$(i_{r,1}, \dots, i_{r,n}, i_d - 1)$	$(\ell + i_d - v) \sigma P(i_{r,1}, \dots, i_{r,n}, i_d)$	$I(\ell + i_d > v)$

Таблица 2.2 — Формирование правой части СУР для состояния $P(i_{r,1}, \dots, i_{r,n}, i_d)$

Событие	Переход из состояния	Интенсивность	Условие
Приём заявки k -го потока	$(i_{r,1}, \dots, i_{r,k} - 1, \dots, i_{r,n}, i_d)$	$P(i_{r,1}, \dots, i_{r,k} - 1, \dots, i_{r,n}, i_d) \lambda_{r,k}$	$I(i_{r,k} > 0, \ell + i_d \leq v)$
Окончание обл. заявки k -го потока	$(i_{r,1}, \dots, i_{r,k} + 1, \dots, i_{r,n}, i_d)$	$P(i_{r,1}, \dots, i_{r,k} + 1, \dots, i_{r,n}, i_d) (i_{r,k} + 1) \alpha_{r,k}$	$I(\ell + b_k + i_d \leq v) + I(\ell + b_k + i_d > v, \ell + b_k \leq v, \ell + b_k + i_d \leq v + w)$
Поступление группы из i и более файлов	$(i_{r,1}, \dots, i_{r,n}, i_d - i)$	$P(i_{r,1}, \dots, i_{r,n}, i_d - i) \lambda_d$	$f_i + I(\ell + i_d = v + w) \sum_{j=i+1}^{v+w} f_j$
Окончание процедуры пересылки файла	$(i_{r,1}, \dots, i_{r,n}, i_d + 1)$	$P(i_{r,1}, \dots, i_{r,n}, i_d + 1) (v - \ell) \alpha_d$	$I(\ell + i_d + 1 \leq v + w)$
Окончание ожидания начала обл. файла	$(i_{r,1}, \dots, i_{r,n}, i_d + 1)$	$P(i_{r,1}, \dots, i_{r,n}, i_d + 1) (\ell + i_d + 1 - v) \sigma$	$I(\ell + i_d + 1 \leq v + w, \ell + i_d + 1 > v)$

Переход случайного процесса $r(t)$ в состояние $(i_{r,1}, \dots, i_{r,n}, i_d)$ может произойти в результате поступления заявки на передачу группы файлов. Для реализации события необходимо учесть возможное изменение скорости передачи эластичных данных и размеры

левая и правая части уравнения СУР для состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$ определяются из следующего соотношения:

$$\begin{aligned}
& P(i_{r,1}, \dots, i_{r,n}, i_d) \left\{ \sum_{k=1}^n \left(\lambda_{r,k} I(\ell + i_d + b_k \leq v) + i_{r,k} \alpha_{r,k} I(i_{r,k} > 0) \right) + \right. \\
& \left. + \lambda_d I(\ell + i_d + 1 \leq v + w) + (v - \ell) \alpha_d I(i_d > 0) + (\ell + i_d - v) \sigma I(\ell + i_d > v) \right\} = \\
& = \sum_{k=1}^n P(i_{r,1}, \dots, i_{r,k} - 1, \dots, i_{r,n}, i_d) \lambda_{r,k} I(i_{r,k} > 0, \ell + i_d \leq v) + \\
& + \sum_{i=1}^{i_d} P(i_{r,1}, \dots, i_{r,n}, i_d - i) \lambda_d \left(f_i + I(\ell + i_d = v + w) \sum_{j=i+1}^{v+w} f_j \right) + \\
& + \sum_{k=1}^n P(i_{r,1}, \dots, i_{r,k} + 1, \dots, i_{r,n}, i_d) (i_{r,k} + 1) \alpha_{r,k} \times \\
& \times \left(I(\ell + b_k + i_d \leq v) + I(\ell + b_k + i_d > v, \ell + b_k \leq v, \ell + b_k + i_d \leq v + w) \right) + \\
& + P(i_{r,1}, \dots, i_{r,n}, i_d + 1) \times \left((v - \ell) \alpha_d I(\ell + i_d + 1 \leq v + w) + \right. \\
& \left. + (\ell + i_d + 1 - v) \sigma I(\ell + i_d + 1 \leq v + w, \ell + i_d + 1 > v) \right).
\end{aligned} \tag{2.8}$$

Для значений $P(i_{r,1}, \dots, i_{r,n}, i_d)$ выполнено условие нормировки

$$\sum_{(i_{r,1}, \dots, i_{r,n}, i_d) \in S} P(i_{r,1}, \dots, i_{r,n}, i_d) = 1.$$

Полученную форму записи СУР удобно использовать для формирования СУР на любом алгоритмическом языке. Для этого достаточно заменить результат действия индикаторной функции (2.7) на выполнение условного оператора. Алгоритм решения системы уравнений равновесия итерационным методом Гаусса-Зейделя рассмотрен в следующем разделе диссертации.

2.4 Частные случаи обобщенной модели узла доступа

Рассмотрим несколько частных случаев исследуемой модели узла доступа. В некоторых из них процесс поступления и обслуживания заявок упрощается и значения характеристик могут быть найдены более простыми методами. В данное семейство входят четыре модели. В дальнейшем они найдут применение при оценке характеристик обслуживания нагрузки online-услуг и файлов данных, допускающих переменную скорость передачи в рассматриваемых частных случаях. Другая область применения моделей — разработка приближённых методов расчёта показателей совместного обслуживания запросов в рассматриваемой модели узла доступа.

1. Эта модель отличается от модели, рассмотренной в подразделе 2.2, изменением процедуры распределения ресурса узла доступа между файлами, попавшими на передачу. В рассматриваемой ситуации распределение ресурсов осуществляется с учетом

приоритетного обслуживания трафика сервисов реального времени, то есть исключительно в рамках возможностей одного канала. При этом общий принцип обработки и совместного обслуживания нагрузки online-услуг и файлов данных, допускающих переменную скорость передачи остается прежним. Следует отметить, что в рассматриваемой модели все типы запросов обрабатываются в соответствии с правилами, установленными для нагрузки online-услуг. При этом поступающие файлы и сессии передачи трафика реального времени имеют одинаковый уровень приоритета в занятии ресурса. Правда файлы, имея возможность ожидать начала обслуживания, имеют больше возможностей его завершить нежели сессии передачи трафика реального времени. Величины характеристик оцениваются из решения системы уравнений равновесия методом последовательных замещений. Данная модель и алгоритм оценки показателей обслуживания поступающих запросов рассмотрены в подразделе 3.3.1. Результаты исследования модели будут использоваться в подразделе 3.3.2 для анализ свойств совместного обслуживания гетерогенного трафика. Рассмотренная модель также позволяет оценить эффективность использования дисциплины Processor Sharing при передаче файлов с эластичными свойствами.

2. Данная модель является дальнейшим упрощением рассмотренной выше первой модели. Упрощение заключается в отсутствии возможности ожидания заблокированного файла. В остальном формирование входных потоков и организация их обслуживания не меняются. Величины характеристик могут быть простым рекурсивным алгоритмом. Данная модель и алгоритм оценки показателей обслуживания поступающих запросов рассмотрены в подразделе 3.3.3. Результаты исследования модели будут использоваться для анализа свойств совместного обслуживания гетерогенного трафика.
3. В этом частном случае рассматривается процесс обслуживания только трафика сервисов реального времени. Интенсивность поступления файлов принята равной нулю. Величины характеристик рассчитываются простым рекурсивным алгоритмом. Результаты исследования модели будут использоваться в подразделе 3.4 при разработке приближенных алгоритмов оценки обобщенной модели узла доступа.
4. В последнем из рассматриваемых частных случаев исследуется процесс обслуживания только эластичного трафика данных. Интенсивность поступления заявок на передачу сессий реального времени принята равной нулю. Величины характеристик рассчитываются простым рекурсивным алгоритмом. Результаты исследования модели будут использоваться в подразделе 3.4 при разработке приближенных алгоритмов оценки обобщенной модели узла доступа.

2.5 Выводы по результатам второго раздела

1. Разработана и исследована функциональная модель совместного обслуживания гетерогенного трафика современных коммуникационных приложений в беспроводных узлах доступа. Рассмотрены функциональные характеристики поступления заявок

и принципы разделения ресурса передачи информации между ними. Трафик сервисов реального времени представляет из себя информационные потоки, относящиеся к передаче речевых сообщений и разного рода видеоинформации, а трафик данных это — файлы, содержащие результаты наблюдений и другую подобную этой информацию. Предполагается, что файлы могут поступать группами случайного размера. Тем самым учитывается импульсный характер этого вида трафика. В модели также предполагается, что нагрузка online-услуг обладает приоритетом в занятии ресурса передачи информации, который выражается в уменьшении в ситуации необходимости скорости передачи файлов до некоторого заранее выбранного минимального значения. При этом файлы в случае блокировки имеют возможность ограниченного ожидания начала обслуживания.

2. Изменение состояния числа заявок, находящихся на обслуживании, предлагается исследовать с использованием марковских процессов. Этот подход позволяет использовать эффективные численные методы, развитые для данного семейства случайных процессов, для оценки характеристик совместного обслуживания трафика. Чтобы прийти к этому результату при построении математической модели узла доступа предполагалось использование экспоненциально распределенных независимых случайных величин при формировании длительностей интервалов времени между осуществлением всех событий.
3. На основе разработанной математической модели узла были выведены формулы для оценки ключевых параметров совместной обработки входящих информационных потоков. К ним относятся вероятность блокировки поступающих запросов и математическое ожидание задействованной пропускной способности узла доступа для каждого анализируемого потока, т.е. запросов на установление сессий пересылки нагрузки online-услуг и запросов на пересылку файлов данных, допускающих переменную скорость передачи. Для трафика данных также определены выражения, позволяющие рассчитать среднее количество файлов в очереди, средний объем ресурса узла доступа, затрачиваемого на передачу одного файла, и другие показатели. Все характеристики модели вычисляются на основе стационарных вероятностей её состояний.
4. Предлагается определять стационарные вероятности модели путем решения системы линейных уравнений равновесия. Из-за большого числа состояний и отсутствия специальных свойств у матрицы этой системы её решение требует применения численных методов линейной алгебры. Для реализации этих методов система уравнений равновесия представлена в виде единого соотношения. Отдельные уравнения системы формируются в ходе итерационного процесса, перебирающего целочисленные компоненты состояний модели.
5. Рассмотрено несколько частных случаев исследуемой модели узла доступа. В некоторых из них процесс поступления и обслуживания заявок упрощается и значения характеристик могут быть найдены более простыми методами. В это семейство включены четыре модели. В дальнейшем они найдут применение при оценке характеристик

обслуживания нагрузки online-услуг и файлов данных, допускающих переменную скорость передачи в рассматриваемых частных случаях. Другая область применения моделей — разработка приближённых методов расчёта показателей совместного обслуживания запросов в рассматриваемой модели узла доступа. Детальное рассмотрение этих моделей представлено в подразделе 3.4.

Раздел 3

Алгоритмы оценки характеристик обобщенной модели узла

3.1 Введение к разделу 3

В данном разделе будут построены алгоритмы оценки характеристик обслуживания гетерогенного трафика в мультисервисных узлах доступа. Для оценки характеристик будет использована обобщенная математическая модель узла и семейство ее частных случаев, построенных в разделе 2. Анализ пропускной способности мультисервисных точек доступа представляет собой нетривиальную задачу. Основные трудности связаны с огромным количеством возможных состояний системы, а также с учетом взаимозависимостей между процессами обработки трафика в разных участках сети. Универсальный характер модели требует привлечения численных методов решения систем уравнений баланса. Один из эффективных подходов это — применение метода последовательных замещений Гаусса-Зейделя. Формулировка метода и перечень процедур, повышающих эффективность его применения применительно к построенной модели узла, приведены в подразделе 3.2. В ряде частных случаев исследуемой модели удалось получить эффективные расчетные алгоритмы, которые позволяют вести оценку характеристик с использованием рекурсий по числу занятых единиц ресурса. Эти результаты обсуждаются в подразделе 3.3.

Разработка инженерных методик решения задачи по анализу условий дифференцированного обслуживания мультисервисного трафика в узлах доступа основана на использовании достаточно обоснованных приближенных алгоритмов оценки характеристик узла. Метод с отмеченными свойствами построен в подразделе 3.4. Он основан на технике декомпозиции исходной модели на отдельные модули, где каждый модуль представлен и исследован в подразделе 3.3, в качестве частного случая.

3.2 Точная оценка характеристик узла методом Гаусса-Зейделя

3.2.1 Формулировка метода

Матрица СУР (2.8) не обладает какими-либо специфическими свойствами, наличие которых упрощает решение СУР анализируемой модели мультисервисного узла доступа. По мнению экспертов в области численных расчетов характеристик систем связи [16,34,32–35] в этой ситуации для вычисления значений $p(i_{r,1}, \dots, i_{r,n}, i_d)$ рекомендуется применять итерационный метод Гаусса-Зейделя. Выбор алгоритма основан на следующих свойствах предлагаемой вычислительной процедуры. Реализация метода практически не зависит от размеров матрицы СУР. Метод Гаусса-Зейделя использует свойство разреженности мат-

рицы, т.е. наличие большого числа нулевых элементов в структуре. Наличие этой характеристики позволяет не хранить элементы матрицы в памяти компьютера, а вычислять их значения в момент обращения к матрице. Как будет показано далее реализация алгоритма Гаусса-Зейделя значительно упрощается при использовании рекурсивной формы представления СУР в виде системы соотношений (2.8), полученной в предыдущем подразделе. Основываясь на результатах [16,34,32–35], приведем основные расчетные выражения, которые используются при решении СУР (2.8) итерационным методом Гаусса-Зейделя.

Реализация метода Гаусса-Зейделя основана на построении последовательности приближений $P^{(j)}(i_{r,1}, \dots, i_{r,n}, i_d)$ к неизвестной вероятности $P(i_1, \dots, i_n, d)$ произвольного состояния исследуемой модели узла $(i_1, \dots, i_n, d)$ из множества S . Здесь j — порядковый номер приближения. Обозначим через $L(i_{r,1}, \dots, i_{r,n}, i_d)$ коэффициент при вероятности $P(i_{r,1}, \dots, i_{r,n}, i_d)$ в левой части (2.8), а через $R(i_{r,1}, \dots, i_{r,n}, i_d)$ обозначим правую часть (2.8). Используя введенные обозначения, СУР в форме (2.8) можно переписать в виде равенства

$$P(i_{r,1}, \dots, i_{r,n}, i_d) = \frac{R(i_{r,1}, \dots, i_{r,n}, i_d)}{L(i_{r,1}, \dots, i_{r,n}, i_d)}, \quad (i_1, \dots, i_n, d) \in S. \quad (3.1)$$

Соотношение (3.1) легко переписать в рекурсию Гаусса-Зейделя для вычисления значений $(j + 1)$ -го приближения к $P(i_{r,1}, \dots, i_{r,n}, i_d)$ по найденным к $(j + 1)$ -му шагу величинам (j) -го и $(j + 1)$ -го приближений к массиву вероятностей $P(i_{r,1}, \dots, i_{r,n}, i_d)$, где $(i_1, \dots, i_n, d) \in S$. Расчетное выражение для $P^{(j+1)}(i_{r,1}, \dots, i_{r,n}, i_d)$ для каждого фиксированного набора значений $i_{r,1}, \dots, i_{r,n}, i_d$ из приведенного ниже перечня

$$\begin{aligned}
i_{r,1} &= 0, 1, \dots, \left\lfloor \frac{v}{b_1} \right\rfloor; & i_{r,2} &= 0, 1, \dots, \left\lfloor \frac{v - i_{r,1}b_1}{b_2} \right\rfloor; \\
\vdots & & \vdots & \\
i_{r,n} &= 0, 1, \dots, \left\lfloor \frac{v - i_{r,1}b_1 - \dots - i_{r,n-1}b_{n-1}}{b_n} \right\rfloor; \\
i_d &= 0, 1, \dots, v + w - i_{r,1}b_1 - \dots - i_{r,n}b_n
\end{aligned}$$

имеет вид:

$$P^{(j+1)}(i_{r,1}, \dots, i_{r,n}, i_d) = \frac{R^{(j+1,j)}(i_{r,1}, \dots, i_{r,n}, i_d)}{L^{(j+1,j)}(i_{r,1}, \dots, i_{r,n}, i_d)}, \quad (i_1, \dots, i_n, d) \in S, \quad (3.2)$$

где функции $L^{(j+1,j)}(i_{r,1}, \dots, i_{r,n}, i_d)$ и $R^{(j+1,j)}(i_{r,1}, \dots, i_{r,n}, i_d)$ определяются из соотношений:

$$L^{(j+1,j)}(i_{r,1}, \dots, i_{r,n}, i_d) = \quad (3.3)$$

$$\begin{aligned}
&= \sum_{k=1}^n \left(\lambda_{r,k} I(\ell + i_d + b_k \leq v) + i_k \alpha_{r,k} I(i_{r,k} > 0) \right) + \\
&+ \lambda_d I(\ell + i_d + 1 \leq v + w) + (v - \ell) \alpha_d I(i_d > 0) + (\ell + i_d - v) \sigma I(\ell + i_d > v); \\
&R^{(j+1,j)}(i_{r,1}, \dots, i_{r,n}, i_d) = \\
&= \sum_{k=1}^n P^{(j+1,j)}(i_{r,1}, \dots, i_{r,k} - 1, \dots, i_{r,n}, i_d) \lambda_{r,k} I(i_{r,k} > 0, \ell + i_d \leq v) +
\end{aligned} \tag{3.4}$$

$$\begin{aligned}
& + \sum_{i=1}^{i_d} P^{(j+1,j)}(i_{r,1}, \dots, i_{r,n}, i_d - i) \lambda_d \left(f_i + I(\ell + i_d = v + w) \sum_{j=i+1}^{v+w} f_j \right) + \\
& + \sum_{k=1}^n P^{(j+1,j)}(i_{r,1}, \dots, i_{r,k} + 1, \dots, i_{r,n}, i_d) (i_{r,k} + 1) \alpha_{r,k} \times \\
& \times \left(I(\ell + b_k + i_d \leq v) + I(\ell + b_k + i_d > v, \ell + b_k \leq v, \ell + b_k + i_d \leq v + w) \right) + \\
& + P^{(j+1,j)}(i_{r,1}, \dots, i_{r,n}, i_d + 1) \times \left((v - \ell) \alpha_d I(\ell + i_d + 1 \leq v + w) + \right. \\
& \left. + (\ell + i_d + 1 - v) \sigma I(\ell + i_d + 1 \leq v + w, \ell + i_d + 1 > v) \right).
\end{aligned}$$

3.2.2 Повышение эффективности расчетной процедуры

В процессе вычисления $P^{(j+1)}(i_{r,1}, \dots, i_{r,n}, i_d)$, следует использовать найденные к этому моменту компоненты $(j+1)$ -го приближения $P^{(j+1)}(i_{r,1}, \dots, i_{r,n}, i_d)$. Если окажется, что эти компоненты еще не рассчитаны, то уже вычисленные компоненты j -го приближения $P^{(j)}(i_{r,1}, \dots, i_{r,n}, i_d)$. Реализация этого подхода позволяет использовать один массив $P(i_{r,1}, \dots, i_{r,n}, i_d)$ размерности $(n+1)$ для хранения в памяти компьютера компонент $(j+1)$ -го и (j) -го приближений. Число элементов массива равно числу состояний в пространстве S , определяемом соотношениями (2.1). Это свойство повышает эффективность рассмотренной вычислительной процедуры.

В качестве начального приближения в (3.1) можно взять следующий массив значений

$$P^{(0)}(i_{r,1}, \dots, i_{r,n}, i_d) = 1, \quad (i_{r,1}, \dots, i_{r,n}, i_d) \in S.$$

Опыт применения итерационного метода Гаусса–Зейделя к численному решению системы уравнений равновесия может приводить к сходящейся процедуре, а может и нет. Следует отметить, что в большинстве интересных с точки зрения практических приложений сходимость имеет место. Результат сходимости необходимо проверять косвенными методами. Для этого можно использовать два подхода. Первый подход традиционный. Он основан на достижении требуемой малости нормированной разности между двумя последовательными приближениями к вектору неизвестных вероятностей. Для этого проверяется выполнение соотношения

$$\frac{\sum_{(i_{r,1}, \dots, i_{r,n}, i_d) \in S} |P^{(j+1)}(i_{r,1}, \dots, i_{r,n}, i_d) - P^{(j)}(i_{r,1}, \dots, i_{r,n}, i_d)|}{\sum_{(i_{r,1}, \dots, i_{r,n}, i_d) \in S} P^{(j+1)}(i_{r,1}, \dots, i_{r,n}, i_d)} \leq \varepsilon. \quad (3.5)$$

Выбор значения ε обусловлен ожидаемой погрешностью оценки точных значений вероятностей стационарных состояний. Обычно ε выбирается из интервала $10^{-8} \dots 10^{-10}$. Итерационный цикл оканчивается после первого выполнения приведенного выше неравенства.

Второй подход основан на проверке выполнения каких-либо известных теоретических зависимостей между вычисленными значениями характеристик. К таковым, например, относятся выражения (2.6), полученные в результате преобразований СУР (2.8) или после применения формулы Литтла. Если сходимость достигнута, то равенства (2.6) должны

выполняться с точностью достигнутой в процессе реализации итерационной процедуры. Примерно она определяется используемым значением ε .

Данный результат представляет собой необходимое, но недостаточное условие сходимости применяемого метода последовательных замещений решения СУР. Как показано в [34], если сходимость в форме (3.1) отсутствует, систему (2.8) можно преобразовать к эквивалентному виду, для которого сходимость метод последовательных замещений Гаусса-Зейделя доказывается строго на основании свойств матрицы СУР. Для этого одну из неизвестных вероятностей в (2.8), например, $P(0, \dots, 0)$, фиксируют (скажем, равной единице) и исключают из системы уравнение, соответствующее этому состоянию. После таких преобразований сумма элементов хотя бы в одном столбце модифицированной матрицы СУР становится положительной. Это гарантирует слабое диагональное преобладание, что, согласно [34], обеспечивает сходимость итерационного процесса. Однако скорость сходимости в этом случае оказывается на порядки ниже, чем при использовании исходной формы (3.1). По этой причине подобная модификация применяется только в исключительных ситуациях, когда метод (3.1) расходится.

В качестве значений искомых вероятностей $P(i_{r,1}, \dots, i_{r,n}, i_d)$, $(i_{r,1}, \dots, i_{r,n}, i_d) \in S$ стационарных состояний модели $(i_{r,1}, \dots, i_{r,n}, i_d)$ берутся нормированные компоненты последнего найденного вектора последовательных приближений. Обозначим номер этого приближения через ℓ . Тогда

$$P(i_{r,1}, \dots, i_{r,n}, i_d) = \frac{P^{(\ell)}(i_{r,1}, \dots, i_{r,n}, i_d)}{\sum_{(i_{r,1}, \dots, i_{r,n}, i_d) \in S} P^{(\ell)}(i_{r,1}, \dots, i_{r,n}, i_d)},$$

$$(i_{r,1}, \dots, i_{r,n}, i_d) \in S.$$

Далее в соответствии с определениями, приведенными в подразделе 2.3.2, рассчитываются величины характеристик качества обслуживания поступающих запросов.

3.3 Оценка характеристик узла в частных случаях

3.3.1 Обслуживание файлов в режиме сервисов реального времени

Рассматриваемая модель узла доступа сохраняет все предположения о характере поступления и обслуживания заявок, сформулированные в подразделе 2.3 при построении обобщенной модели узла, за исключением процедуры распределения ресурса при обслуживании файлов. Теперь каждый поступивший файл обслуживается с использованием передаточных возможностей одного виртуального канала, т.е. по правилам обслуживания трафика сервисов реального времени. В данной модели сохраняются обозначения всех параметров и показателей качества обслуживания входящих заявок. Ниже представлены только те характеристики, формулировки которых отличаются от анализируемой комплексной модели узла доступа. Параметры исследуемой математической модели узла доступа представлены на рисунке 3.1.

Как уже было сказано, единственным отличием анализируемой модели от обобщенной модели узла, рассмотренной в подразделе 2.3, является процедура использования ресурса

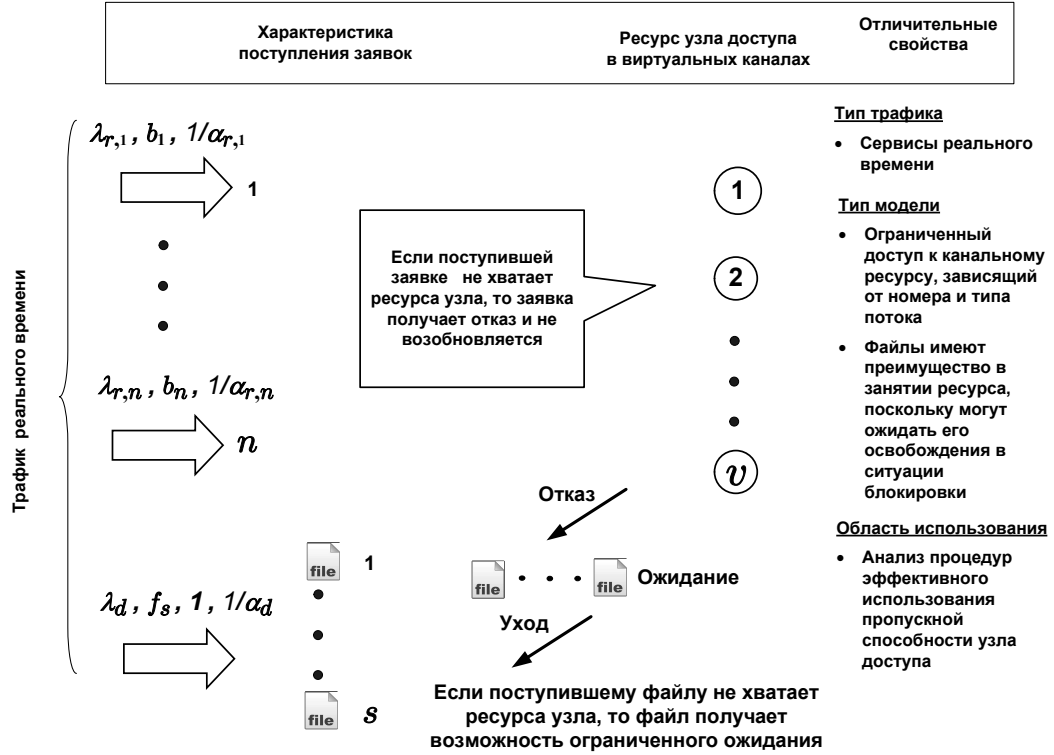


Рисунок 3.1 — Математическая модель узла доступа с обслуживанием файлов в режиме сервисов реального времени

узла при передаче поступившего файла. Теперь она выполняется только на основе возможностей одного виртуального канала.

Изменение состояний данной модели характеризуется марковским процессом $r(t) = (i_{r,1}(t), \dots, i_{r,n}(t), i_d(t))$, заданным на конечном множестве состояний S . Это множество состоит из векторов $(i_{r,1}, \dots, i_{r,n}, i_d)$, где каждая компонента изменяется в пределах, заданных условиями (2.1). Смысл компонент процесса $r(t)$ и состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$ совпадает с определением этих выражений, приведенными в подразделе 2.3. Обозначим через $p(i_{r,1}, \dots, i_{r,n}, i_d)$ вероятности стационарных состояний модели и приведем с их помощью определения основных характеристик обслуживания поступающих потоков заявок.

Характеристики обслуживания заявок k -го потока нагрузки онлайн-услуг определяются следующими выражениями:

- вероятность блокировки поступивших заявок — $\pi_{r,k}$;
- математическое ожидание количества используемых единиц ресурса передачи информации — $m_{r,k}$.

Качество обслуживания запросов на пересылку файлов, допускающих переменную скорость передачи характеризуется следующими показателями:

- $\pi_{d,t}$ — вероятность блокировки из-за недостаточности пропускной способности узла доступа для приема поступившего файла;

- $\pi_{d,c}$ — вероятность блокировки из-за недостаточности пропускной способности узла доступа для приема поступившего файла и нехватки в этой ситуации позиций для постановки файла на ожидание;
- π_d — вероятность блокировки файлов по всем причинам анализируемым в построенной модели узла доступа;
- m_d — математическое ожидание количества единиц ресурса узла доступа, используемых на пересылку файлов;
- L_q — математическое ожидание количества файлов, ожидающих начала обслуживания;
- L_s — математическое ожидание количества файлов, находящихся в состоянии передачи;
- $L = L_q + L_s$ — математическое ожидание количества файлов в системе (ожидание + обслуживание).

Значения введенных характеристик можно рассчитать, если известны величины предельных вероятностей $p(i_{r,1}, \dots, i_{r,n}, i_d)$ для всех состояний модели $(i_{r,1}, \dots, i_{r,n}, i_d) \in S$. Для состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$ обозначим через ℓ число единиц ресурса узла доступа, используемых на пересылку нагрузки онлайн-услуг

$$\ell = i_{r,1}b_1 + \dots + i_{r,n}b_n.$$

Предполагается, что максимальное число файлов m_g в группе не превышает $v + w$.

Формальные выражения для введенных выше характеристик $\pi_{r,k}$, $m_{r,k}$, $k = 1, \dots, n$, $\pi_{d,t}$, $\pi_{d,c}$, π_d , L_q , L , h_d через значения вероятностей $p(i_{r,1}, \dots, i_{r,n}, i_d)$ совпадают с выражениями для аналогичных по смыслу характеристик обобщенной модели узла доступа. Они сформулированы в подразделе 2.3.2 в виде соотношений (2.2), (2.4), (2.5), и по этой причине здесь не приводятся. Изменяются определения для m_d и L_s . Приведем эти выражения

$$\begin{aligned} m_d &= \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d \leq v\}} p(i_{r,1}, \dots, i_{r,n}, i_d) i_d + \\ &+ \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d > v\}} p(i_{r,1}, \dots, i_{r,n}, i_d) (v - \ell); \\ L_s &= m_d. \end{aligned} \tag{3.6}$$

Величины характеристик модели рассчитываются согласно введенным определениям с использованием значений $p(i_{r,1}, \dots, i_{r,n}, i_d)$. Для этого необходимо составить и решить систему уравнений равновесия. Действуя стандартным методом, подробно изложенным для обобщенной модели узла доступа, в подразделе 2.3, получаем такое соотношение

$$P(i_{r,1}, \dots, i_{r,n}, i_d) \left\{ \sum_{k=1}^n \left(\lambda_{r,k} I(\ell + i_d + b_k \leq v) + i_k \alpha_{r,k} I(i_{r,k} > 0) \right) + \right. \tag{3.7}$$

$$\begin{aligned}
& +\lambda_d I(\ell + i_d + 1 \leq v + w) + i_d \alpha_d I(\ell + i_d \leq v, i_d > 0) + \\
& + ((v - \ell) \alpha_d + (\ell + i_d - v) \sigma) I(\ell + i_d > v) \Big\} = \\
& = \sum_{k=1}^n P(i_{r,1}, \dots, i_{r,k} - 1, \dots, i_{r,n}, i_d) \lambda_{r,k} I(i_{r,k} > 0, \ell + i_d \leq v) + \\
& + \sum_{i=1}^{i_d} P(i_{r,1}, \dots, i_{r,n}, i_d - i) \lambda_d \left(f_i + I(\ell + i_d = v + w) \sum_{j=i+1}^{v+w} f_j \right) + \\
& + \sum_{k=1}^n P(i_{r,1}, \dots, i_{r,k} + 1, \dots, i_{r,n}, i_d) (i_{r,k} + 1) \alpha_{r,k} \times \\
& \times \left(I(\ell + b_k + i_d \leq v) + I(\ell + b_k + i_d > v, \ell + b_k \leq v) \right) + \\
& + P(i_{r,1}, \dots, i_{r,n}, i_d + 1) \times \left((i_d + 1) \alpha_d I(\ell + i_d + 1 \leq v) + \right. \\
& \left. + ((v - \ell) \alpha_d + (\ell + i_d + 1 - v) \sigma) I(\ell + i_d + 1 \leq v + w, \ell + i_d + 1 > v) \right).
\end{aligned}$$

Для значений $P(i_{r,1}, \dots, i_{r,n}, i_d)$ выполнено условие нормировки

$$\sum_{(i_{r,1}, \dots, i_{r,n}, i_d) \in S} P(i_{r,1}, \dots, i_{r,n}, i_d) = 1.$$

Здесь $I(\cdot)$ — индикаторная функция, определяемая соотношением (2.7).

3.3.2 Анализ эффективности дисциплины Processor Sharing при обслуживании эластичного трафика

Модель узла доступа, рассмотренная в подразделе 3.3.1, отличается от обобщенной модели узла изменением процедуры распределения ресурса при обслуживании поступающих файлов. Теперь каждый файл передается с использованием возможностей одного канала. Свойство эластичности при этом во внимание не принимается. Модифицированная подобным образом модель дает возможность провести исследование эффективности применения дисциплины Processor Sharing при обслуживании эластичного трафика.

Рассмотрим модель узла доступа со следующими заданными параметрами:

$$C = 50 \text{ Мбит/с};$$

$$n = 2;$$

$$c_1 = 1 \text{ Мбит/с};$$

$$c_2 = 2 \text{ Мбит/с};$$

$$c = 1 \text{ Мбит/с};$$

$$F = 1 \text{ Мбит.}$$

Отсюда:

$$v = 50;$$

$$b_1 = 1;$$

$$b_2 = 2;$$

$$\frac{1}{\alpha_d} = 1 \text{ секунда (эта величина принята за единицу времени)};$$

$$\begin{aligned}\alpha_d &= 1; \\ \alpha_{r,1} &= 0,1 \text{ и} \\ \alpha_{r,2} &= 0,1.\end{aligned}$$

Интенсивность предложенной нагрузки на один виртуальный канал ρ определяется по формуле:

$$\rho = \frac{1}{v} \left(\frac{\lambda_{r,1}}{\alpha_{r,1}} b_1 + \frac{\lambda_{r,2}}{\alpha_{r,2}} b_2 + \frac{\lambda_d}{\alpha_d} \bar{b} \right). \quad (3.8)$$

При расчетах предполагается, что все три потока вносят равный вклад в потенциальную загрузку узла. Отсюда интенсивности поступления запросов для каждого потока определяются как:

$$\lambda_{r,1} = \frac{v\rho\alpha_{r,1}}{3b_1}; \quad \lambda_{r,2} = \frac{v\rho\alpha_{r,2}}{3b_2}; \quad \lambda_d = \frac{v\rho\alpha_d}{3\bar{b}}. \quad (3.9)$$

Параметры группового поступления файлов определяются из выражений:

$$\begin{aligned}f_s &= 1/30, \quad s = 1, 2, \dots, 30 \\ (\bar{b} &= 15,5); \\ w &= 30; \quad \sigma = 0,1.\end{aligned}$$

Характеристики совместной передачи рассмотренных информационных потоков на рисунках обозначены символом RT, в ситуации, если пересылка файлов шла на условиях обслуживания трафика сервисов реального времени, т.е. с использованием возможностей одного канала и символом PS, если пересылка файлов шла по правилам дисциплины Processor Sharing, т.е. на всех свободных каналах. Для удобства обсуждения полученных результатов обозначим первый сценарий распределения ресурса передачи как сценарий RT, соответственно второй — это сценарий PS.

На рисунке 3.2 приведены значения $\pi_{r,1}$ доли потерянных заявок 1-го потока (левая часть) и значения $\pi_{r,2}$ доли потерянных заявок 2-го потока (правая часть) в зависимости от использования/не использования дисциплины Processor Sharing при обслуживании эластичного трафика и роста нагрузки на канал ρ .

Рисунок 3.3 построен с использованием тех же входных данных, что и рисунок 3.2, показаны значения π_d доли потерянных файлов (левая часть) и значения h_d среднего времени доставки файла (правая часть) в зависимости от использования/не использования дисциплины равномерного распределения ресурса (Processor Sharing) при обслуживании файлов, допускающих переменную скорость передачи и роста нагрузки на канал ρ .

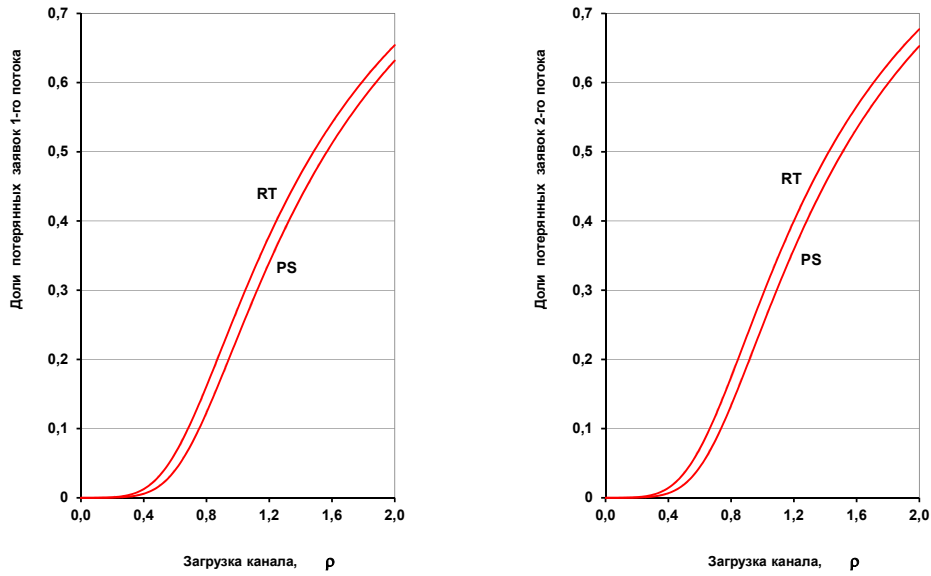


Рисунок 3.2 — Доля потерянных заявок двух потоков в зависимости от использования дисциплины Processor Sharing при обслуживании файлов, допускающих переменную скорость передачи и роста нагрузки на канал ρ

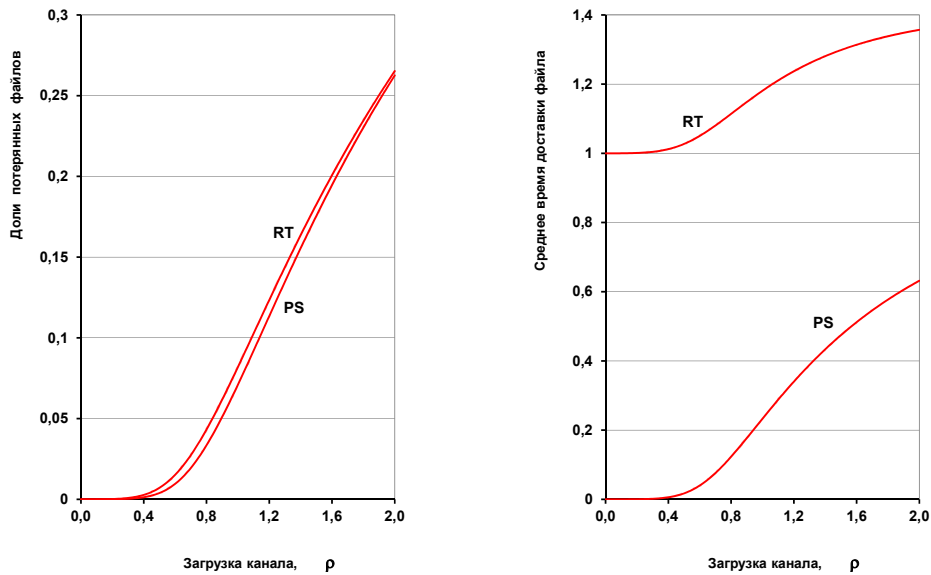


Рисунок 3.3 — Значения π_d доли потерянных файлов (левая часть) и значения h_d среднего времени доставки файла в зависимости от использования/не использования дисциплины Processor Sharing при обслуживании файлов, допускающих переменную скорость передачи и роста нагрузки на канал ρ

На рисунке 3.4 представлены результаты для тех же входных параметров, что и на

рисунках 3.2 и 3.3. На левой части графика отображена доля времени занятости канала $\delta = \frac{m_{r,1}b_1 + m_{r,2}b_2 + m_d}{v}$ в зависимости от применения (или отсутствия) дисциплины равномерного распределения ресурса (Processor Sharing) при обслуживании файлов, допускающих переменную скорость передачи. На правой части показано среднее число каналов b_f , задействованных при передаче одного файла с использованием Processor Sharing. Зависимости построены при варьировании нагрузки на канал ρ .

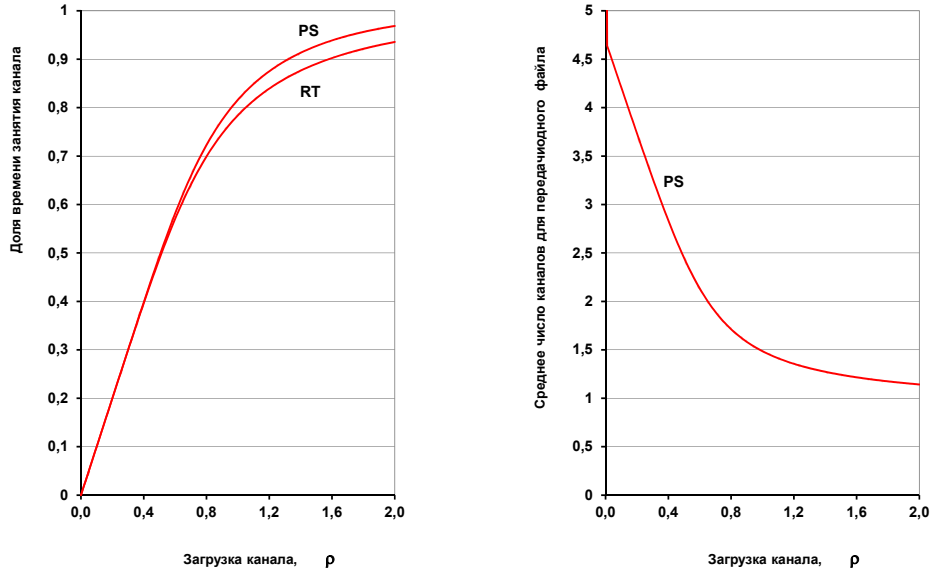


Рисунок 3.4 — Значения δ доли времени занятости канала (левая часть) в зависимости от использования/не использования дисциплины равномерного распределения ресурса (Processor Sharing) при обслуживании файлов, допускающих переменную скорость передачи и роста нагрузки на канал ρ и значения b_f среднего числа каналов, используемых при передаче одного файла, и использования дисциплины Processor Sharing (правая часть)

Из анализа численных данных, показанных на рисунках 3.2–3.4, можно сделать следующие выводы:

1. Во всех случаях характеристики, полученные в результате реализации сценария RT, дают оценки сверху для соответствующих характеристик, полученных в процессе реализации сценария PS. Этому факту можно дать следующее объяснение. Ускоренная передача файлов в режиме PS приводит к более раннему уходу файлов из обслуживания. Освобождение ресурса передачи дает больше шансов всем поступающим в узел доступа заявкам попасть на обслуживание и как следствие уменьшает потери всех видов запросов и увеличивает коэффициент занятия канала δ . Это свойство позволяет использовать характеристики модели с режимом RT для оценки характеристик модели с обслуживанием файлов в режиме PS. Данный результат применяется в подразделе 4.4 при формулировке основных этапов планирования ресурса узла доступа.

2. Режим PS демонстрирует наибольшую эффективность при минимальных потерях, когда файл передаётся посредством большого числа свободных каналов. Однако его производительность снижается в условиях значительных потерь, когда, в соответствии с распределением ресурсов в узле доступа, передача данных фактически осуществляется через один канал. Первое из отмеченных свойств будет использовано в подразделе 4.4 при разработке приближенного метода оценки характеристик совместной передачи эластичных файлов и трафика сервисов реального времени, основанного на технике декомпозиции.
3. Интересные свойства применения режимов RT и PS показывает диаграмма зависимости от ρ значения доли времени занятости канала. Достаточно в большом интервале изменения ρ от нуля и далее значения δ для режимов RT и PS практически одинаковы. Это связано с тем, что в этой области изменения параметров модели ускорение передачи файлов приводит к пропорциональному уменьшению времени их пребывания на обслуживании. Таким образом, работа узла доступа по пересылке файлов, которая характеризуется долей времени занятости канала, меняется не столь значительно.
4. Минимальное среднее время доставки файла в режиме PS наблюдается при стремлении ρ к нулю и составляет $\frac{1}{v}$ секунд. В то же время для режима RT в аналогичных предельных условиях это время достигает 1 секунды.
5. Максимальное среднее число виртуальных каналов для передачи файла достигается при $\rho \rightarrow 0$ и составляет v каналов.

3.3.3 Обслуживание файлов в режиме сервисов реального времени без ожидания

Модель узла доступа построенная и исследованная в предыдущем подразделе, дает возможность анализировать эффективность дисциплины PS. В последующих разделах диссертации она будет использоваться при разработке приближенных методов оценки характеристик обобщенной модели узла доступа. Недостатком модели являются трудности ее численного анализа. Значения характеристик приходится рассчитывать с использованием итерационного метода Гаусса-Зейделя. Это увеличивает время оценки показателей совместного обслуживания заявок, что является критичным при решении задачи планирования пропускной способности узла доступа. В данном подразделе будет рассмотрен частный случай этой модели, который лишен отмеченного недостатка. Для оценки характеристик здесь используется рекурсивный алгоритм.

Представленная модель узла доступа учитывает все допущения, принятые при поступлении и обработке заявок, описанные в подразделе 3.3.1 при создании исходной модели. Однако в ней модифицирован механизм ожидания для файлов, получивших отказ. В данном случае файлы, которым отказано в обслуживании, теряются без возможности повторной обработки. При этом в модели сохранены все ранее введенные обозначения параметров и показателей качества обслуживания заявок. Ниже приведены только те расчётные

выражения для характеристик, которые были изменены по сравнению с моделью из подраздела 3.3.1. Параметры данной математической модели узла доступа представлены на рисунке 3.5.

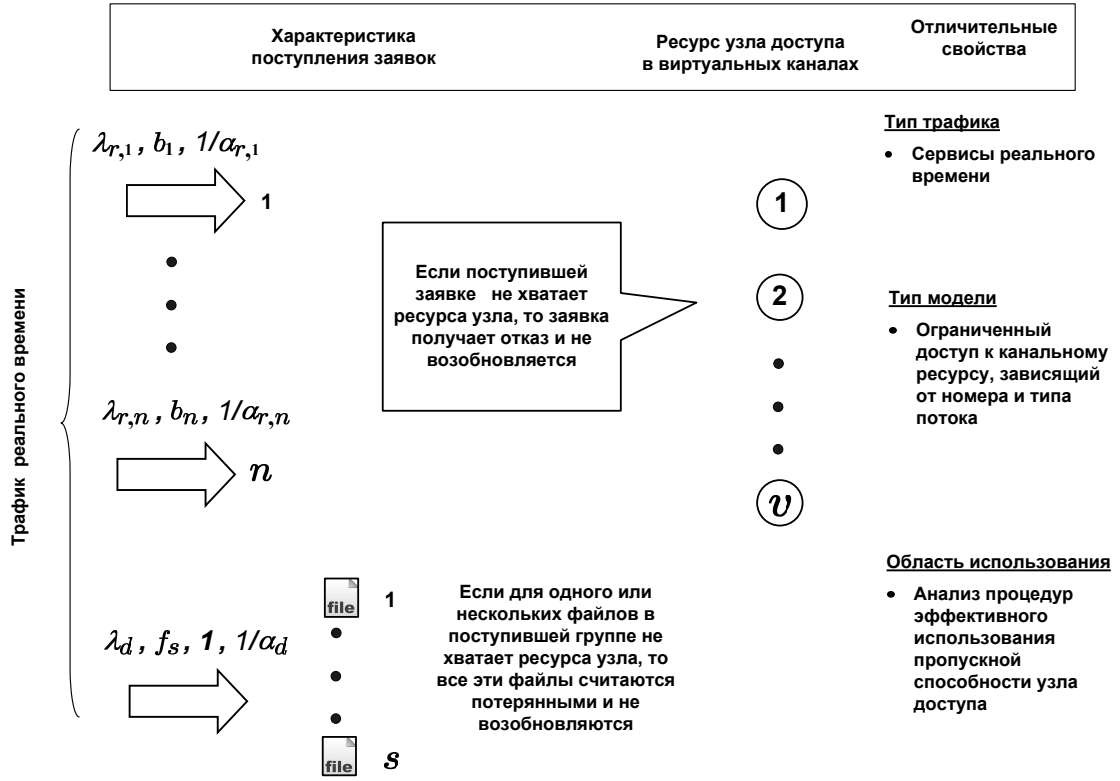


Рисунок 3.5 — Математическая модель узла доступа с обслуживанием файлов в режиме сервисов реального времени без возможности ожидания в ситуации отказа в обслуживании

Изменение состояний данной модели характеризуется марковским процессом $r(t) = (i_{r,1}(t), \dots, i_{r,n}(t), i_d(t))$, заданным на конечном множестве состояний S , которое включает в себя состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$, с компонентами, меняющимися в пределах, определенных соотношениями (2.1) для $w = 0$. Смысл компонент процесса $r(t)$ и состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$ совпадает с определением этих выражений, приведенными в подразделе 2.3. Обозначим через $p(i_{r,1}, \dots, i_{r,n}, i_d)$ вероятности стационарных состояний модели и приведем с их помощью определения основных характеристик обслуживания поступающих потоков заявок.

Характеристики обслуживания заявок k -го потока нагрузки онлайн-услуг определяются следующими выражениями:

- вероятность блокировки поступивших заявок — $\pi_{r,k}$;
- математическое ожидание количества используемых единиц ресурса передачи ин-

формации — $m_{r,k}$.

Качество обслуживания запросов на пересылку файлов, допускающих переменную скорость передачи характеризуется следующими показателями:

- $\pi_{d,t}$ — вероятность блокировки из-за недостаточности пропускной способности узла доступа для приема поступившего файла;
- π_d — вероятность блокировки файлов по всем причинам анализируемым в построенной модели узла доступа;
- m_d — математическое ожидание количества единиц ресурса узла доступа, используемых на пересылку файлов;

Для расчета этих характеристик необходимо знать значения стационарных вероятностей $p(i_{r,1}, \dots, i_{r,n}, i_d)$ всех возможных состояний модели $(i_{r,1}, \dots, i_{r,n}, i_d) \in S$. Введем обозначение: для состояния $(i_{r,1}, \dots, i_{r,n}, i_d)$ через ℓ обозначим количество виртуальных каналов, занятых трафиком сервисов реального времени: $\ell = i_{r,1}b_1 + \dots + i_{r,n}b_n$. Также предположим, что максимальное число файлов m_g в поступающей группе равно v .

Формальные выражения для введенных выше характеристик $\pi_{r,k}$, $m_{r,k}$, $k = 1, \dots, n$ через значения вероятностей $p(i_{r,1}, \dots, i_{r,n}, i_d)$ совпадают с выражениями для аналогичных по смыслу характеристик обобщенной модели узла доступа. Они сформулированы в подразделе 2.3.2 в виде соотношений (2.2) и по этой причине здесь не приводятся. Изменяются определения для $\pi_{d,t}$, π_d , m_d . Приведем эти выражения

$$\pi_{d,t} = \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d = v\}} p(i_{r,1}, \dots, i_{r,n}, i_d); \quad (3.10)$$

$$\pi_d = \frac{1}{b} \times \sum_{i=0}^{i_d} \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d = v\}} p(i_{r,1}, \dots, i_{r,n}, i_d - i) \sum_{s=i+1}^v f_s(s - i).$$

$$m_d = \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid \ell + i_d \leq v\}} p(i_{r,1}, \dots, i_{r,n}, i_d) i_d;$$

Величины характеристик модели рассчитываются согласно введенным определениям с использованием значений $p(i_{r,1}, \dots, i_{r,n}, i_d)$. Для этого необходимо составить и решить систему уравнений равновесия. Действуя стандартным методом, подробно изложенным для обобщенной модели узла доступа, в подразделе 2.4, получаем такое соотношение

$$P(i_{r,1}, \dots, i_{r,n}, i_d) \left\{ \sum_{k=1}^n \left(\lambda_{r,k} I(\ell + i_d + b_k \leq v) + i_k \alpha_{r,k} I(i_{r,k} > 0) \right) + \right. \quad (3.11)$$

$$\left. + \lambda_d I(\ell + i_d + 1 \leq v) + i_d \alpha_d I(i_d > 0) \right\} =$$

$$\begin{aligned}
&= \sum_{k=1}^n P(i_{r,1}, \dots, i_{r,k} - 1, \dots, i_{r,n}, i_d) \lambda_{r,k} I(i_{r,k} > 0) + \\
&+ \sum_{i=1}^{i_d} P(i_{r,1}, \dots, i_{r,n}, i_d - i) \lambda_d \left(f_i + I(\ell + i_d = v) \sum_{j=i+1}^v f_j \right) + \\
&+ \sum_{k=1}^n P(i_{r,1}, \dots, i_{r,k} + 1, \dots, i_{r,n}, i_d) (i_{r,k} + 1) \alpha_{r,k} I(\ell + b_k + i_d \leq v) + \\
&+ P(i_{r,1}, \dots, i_{r,n}, i_d + 1) (i_d + 1) \alpha_d I(\ell + i_d + 1 \leq v).
\end{aligned}$$

Для значений $P(i_{r,1}, \dots, i_{r,n}, i_d)$ выполнено условие нормировки

$$\sum_{(i_{r,1}, \dots, i_{r,n}, i_d) \in S} P(i_{r,1}, \dots, i_{r,n}, i_d) = 1.$$

Здесь $I(\cdot)$ — индикаторная функция, определяемая соотношением (2.7).

Для построенной модели оценка характеристик может быть выполнена не только путем решения СУР (3.11), но и другим более эффективным способом. Обозначим через $p(i)$ вероятность занятости i виртуальных каналов всеми типами запросов, поступающих в узел доступа

$$p(i) = \sum_{\{(i_{r,1}, \dots, i_{r,n}, i_d) \in S \mid i_{r,1}b_1 + \dots + i_{r,n}b_n + i_d = i\}} p(i_{r,1}, \dots, i_{r,n}, i_d), \quad (i_{r,1}, \dots, i_{r,n}, i_d) \in S.$$

Выполнив ряд алгебраических преобразований СУР (3.11) и воспользовавшись результатами [57], можно показать, что для ненормированных значений $P(i)$ выполняются следующие рекурсивные соотношения

$$\begin{aligned}
P(i) = \frac{1}{i} \left\{ \sum_{k=1}^n a_{r,k} b_k P(i - b_k) + a_d \left(P(0)(f_i + f_{i+1} + \dots + f_v) + \right. \right. \\
\left. \left. + P(1)(f_{i-1} + f_i + \dots + f_v) + \dots + P(i-1)(f_1 + f_2 + \dots + f_v) \right) \right\},
\end{aligned} \tag{3.12}$$

где $a_{r,k} = \frac{\lambda_{r,k}}{\mu_{r,k}}$, $k = 1, 2, \dots, n$, $a_d = \frac{\lambda_d}{\mu_d}$. В (3.12) величина i меняется от 0 до v .

Полученная форма рекурсии позволяет рассчитать все значения $p(i)$, $i = 1, \dots, v$, используя следующий алгоритм. Он состоит из следующих этапов.

1. Примем начальное значение $P(0) = 1$.
2. Для каждого i от 1 до v вычислим ненормированные вероятности $P(i)$ с помощью рекуррентного соотношения

$$\begin{aligned}
P(i) = \frac{1}{i} \times \left\{ \sum_{k=1}^n a_{r,k} b_k P(i - b_k) + \right. \\
\left. + a_d \left(P(0) \sum_{j=i}^v f_j + P(1) \sum_{j=i-1}^v f_j + \dots + P(i-1) \sum_{j=1}^v f_j \right) \right\}.
\end{aligned} \tag{3.13}$$

3. Определяем нормировочную константу и нормированные вероятности $p(i)$:

$$N = \sum_{i=0}^v P(i), \quad p(i) = \frac{P(i)}{N}, \quad i = 0, 1, \dots, v.$$

Из вида рекурсивного выражения следует, что характеристики модели зависят не от конкретных значений $\lambda_{r,k}$, $\mu_{r,k}$, $k = 1, \dots, n$ и λ_d , α_d а от их отношений $a_{r,k}$, $k = 1, \dots, n$ и a_d .

Покажем, что знание величин $p(i)$, $i = 0, 1, \dots, v$ достаточно для вычисления основных характеристик совместного обслуживания поступающих запросов. Приведем формулы для показателей обслуживания трафика сервисов реального времени

$$\pi_{r,k} = \sum_{i=v-b_k+1}^v p(i); \quad m_{r,k} = a_{r,k} b_k (1 - \pi_{r,k}), \quad k = 1, 2, \dots, n. \quad (3.14)$$

Перейдем к оценке показателей обслуживания поступающих групп файлов. Вероятность потерь по времени $\pi_{d,t} = p(v)$. Вероятность потери файлов π_d определяется как частное от деления интенсивности потерянных файлов на интенсивность поступивших файлов. В состоянии (i) узел доступа имеет $v - i$ свободных виртуальных каналов. Поступление в это состояние узла группы из $v - i + 1$, $v - i + 2, \dots, v$ файлов, приведет к потере соответственно $1, 2, \dots, i$ файлов. Среднее число файлов, потерянных во всех состояниях (i) , $i = 1, \dots, v$ равно

$$\sum_{i=1}^v p(i) (f_{v-i+1} \cdot 1 + f_{v-i+2} \cdot 2 + \dots + f_v \cdot i) = \sum_{i=1}^v \sum_{k=0}^{i-1} f_{v-k} (i - k).$$

В результате, для оценки доли потерянных файлов получаем такое соотношение

$$\pi_d = \frac{1}{\bar{b}} \sum_{i=1}^v p(i) \sum_{k=0}^{i-1} f_{v-k} (i - k). \quad (3.15)$$

Величина m_d определяется из соотношения

$$m_d = a_d \bar{b} (1 - \pi_d). \quad (3.16)$$

Реализация полученных рекурсивных соотношений при оценке характеристик модели не вызывает затруднений. Как отмечалось ранее при рассмотрении численных данных на рисунках 3.2–3.4, выведенные в этом разделе формулы для расчёта характеристик передачи трафика сервисов реального времени и файловых данных представляют собой верхние оценки соответствующих параметров обслуживания запросов в обобщённой модели узла доступа, которая была введена и изучена в разделе 2. Это свойство далее будет использовано при разработке приближенных методов оценки характеристик обобщенной модели, основанных на реализации техники декомпозиции.

В двух последующих подразделах анализируются два специальных случая обобщенной модели. Первая модель изучает процесс обслуживания исключительно трафика сервисов реального времени, а вторая — только эластичного трафика данных с учетом возможного устаревания передаваемой информации.

3.3.4 Анализ характеристик трафика реального времени

Как отмечалось ранее, данный частный случай вытекает из обобщенной модели и из модели узла, рассмотренной в предыдущем подразделе, при условии $\lambda_d = 0$ в обеих моделях. Полученная в результате этих действий модель хорошо известна в теории телетрафика и носит название мультисервисной модели Эрланга [16,34,57]. Параметры анализируемой математической модели узла доступа, показаны на рисунке 3.6.

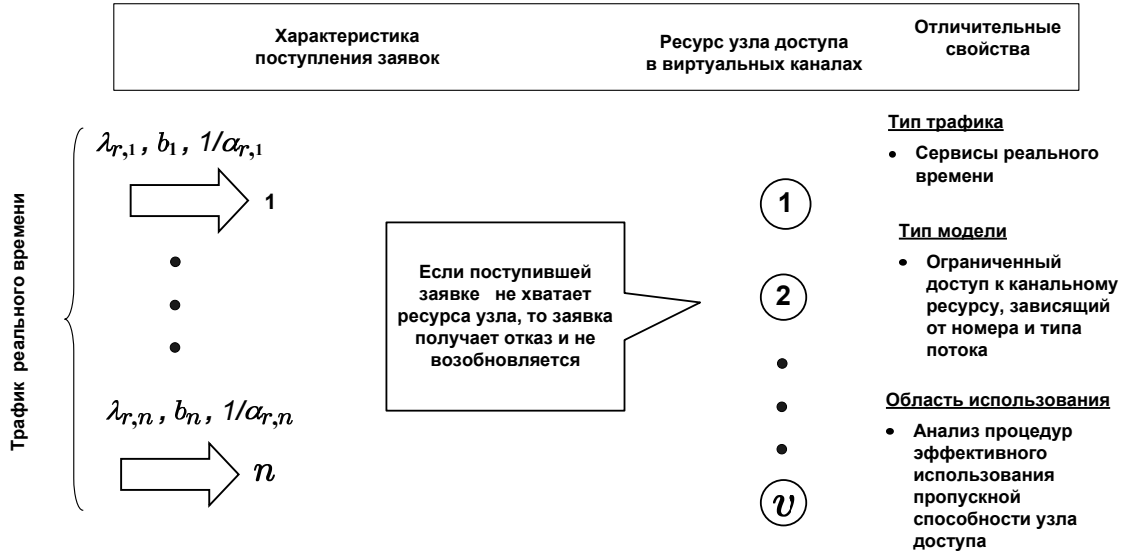


Рисунок 3.6 — Математическая модель узла доступа в отсутствии запросов на передачу файлов

Характеристики качества обслуживания трафика сервисов реального времени $\pi_{r,k}$, $m_{r,k}$, $k = 1, \dots, n$, находятся из выражений (3.14), после подстановки в (3.14) значений $p(i)$, полученных после реализации рекурсии (3.12) при $a_d = 0$. Алгоритм вычисления $p(i)$ имеет вид [16,34,32–35]:

1. Примем $P(0) = 1$.
2. Для каждого i от 1 до v последовательно рассчитываем $P(i)$ по рекуррентной формуле:

$$P(i) = \frac{1}{i} \left\{ \sum_{k=1}^n a_{r,k} b_k P(i - b_k) \right\}.$$

3. Вычисляем нормировочную константу N и нормированные вероятности $p(i)$:

$$N = \sum_{i=0}^v P(i), \quad p(i) = \frac{P(i)}{N}, \quad i = 0, 1, \dots, v.$$

4. Характеристики обслуживания запросов на передачу трафика сервисов реального времени рассчитываются по формулам (3.14).

3.3.5 Оценка характеристик эластичного трафика

Рассматриваемый частный случай вытекает из комплексной модели узла доступа, а также модели узла, введенной и проанализированной в предыдущем подразделе. Чтобы ее получить достаточно принять, что интенсивности поступления запросов на передачу нагрузки онлайн-услуг равны нулю $\lambda_{r,k} = 0$, $k = 1, 2, \dots, n$. Математическая формулировка поступления и обслуживания групп файлов данных, допускающих переменную скорость передачи, соответствует изложению в подразделе 2.2.3. Эту часть описания модели пропустим сохранив обозначения для параметров формирования потоков файлов. Перейдем к оценке характеристик их обслуживания. Обозначим их теми же символами, что были использованы для обобщенной модели и найдем выражения для их оценки. В отличие от обобщенной модели, где значения характеристик находятся из решения системы уравнений итерационным методом Гаусса-Зейделя, в рассматриваемом частном случае величины характеристик получаются из простых рекурсивных формул. Параметры анализируемой математической модели узла доступа, показаны на рисунке 3.7.

Пусть $S = \{(i) | i = 0, 1, \dots, v + w\}$ — множество возможных состояний данной модели. Изменение состояний во времени определяется случайным процессом $r(t) = i(t)$, где $i(t)$ — число файлов, которые в момент времени t либо передаются, либо ожидают начала обслуживания. Поскольку все случайные величины в модели независимы и имеют экспоненциальное распределение, процесс $r(t)$ является марковским. Обозначим стационарные вероятности состояний модели через $p(i)$. Эти вероятности отражают долю времени, которое система находится в состоянии с i файлами (передаваемыми или ожидающими в очереди), и позволяют исследовать параметры обслуживания поступающих файлов. Будем считать, что предельное количество файлов в группе определяется как $m_g = v + w$.

Предположим, что вероятности $p(i)$ известны, и на их основе выведем выражения для оценки параметров обслуживания эластичного трафика. В практических приложениях модели наиболее важны следующие показатели (см. подраздел 2.2.5):

- $\pi_{d,t}$ — вероятность того, что ресурс узла доступа будет недоступен для приёма файла в момент его поступления;
- $\pi_{d,c}$ — вероятность потери файла из-за недоступности ресурса узла доступа для его приёма, передачи или ожидания;
- π_d — общая вероятность потери файла с учётом всех факторов, рассматриваемых в модели;

- b_f — средний объём ресурса, требуемый для передачи одного файла (в виртуальных каналах);
- c_d — средний объём ресурса, необходимый для передачи одного файла (в бит/с).

Вывод расчетных выражений для введенных характеристик следует преобразованиям, выполненным в подразделе 2.2.5, по этой причине эту часть исследования не будем приводить. Сформулируем окончательные расчетные выражения:

$$\begin{aligned}
 \pi_{d,t} &= \sum_{i=v}^{v+w} p(i); \\
 \pi_{d,c} &= \frac{1}{b} \left(\sum_{i=1}^{v+w} p(i) \sum_{k=0}^{i-1} f_{v+w-k}(i-k) \right); \\
 \pi_d &= \frac{1}{\lambda b} \left(\lambda \sum_{i=1}^{v+w} p(i) \sum_{k=0}^{i-1} f_{v+w-k}(i-k) + \sum_{i=v+1}^{v+w} p(i)(i-v)\sigma \right); \\
 L_q &= \sum_{i=v+1}^{v+w} p(i)(i-v), \quad L = \sum_{i=1}^{v+w} p(i)i, \quad L_s = \sum_{i=1}^v p(i)i + \sum_{i=v+1}^{v+w} p(i)v; \\
 h_d &= \frac{L}{\lambda b(1 - \pi_{d,c})}; \quad m_d = \sum_{i=1}^v p(i)v + \sum_{i=v+1}^{v+w} p(i)v = 1 - p(0)v; \\
 z_d &= m_d c; \quad b_f = \frac{m_d}{L_s}; \quad c_d = \frac{z_d}{L_s}.
 \end{aligned} \tag{3.17}$$

Чтобы оценить параметры обслуживания файлов, заданные соотношениями (3.17), требуется построить и решить систему уравнений статистического баланса. Применяя классическую методику (как, например, в работе [34]), приходим к следующей системе уравнений:

$$\begin{aligned}
 P(0)\lambda &= P(1)v\mu_d, \quad i = 0; \\
 P(1)(\lambda + v\mu_d) &= P(0)\lambda f_1 + P(2)v\mu_d, \quad i = 1; \\
 P(2)(\lambda + v\mu_d) &= (P(0)f_2 + P(1)f_1)\lambda + P(3)v\mu_d, \quad i = 2; \\
 P(i)(\lambda + v\mu_d) &= (P(0)f_i + P(1)f_{i-1} + \dots + P(i-1)f_1)\lambda + P(i+1)v\mu_d, \\
 i &= 3, 4, \dots, v-1; \\
 P(v)(\lambda + v\mu_d) &= (P(0)f_v + P(1)f_{v-1} + \dots + P(v-1)f_1)\lambda + \\
 &\quad + P(v+1)(v\mu_d + \sigma), \quad i = v; \\
 P(v+1)(\lambda + v\mu_d + \sigma) &= (P(0)f_{v+1} + P(1)f_v + \dots + P(v)f_1)\lambda + \\
 &\quad + P(v+2)(v\mu_d + 2\sigma), \quad i = v+1; \\
 &\dots\dots\dots \\
 P(v+w-1)(\lambda + v\mu_d) + (w-1)\sigma &=
 \end{aligned} \tag{3.18}$$

$$\begin{aligned}
& + (P(0)f_{v+w-1} + P(1)f_{v+w-2} + \dots + P(v+w-2)f_1)\lambda + \\
& + P(v+w)(v\mu_d + w\sigma), \quad i = v+w-1; \\
P(v+w)(v\mu_d + w\sigma) & = (P(0)f_{v+w} + P(1)(f_{v+w-1} + f_{v+w}) + \dots + \\
& + P(v+w-1)(f_1 + f_2 + \dots + f_{v+w}))\lambda, \quad i = v+w.
\end{aligned}$$

Величины $P(i)$ удовлетворяют условию нормировки

$$\sum_{i=0}^{v+w} P(i) = 1.$$

Разработанная структура системы уравнений равновесия позволяет путем алгебраических преобразований построить рекурсивный алгоритм оценки значений вероятностей отдельных состояний модели. Выполним необходимые действия. Просуммируем (3.18) по i от 0 до $j-1$. После сокращения одинаковых слагаемых получаем соотношение следующего вида:

$$\begin{aligned}
\lambda \sum_{i=0}^{j-1} P(i) + \sum_{i=0}^{j-1} P(i)v\mu_d + \sum_{i=v+1}^{j-1} P(i)(i-v)\sigma & = \\
= \lambda \sum_{i=1}^{j-1} \sum_{k=0}^{i-1} P(k)f_{i-k} + \sum_{i=0}^j P(i)v\mu_d + \sum_{i=v+1}^j P(i)(i-v)\sigma.
\end{aligned} \tag{3.19}$$

В приведенном выражении преобразуем двойную сумму в правой части

$$\begin{aligned}
& P(0)f_1 + (P(0)f_2 + P(1)f_1) + (P(0)f_3 + P(1)f_2 + P(2)f_1) + \dots + \\
& + (P(0)f_{j-1} + P(1)f_{j-2} + \dots + P(j-2)f_1) = \\
& = P(0)(f_1 + f_2 + \dots + f_{j-1}) + \\
& + P(1)(f_1 + f_2 + \dots + f_{j-2}) + \dots + P(j-2)f_1 = \\
& = P(0) \sum_{i=1}^{j-1} f_i + P(1) \sum_{i=1}^{j-2} f_i + \dots + P(j-2) \sum_{i=1}^1 f_i.
\end{aligned}$$

Подставив найденное выражение в (3.19) и воспользовавшись соотношением $f_1 + \dots + f_{v+w} = 1$, получаем после приведения подобных членов рекурсивное соотношение, связывающее значения $P(j)$ ненормированных вероятностей отдельных состояний модели

$$\begin{aligned}
& (v\mu_d + (j-v)\sigma I(j > v))P(j) = \\
& = \lambda \left(P(0) \sum_{i=j}^{v+w} f_i + P(1) \sum_{i=j-1}^{v+w} f_i + \dots + P(j-1) \sum_{i=1}^{v+w} f_i \right).
\end{aligned} \tag{3.20}$$

Здесь $I(\cdot)$ — индикаторная функция, заданная выражением (2.7).

Алгоритм вычисления $p(j)$ включает в себя такие этапы.

1. Принимаем $P(0) = 1$.

2. Для каждого j от 1 до $v + w$ вычисляем ненормированные вероятности $P(j)$ по формуле:

$$P(j) = \frac{\lambda}{v\mu_d + (j - v)\sigma I(j > v)} \times \left(P(0) \sum_{i=j}^{v+w} f_i + P(1) \sum_{i=j-1}^{v+w} f_i + \dots + P(j-1) \sum_{i=1}^{v+w} f_i \right), \quad (3.21)$$

которая выводится из (3.20).

3. Находим константу нормировки:

$$N = \sum_{j=0}^{v+w} P(j).$$

4. Нормируем вероятности:

$$p(j) = \frac{P(j)}{N}, \quad j = 0, 1, \dots, v + w.$$

5. По формулам (3.17) определяем показатели эффективности обработки файлов.

3.4 Приближенная оценка характеристик передачи эластичного трафика

Разработка инженерных методик решения задачи планирования необходимого объема ресурса узла доступа основана на использовании достаточно обоснованных приближенных алгоритмов оценки характеристик узла. Предложенный метод основан на технике декомпозиции исходной модели на отдельные модули, где каждый модуль представлен и исследован в подразделе 3.3, в качестве частного случая. При построении расчетного алгоритма будем опираться на свойства модели, которые появляются у ее характеристик при формулировке задачи планирования. Главное, на что надо обратить внимание — это то обстоятельство, что оценка необходимого ресурса ведется в области малых потерь.

Для оценки среднего времени передачи файла предлагается следующий подход. Сначала определяется количество виртуальных каналов m_r , занятых трафиком сервисов реального времени. Затем анализируются характеристики передачи эластичного трафика на оставшихся свободных каналах $[v - m_r]$. В расчетах используются данные из подраздела 3.3.5. Среди них определения характеристик (3.17) и процедура их расчета (3.21). Точность оценки определяется уровнем потерь заявок на передачу трафика онлайн-услуг: чем они ниже, тем выше точность. Потери связаны с параметром ρ — потенциальной нагрузкой на канал, определяемой формулой (3.8).

В расчетах применялись следующие параметры: $C = 100$ Мбит/с; $n = 2$; $c_1 = 5$ Мбит/с; $c_2 = 10$ Мбит/с; $c = 1$ Мбит/с; $F = 1$ Мбит. Отсюда $v = 100$ к.е.; $b_1 = 5$ к.е., $b_2 = 10$ к.е.; $\frac{1}{\alpha_d} = 1$ сек. Среднее время передачи файла принято за единицу времени. Отсюда $\alpha_d = 1$. Для трафика онлайн-услуг параметры обслуживания заданы как $\alpha_{r,1} = 0,1$ и $\alpha_{r,2} = 0,1$. Характеристики группового поступления файлов:

- Вероятность $f_s = 1/30$ для $s = 1, 2, \dots, 30$ (среднее $\bar{b} = 15,5$).
- Параметры $w = 10$, $\sigma = 0,1$.

Интенсивности запросов для каждого потока вычисляются по формулам (3.9).

На рисунке 3.8 (левая часть) показано изменение значения h_d среднего времени пребывания файла в узле доступа, найденное точно и приближено. В последнем случае используется приближенное значение t_r , полученное в результате реализации рекурсии (3.13), без учета наличия потока эластичных данных. Ступенчатый характер кривых связан с целочисленным округлением t_r . Результаты вычислений говорят о хорошей точности расчетной процедуры особенно в области малых потерь, которая, как уже говорилось ранее, соответствует характеристикам обслуживания заявок используемым при решении задач планирования. Эти величины соответствуют значениям максимальных потерь на уровне 0,01–0,05.

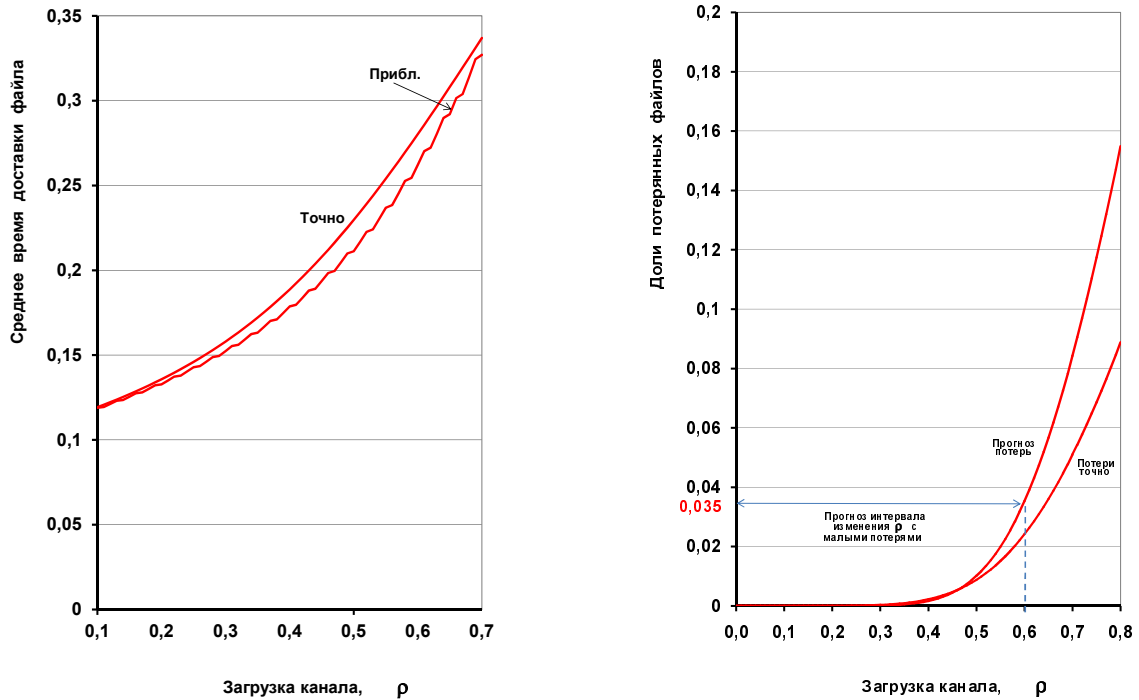


Рисунок 3.8 — Зависимость точного и приближенного расчета h_d среднего времени пребывания файла в узле доступа от изменения ρ (левая часть). Прогноз интервала изменения ρ с малыми (на уровне 0,035) потерями (правая часть)

Рассмотрим теперь, как можно предварительно оценить интервал изменения ρ , где потери поступающих заявок будут достаточно малы, чтобы иметь право воспользоваться при оценке характеристик передачи файлов построенной выше вычислительной процедурой. Для этого воспользуемся результатами подраздела 3.3.1, где численно показано,

что изменение процедуры распределения ресурса при обслуживании файлов с Processor Sharing на дисциплину обслуживания трафика сервисов реального времени (т.е. передача файлов ведется на минимальном ресурсе в один канал) при сохранении прочих параметров потоков приводит к оценкам сверху для величин потерь. Эти оценки могут быть рассчитаны с использованием рекурсии (3.13). При этом групповой поток файлов меняется на пуассоновский ординарный поток файлов интенсивности $\lambda_d \bar{b}$. Результаты реализации этой процедуры показаны на рисунке 3.8 (правая часть). Приведены значения максимальных потерь, найденные точно из решения СУР (2.8), и полученные в результате сформулированной приближенной процедуры. Оценивалась область изменения ρ , соответствующая значениям на уровне 0,035.

Результаты численного анализа, частично представленные на рисунке 3.8, показывают, что предложенный метод обладает приемлемой для практических приложений точностью. Разработанные процедуры будут использованы в разделе 4 при решении задачи планирования необходимой величины ресурса узла доступа.

3.5 Выводы по результатам третьего раздела

1. Проведено исследование применения метода Гаусса-Зейделя для определения параметров узла путём решения системы уравнений баланса. Определены зависимости, позволяющие контролировать точность вычисления характеристик.
2. Предложено и исследовано несколько частных случаев исходной модели. Они отличаются механизмом распределения ресурса передачи информации и числом учитываемых особенностей обслуживания поступающих запросов. Эти модели представляют самостоятельный интерес и могут использоваться при разработке приближенных методов расчета исследуемой модели узла.
3. Показано, что пересылка файлов по правилам трафика сервисов реального времени только на основе ресурса одного канала приводит к оценкам сверху для характеристик исходной модели. Этот результат будет использован для обоснования процедур планирования необходимой производительности узла доступа. Установлено, что в рассматриваемом частном случае и в отсутствии возможности ожидания значения характеристик могут быть найдены с помощью простых рекурсивных выражений по числу занятых каналов. Построен алгоритм реализации рекурсии. Сформулированы выражения для оценки характеристик.
4. Для приближенной оценки характеристик узла предложена процедура декомпозиции модели на две составляющие. В каждой из них рассматривается по отдельности процесс обслуживания трафика сервисов реального времени и пересылка эластичного трафика файлов. Показано, что в каждом из исследуемых частных случаев расчет значений характеристик обслуживания поступающих запросов может быть выполнен с помощью простых рекурсивных формул. Построен алгоритм реализации рекурсии. Сформулированы выражения для оценки характеристик.

5. Область применения предложенной расчетной процедуры — оценка характеристик передачи эластичных файлов при решении задачи планирования необходимого ресурса узла доступа. Поскольку задача планирования решается в области малых потерь, то для оценки среднего времени передачи файла предлагается поступить следующим образом. На первом этапе оценивается количество единиц ресурса передачи информации m_r , занятых трафиком онлайн-услуг. Далее анализируются характеристики передачи данных, допускающие переменную скорость передачи, на оставшихся свободных каналах $[v - m_r]$. Очевидно, что чем ниже уровень потерь заявок на передачу трафика онлайн-услуг, тем точнее будет аппроксимация. Уровень потерь непосредственно зависит от параметра ρ — потенциальной нагрузки на канал.
6. Численно исследована погрешность оценки в зависимости от ρ . Результаты вычислений говорят о хорошей точности расчетной процедуры особенно в области малых потерь, которая, как уже говорилось ранее, соответствует характеристикам обслуживания заявок используемым при решении задач планирования. Эти величины соответствуют значениям максимальных потерь на уровне 0,01—0,05, обеспечивая погрешность 10%–20%.
7. Построена и численно исследована процедура предварительной оценки интервала изменения ρ , где потери поступающих заявок будут достаточно малы, чтобы иметь право воспользоваться при оценке характеристик передачи файлов предложенной процедурой декомпозиции. Для этого применяется модель узла, обеспечивающая верхний характер оценок потерь заявок, построенная в подразделе 3.3.1.

Раздел 4

Оценка пропускной способности мультисервисного узла доступа в условиях учета особенностей формирования нагрузки

4.1 Введение к разделу 4

Построенная в разделе 2 обобщенная модель узла доступа дает возможность определить основные показатели обслуживания мультисервисного трафика. Для решения этой задачи используются точные и приближенные алгоритмы, построенные и исследованные в разделе 3. Важно отметить, что наличие нескольких потоков заявок, участвующих в занятии ресурса передачи информации узла доступа, вносит определенные коррективы в методики планирования необходимой по нагрузке пропускной способности узла. Исследованию особенностей планирования и эффективного распределения в этих условиях ресурса узла доступа и посвящен материал данного раздела.

В подразделе 4.2 численно анализируется зависимость доли потерянных заявок и коэффициента занятия канала от величины предложенной нагрузки на канал применительно к условиям формирования потоков заявок и их обслуживания. Рассмотрены следующие сценарии: наличие гетерогенных свойств поступающих потоков трафика, групповое поступление файлов, наличие ожидания и старения передаваемой информации, а также эластичности пересылаемых файлов.

В следующем подразделе исследуется использование концепции Network Slicing для создания условий по дифференцированному обслуживанию поступающих информационных потоков. Решается задача оценки величины ресурса слайса, обеспечивающего заданную долю потерь для заявок каждого потока поступающего в узел доступа.

В подразделе 4.4 исследуются особенности решения задачи планирования ресурса узла доступа. Рассмотрено влияние на оценку ресурса группового характера поступления файлов и учета свойства эластичности. Анализируются точные и приближенные алгоритмы решения сформулированной задачи. В следующем подразделе предложен способ борьбы с перегрузкой перегрузкой передаточных возможностей узла, основанной на выгрузке трафика в другие узлы. Оценка параметров выгрузки выполняется с помощью построенной модели и алгоритмов расчета ее вероятностных характеристик. В заключительном подразделе сформулированы выводы по результатам материалов, представленных в разделе 4.

4.2 Анализ влияния неоднородности входных потоков на оценку ресурса

4.2.1 Общие положения

Как уже было отмечено в подразделе 2.2.2, наличие свойства гетерогенности у поступающих информационных потоков сильно влияет на значения характеристик их совместного обслуживания на общем ресурсе узла доступа. Это влияние проявляет себя в разной степени и его несомненно надо учитывать при решении задачи планирования величины пропускной способности узла, обеспечивающей заданное качество обслуживания пользователей услуг связи.

Свойство гетерогенности, т.е. неоднородности смеси поступающих в узел информационных потоков, не имеет четкого определения в литературе по теории телетрафика. Исходя из смысла этого понятия, наличие свойства гетерогенности означает присутствие существенного разброса в требованиях к ресурсу передачи информации у поступающих заявок. В данном диссертационном исследовании будем считать, что мультисервисный трафик гетерогенный, если требования к пропускной способности у поступающих заявок различаются более, чем в десять раз.

Назовем трафик с малыми требованиями к пропускной способности легким, а с большими требованиями — тяжелым. Известно [34], что совместное обслуживание легкого и тяжелого трафика приводит к вытеснению последнего из обслуживания. Изучим это свойство более детально. Для этого воспользуемся возможностями описания свойств поступающих потоков заявок, которые встроены в обобщенную модель узла доступа, построенную в разделе 2. Исследуем зависимость доли потерянных заявок и коэффициента занятия канала от величины предложенной нагрузки на канал применительно к условиям формирования потоков заявок и их обслуживания. Влияние гетерогенных свойств трафика будет проанализировано для следующих моделей формирования входных потоков. В первой группе рассмотрены следующие четыре сценария поступления запросов на информационное обслуживание.

1. Основная модель. Здесь отсутствуют перечисленные ниже особенности формирования входных потоков и использования ресурса.
2. Дополнительно к п.1 учет ожидания заблокированных файлов и старения передаваемой информации.
3. Дополнительно к п.2 учет группового характера поступления файлов.
4. Дополнительно к п.3 учет мультисервисного характера поступления заявок на передачу трафика сервисов реального времени.

Во всех перечисленных моделях формирования информационной нагрузки из первой группы предполагается, что поступающие файлы обслуживаются по правилам трафика сервисов реального времени, т.е. путем предоставления фиксированной скорости передачи информации в размере одного виртуального канала на все время соединения. Во

второй группе рассматриваются те же самые четыре сценария поступления запросов на информационное обслуживание, но обслуживание файлов уже происходит по правилам дисциплины Processor Sharing (см. подраздел 2.2.3). Более детально входные параметры поступающих информационных потоков будут рассмотрены ниже.

4.2.2 Обслуживание файлов по правилам сервисов реального времени

Первая группа моделей исследуется при следующих входных параметрах:

- Пропускная способность $C = 100$ Мбит/с.
- Количество потоков $n = 1$.
- Скорость первого потока $c_1 = 20$ Мбит/с.

Скорость виртуального канала c установлена на уровне 1 Мбит/с, а средний размер передаваемого файла F выбран таким, что $F = 1$ Мбит. Важно отметить, что в данной группе моделей файлы обслуживаются в соответствии с правилами трафика сервисов реального времени, то есть для передачи используется пропускная способность одного канала на протяжении всего сеанса связи.

Из этих допущений следует, что:

- Общее число виртуальных каналов $v = 100$.
- Число каналов, занятых первым потоком, $b_1 = 20$.

Математическое ожидание времени пересылки одного файла одним каналом равно $\frac{1}{\alpha_d} = 1$ секунд. Этот промежуток времени в дальнейших вычислениях возьмем за единицу. Следовательно, $\alpha_d = 1$.

Параметр экспоненциального распределения для времени обслуживания трафика сервисов реального времени выбран как $\alpha_{r,1} = 0,1$.

Нагрузка ресурса оценивается через коэффициент ρ , который представляет собой интенсивность предложенного трафика на один виртуальный канал. Этот параметр вычисляется по формуле:

$$\rho = \frac{1}{v} \left(\frac{\lambda_{r,1}}{\alpha_{r,1}} b_1 + \frac{\lambda_d \bar{b}}{\alpha_d} \right). \quad (4.1)$$

В расчетах предполагается, что оба потока создают одинаковую потенциальную нагрузку на пропускную способность узла доступа. Из этого условия получаются выражения для интенсивностей поступления запросов каждого потока:

$$\lambda_{r,1} = \frac{v\rho\alpha_{r,1}}{2b_1}; \quad \lambda_d = \frac{v\rho\alpha_d}{2\bar{b}}. \quad (4.2)$$

Сравнение характеристик интенсивностей входных потоков позволяет заключить, что поступающие потоки трафика online-услуг обладают свойством гетерогенности, сформулированным в начале предыдущего подраздела, поскольку b_1 более, чем в 10 (на самом

деле в 20) раз превосходит объем ресурса, используемого для передачи файлов. Рассмотрим, как наличие этого свойства влияет на основные характеристики совместного обслуживания тяжелого (первый поток) и легкого (поток файлов) трафика. В качестве характеристик выберем $\pi_{r,1}$ и π_d , соответственно, значения потерь заявок трафика реального времени и заявок на передачу файлов, а также $\delta_{r,1}$, δ_d , соответственно, величины значения среднего использования единицы ресурса каждым из анализируемых информационных потоков. Перечисленные характеристики определяются из выражений

$$\delta_{r,1} = \frac{m_{r,1}}{v}; \quad \delta_d = \frac{m_d}{v}. \quad (4.3)$$

Формальные определения характеристик сформулированы через величины предельных вероятностей модели в подразделе 3.3.1. Значения вероятностей рассчитываются посредством решения системы уравнений равновесия (3.7) методом последовательных замещений Гаусса-Зейделя.

На приведенных далее рисунках 4.1 — 4.3 в левой части показаны зависимости $\pi_{r,1}$ (кривая RT) и π_d (кривая DRT) от изменения загрузки единицы ресурса ρ , а в правой части исследуются зависимости от ρ значений $\delta_{r,1}$ и δ_d . Кривые обозначены теми же символами, что были использованы в левой части рисунка 4.1.

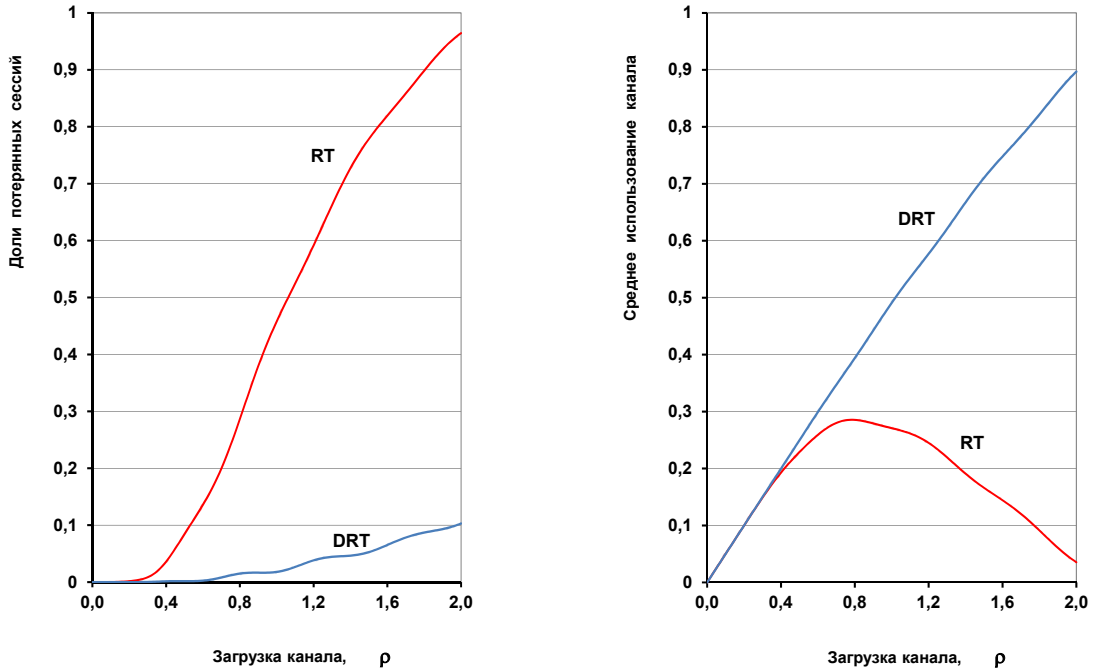


Рисунок 4.1 — Основная модель

На рисунке 4.1 исследуются перечисленные выше зависимости для основной модели узла, формирующей первую группу сценариев поступления запросов на информационное обслуживание, где не учитываются возможности ожидания, группового поступления

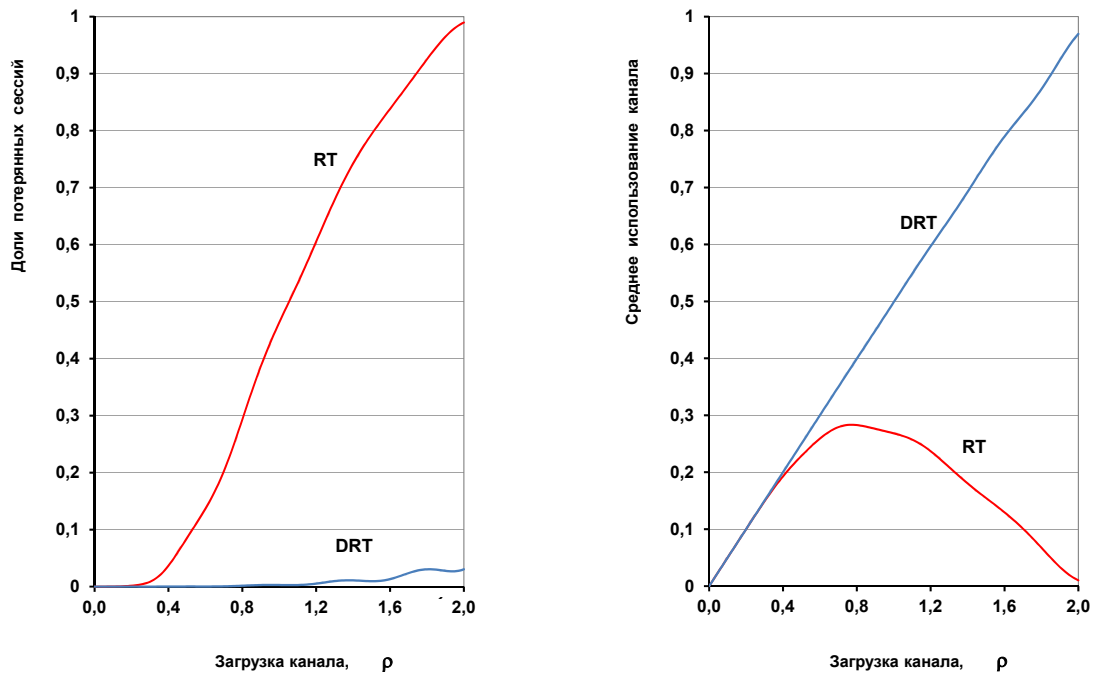


Рисунок 4.2 — Основная модель и ожидание для файлов

файлов и рассматривается один поток трафика сервисов реального времени. На рисунке 4.2 представлены аналогичные зависимости, но для усложненной модели узла, в которой добавлены механизмы ожидания отложенных файлов и учета старения передаваемых данных. Параметры ожидания и старения заданы значениями: $w = 40$, $\sigma = 0,1$. На рисунке 4.3 те же зависимости анализируются для дополнительно усложненной модели узла с учетом ожидания для задержанных файлов, старения передаваемой информации и возможности группового поступления файлов. Параметры формирования групп файлов находятся из выражений: $m_g = 30$, $f_s = \frac{1}{30}$, $s = 1, 2, \dots, m_g$.

На рисунке 4.4, помимо ранее учтенных особенностей формирования входных потоков заявок (использованных при построении данных на рисунке 4.3), предполагается, что трафик реального времени включает два потока (т.е. $n = 2$). Остальные параметры имеют следующие значения:

- $C = 100$ Мбит/с (общая пропускная способность).
- $c_1 = 10$ Мбит/с, $c_2 = 20$ Мбит/с (пропускные способности для двух потоков).
- $c = 1$ Мбит/с (скорость одного виртуального канала).
- $F = 1$ Мбит (средний объем передаваемого файла).

Таким образом:

- $v = 100$ в.к. (общее число виртуальных каналов).

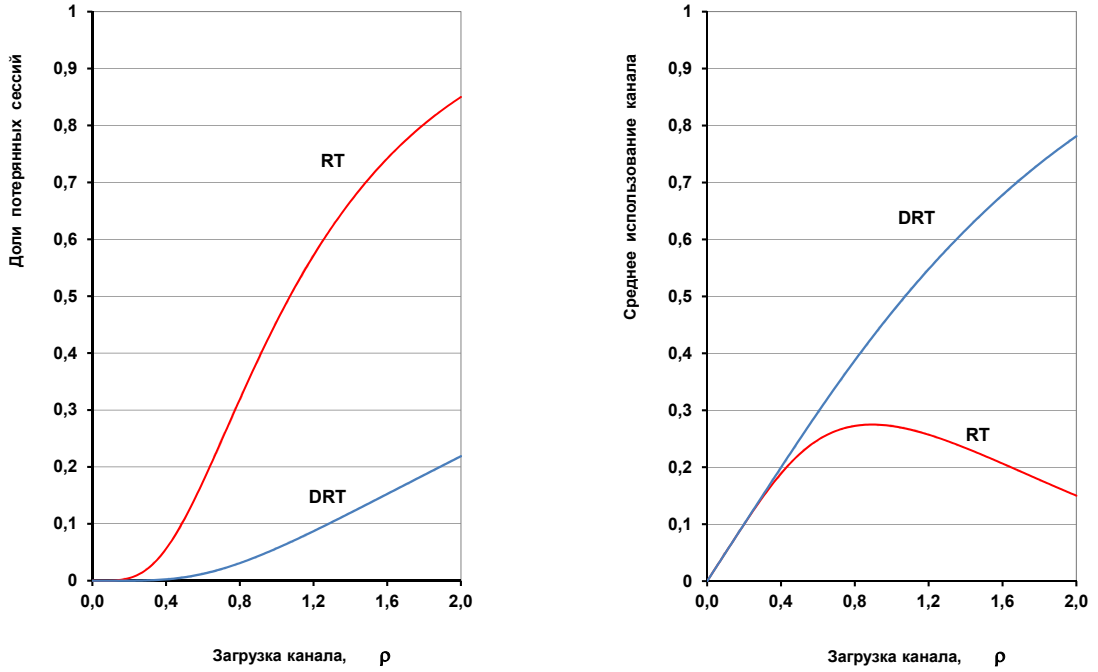


Рисунок 4.3 — Основная модель, ожидание и групповое поступление файлов

- $b_1 = 10$ в.к., $b_2 = 20$ в.к. (выделенные каналы для потоков).
- Среднее время передачи файла одним каналом: $1/\alpha_d = 1$ секунда (принято за единицу времени в расчетах, поэтому $\alpha_d = 1$).
- Параметры экспоненциального распределения времени обслуживания трафика сервисов реального времени заданы как $\alpha_{r,1} = \alpha_{r,2} = 0,1$.

Степень загрузки ресурса оценивается значением ρ интенсивности предложенного трафика на один виртуальный канал. Величина ρ определяется из соотношения

$$\rho = \frac{1}{v} \left(\frac{\lambda_{r,1}}{\alpha_{r,1}} b_1 + \frac{\lambda_{r,2}}{\alpha_{r,2}} b_2 + \frac{\lambda_d}{\alpha_d} \bar{b} \right). \quad (4.4)$$

В ходе расчетов будем исходить из того, что все три потока создают равную нагрузку на пропускную способность точки доступа. На основе этого допущения выведем формулы для расчета интенсивности поступления запросов по каждому из анализируемых потоков.

$$\lambda_{r,1} = \frac{v\rho\alpha_{r,1}}{3b_1}; \quad \lambda_{r,2} = \frac{v\rho\alpha_{r,2}}{3b_2}; \quad \lambda_d = \frac{v\rho\alpha_d}{3\bar{b}}. \quad (4.5)$$

На рисунке 4.4 в левой части показаны зависимости $\pi_{r,1}$ и $\pi_{r,2}$ (кривые RT1 и RT2) и π_d (кривая DRT) от изменения загрузки единицы ресурса ρ , а в правой части исследуются

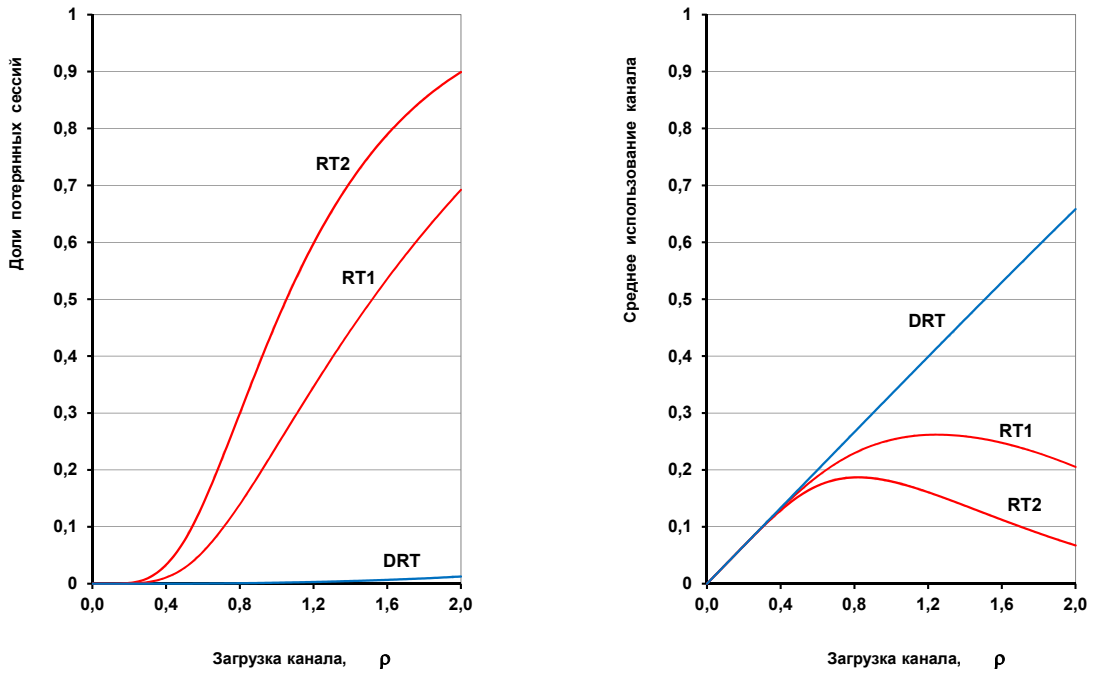


Рисунок 4.4 — Основная модель, ожидание, групповое поступление файлов, два потока RT

зависимости от ρ значений $\delta_{r,1}$, $\delta_{r,2} = \frac{m_{r,2}}{v}$, δ_d (см. (4.3)). Кривые обозначены теми же символами, что были использованы в левой части рисунка 4.4. Результаты анализа численных данных, представленных на рисунках 4.1 — 4.4, будут сформулированы в следующем подразделе, где влияние свойства гетерогенности на оценку ресурса будет исследовано для второй группы сценариев с обслуживанием файлов по правилам дисциплины Processor Sharing (см. подраздел 2.2.3).

4.2.3 Обслуживание файлов по правилам эластичного трафика

Чтобы иметь возможность сравнить действие одних и тех же факторов на изменение сценария распределения ресурса при обслуживании файлов, сохраним значения всех сценариев и параметров формирования входных потоков. Изменение только затронет способ использования ресурса передачи информации узла доступа. На рисунках 4.1 — 4.4, поступающие файлы передавались с использованием передаточных возможностей одного виртуального канала, а на рисунках этого подраздела с использованием всех виртуальных каналов, свободных от обслуживания трафика сервисов реального времени.

Характеристики обслуживания заявок определены в подразделе 2.5.2 с использованием значений стационарных вероятностей, а сами значения вероятностей рассчитаны с помощью решения системы уравнений равновесия (2.8) алгоритмом Гаусса-Зейделя. Детали реализации алгоритма исследованы в подразделе 3.2. Значения входных параметров

поступающих потоков и имеющийся ресурс передачи узла доступа остались теми же, что были использованы при подготовке численного материала, использованного построении соответствующих кривых на рисунках 4.1 — 4.4. В обозначениях зависимостей характеристик передачи файлов, полученных с использованием дисциплины Processor Sharing используется символ DPS. Результаты вычислений представлены на рисунках 4.5 — 4.8.

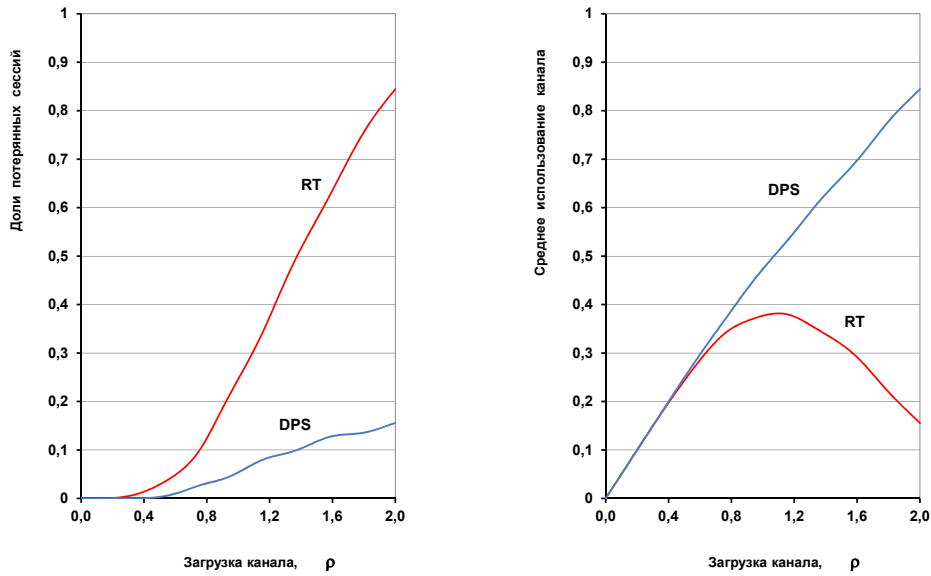


Рисунок 4.5 — Основная модель

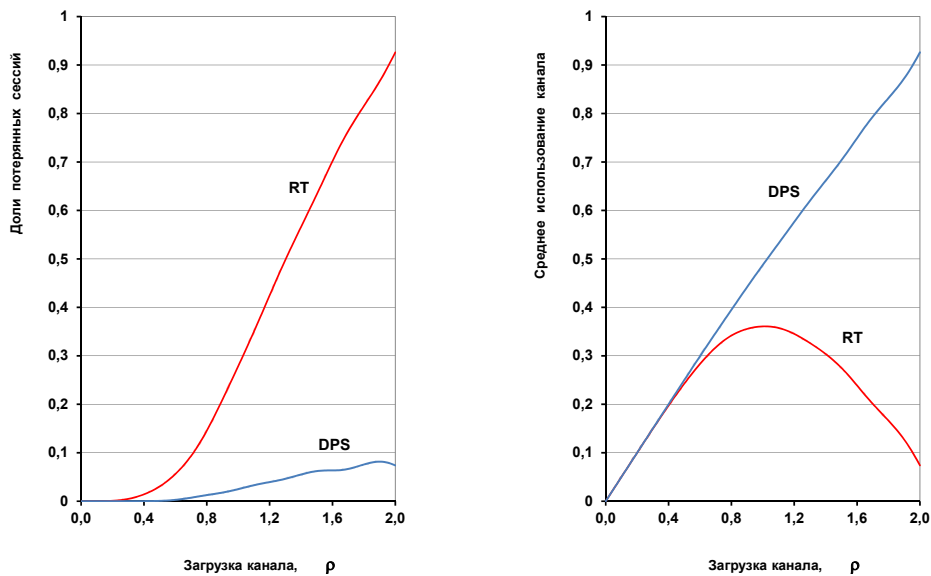


Рисунок 4.6 — Основная модель и ожидание для файлов

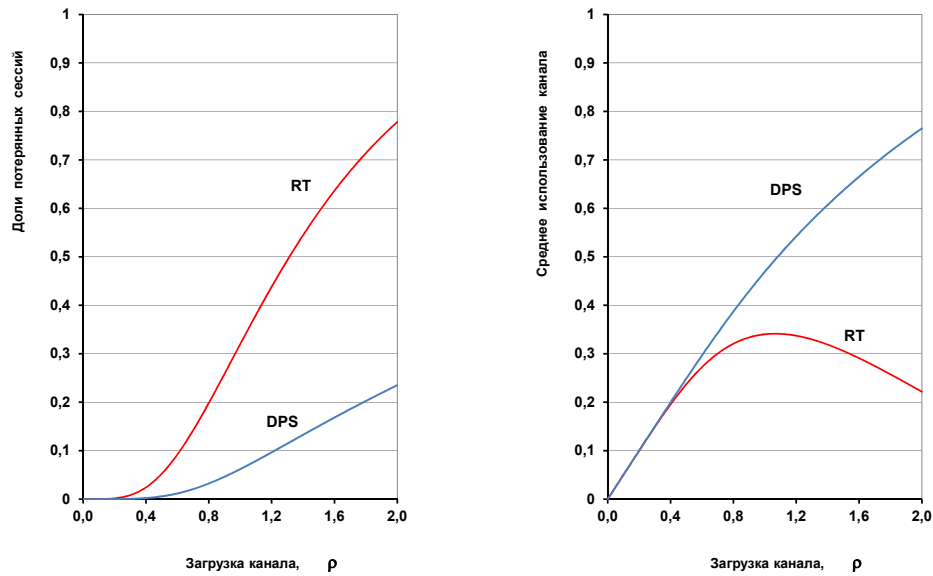


Рисунок 4.7 — Основная модель, ожидание и групповое поступление файлов

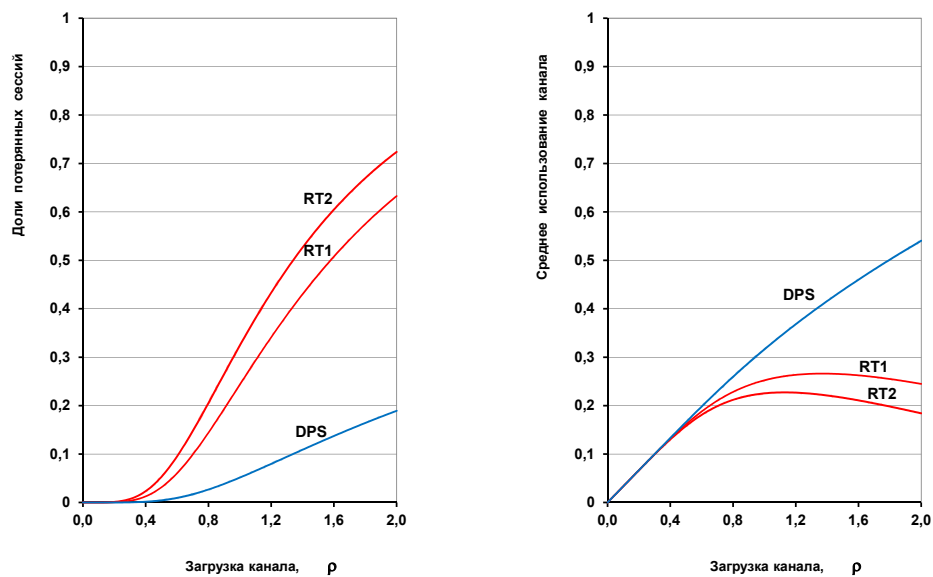


Рисунок 4.8 — Основная модель, ожидание, групповое поступление файлов, два потока RT

Сформулируем выводы по результатам выполненного численного исследования.

1. Во всех рассмотренных случаях наличие свойства гетерогенности влияет на значения основных характеристик совместного обслуживания поступающих потоков заявок, а, следовательно, и на решение задачи планирования ресурса узла доступа, которая осуществляется с использованием этих характеристик.
2. Это влияние зависит от нагрузки и от используемых сценариев формирования потоков заявок и использованных сценариев назначения свободных единиц ресурса пе-

передачи информации, но для всех рассмотренных случаях легкий трафик вытесняет из обслуживания тяжелый трафик.

3. Для небольших величин ρ в интервале от 0 до 0,5, когда поступающие заявки не испытывают конкуренции в занятии ресурса, это влияние почти не заметно. Оно начинает проявлять себя со средних значений ρ , начиная примерно с 0,8 и выше и наиболее сильно сказывается в области перегрузки, когда ρ становится выше 1,2.
4. Из всех изученных в расчетах факторов наиболее сильно влияют на значения характеристик наличие ожидания (см. рисунки 4.2, 4.6) и групповой характер поступления файлов (см. рисунки 4.3, 4.7). В первом случае по сравнению с основной моделью увеличивается доля файлов, попавших на обслуживание, что ведет к более сильному вытеснению тяжелого трафика легким трафиком. Во втором сценарии, по сравнению с исходной моделью, наблюдается увеличение потерь файлов, но при этом частично сокращаются потери запросов на передачу трафика реального времени, а также возрастает уровень использования канала этим видом трафика.
5. Использование дисциплины Processor Sharing уменьшает долю файлов, находящихся на обслуживании. Это по сравнению с основной моделью увеличивает шансы трафика сервисов реального времени попасть на обслуживание. В результате у подобных заявок уменьшаются потери и увеличивается коэффициент использования канала. Это ведет к увеличению потерь заявок на передачу файлов и уменьшению коэффициента использования канала.
6. Наличие более, чем одного потока заявок на передачу трафика сервисов реального времени не приводит к существенному изменению перечисленных выше свойств совместного обслуживания информационных потоков в узле доступа.
7. В целом давая оценку анализируемым факторам отметим, что наличие неоднородности в совместно обслуживаемых потоках трафика и групповой характер поступления файлов увеличивают требование к числу единиц ресурса узла доступа достаточному для пересылки поступающих информационных потоков с заданным качеством. В тоже время наличие возможности ожидания и дисциплины равномерного распределения имеющегося ресурса уменьшают величину требования.

Выполненный анализ показал, что оценка необходимой по нагрузке и качеству обслуживания величины ресурса передачи информации узла доступа должно происходить с учетом всех перечисленных факторов, следовательно, с использованием обобщенной модели узла построенной и исследованной в предыдущих разделах работы. Перейдем к изложению соответствующих результатов.

4.3 Анализ сценариев дифференцированного обслуживания гетерогенного трафика

4.3.1 Использование архитектуры Network Slicing

Результаты численных расчетов, представленные в подразделе 4.2, показывают, что наличие гетерогенных свойств у поступающих информационных потоков сильно сказываются на величинах характеристик их совместного обслуживания. Если перед оператором стоит задача обеспечить поступающие потоки трафика заданными показателями качества обслуживания, то нужно каким-то образом управлять доступом заявок к ресурсу. Реализовать дифференцированное обслуживание можно разными способами. Рассмотрим два сценария: использование возможностей, заложенных в архитектуру Network Slicing и реализация функции контроля за доступом. Последнюю можно связать со степенью загрузки пропускной способности узла. В данном подразделе рассмотрим первый из сформулированных сценариев, а в следующем — второй.

Детали реализации архитектуры Network Slicing рассмотрены в подразделе 1.4.1. Ее применение дает возможность виртуальным образом разделить имеющийся ресурс узла доступа на отдельные фрагменты (слайсы), каждый из которых предназначен для обслуживания группы поступающих информационных потоков с примерно одинаковыми свойствами. К этим свойствам в первую очередь относится требование к скорости передачи информационного потока, связанного с обслуживанием заказанного сервиса, а также другие особенности формирования заявок и их обслуживания. Эти свойства подробно исследованы в подразделе 4.2. Рассмотрим численный пример, который покажет последовательность применения архитектуры Network Slicing в целях создания условий по дифференцированному обслуживанию мультисервисного трафика в узле доступа.

В ходе расчетов были применены следующие параметры мультисервисного узла доступа:

- Общая пропускная способность: $C = 100$ Мбит/с.
- Количество типов трафика: $n = 2$.
- Скорости передачи для каждого типа: $c_1 = 5$ Мбит/с, $c_2 = 10$ Мбит/с.
- Скорость виртуального канала: $c = 1$ Мбит/с.
- Средний размер файла: $F = 1$ Мбит.

Согласно определению, принятому в диссертации, входящие потоки трафика являются гетерогенными, поскольку первый поток требует в 10 раз больше ресурсов, чем минимальная скорость передачи файла. Из указанных допущений следует:

- Общее число виртуальных каналов: $v = 100$.
- Требуемые каналы для потоков: $b_1 = 5$, $b_2 = 10$.
- Среднее время передачи файла одним каналом: $\frac{1}{\alpha_d} = 1$ секунда (принято за единицу времени), откуда $\alpha_d = 1$.

Для заявок на передачу трафика сервисов реального времени параметры обслуживания выбраны как: $\alpha_{r,1} = 0,1$, $\alpha_{r,2} = 0,1$. Характеристики группового поступления файлов:

- Вероятность группового поступления: $f_s = 1/10$, где $s = 1, 2, \dots, 10$ (среднее $\bar{b} = 5,5$)
- Число мест ожидания: $w = 0$.

Обслуживание файлов осуществляется в соответствии с дисциплиной Processor Sharing.

Качество обслуживания мультисервисного трафика оценим значениями $\pi_{r,1}$ и $\pi_{r,2}$ доли потерянных заявок соответственно 1-го и 2-го потоков и величиной потерь файлов π_d . Как и в других рассмотренных примерах расчета характеристик, уровень загрузки пропускной способности узла доступа оценивается значением ρ интенсивности предложенного трафика на один виртуальный канал. Величина ρ определяется из соотношения

$$\rho = \frac{1}{v} \left(\frac{\lambda_{r,1}}{\alpha_{r,1}} b_1 + \frac{\lambda_{r,2}}{\alpha_{r,2}} b_2 + \frac{\lambda_d}{\alpha_d} \bar{b} \right). \quad (4.6)$$

При выполнении расчётов будем исходить из допущения, что все три потока создают равную потенциальную нагрузку на пропускную способность узла доступа. На основании этого предположения можно вывести формулы для расчёта интенсивности поступления запросов для каждого из рассматриваемых потоков

$$\lambda_{r,1} = \frac{v\rho\alpha_{r,1}}{3b_1}; \quad \lambda_{r,2} = \frac{v\rho\alpha_{r,2}}{3b_2}; \quad \lambda_d = \frac{v\rho\alpha_d}{3\bar{b}}. \quad (4.7)$$

Определения рассчитываемых характеристик приведены в подразделе 2.2.5. Их численные значения находились из решения СУР (2.8) алгоритмом Гаусса-Зейделя (см. подраздел 3.2). Значение $\rho = 1,5$. При данном выборе ρ величины интенсивностей находятся из (4.7) и определяются из выражений $\lambda_{r,1} = 1,0$; $\lambda_{r,2} = 0,5$, $\lambda_d = 9,09$. Напомним, что значения интенсивностей приведены в единицу времени, в качестве каковой выбрано среднее время передачи файла одним каналом.

Величины характеристик принимают значения: $\pi_{r,1} = 0,308730$; $\pi_{r,2} = 0,485113$; $\pi_d = 0,245763$. Несмотря на одинаковые значения потенциальной нагрузки для всех трех совместно обслуживаемых потоков величины потерь отличаются достаточно сильно. Также непропорционально нагрузке занят ресурс виртуального канала. Значение средней занятости канала пересылкой трафика реального времени вычисляется из равенства $\delta_i = 0,603$, а для трафика данных эта характеристика определяется из соотношения $\delta_d = 0,377$ и это несмотря на то, что при использовании дисциплины Processor Sharing на передачу данных занят весь свободный ресурс. Приведенные данные еще раз демонстрируют, что неконтролируемый доступ легкого трафика нарушает принципы справедливого обслуживания мультисервисного трафика.

Дальнейшие действия по устранению отмеченных недостатков зависят от тех возможностей, которыми располагает администрация сети. Если изменить имеющееся количество ресурса передачи информации нельзя, то можно средствами Network Slicing разделить имеющийся ресурс так, чтобы каждый из поступающих потоков получал примерно одинаковый уровень обслуживания. Если получаемый уровень обслуживания недостаточен,

то найти дополнительный объем ресурса и далее воспользоваться процедурой Network Slicing для выравнивания показателей качества обслуживания. Рассмотрим решение каждой из отмеченных задач.

Воспользуемся возможностями архитектуры Network Slicing и создадим условия для обслуживания поступающих потоков трафика на имеющемся ресурсе $v = 100$ в.к. с примерно одинаковыми значениями потерь. С этой целью объединим оба потока трафика реального времени в одну группу и найдем величину v_1 пропускной способности 1-го слайса, который будет обслуживать заявки этих потоков с уровнем потерь $\pi_1 \approx \pi_2$, где π_2 величина потерь файлов на остаточной части ресурса $v_2 = 100 - v_1$, который назовем 2-ым слайсом. Результаты вычислений показаны на рисунке 4.9.

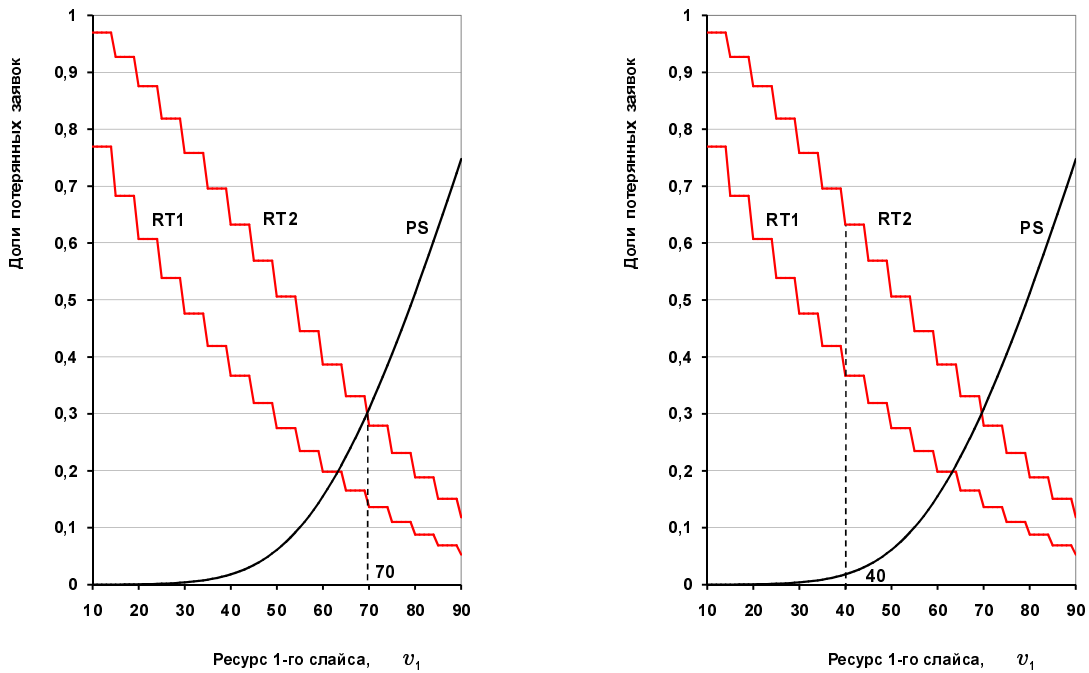


Рисунок 4.9 — Оценка объема слайсов, обеспечивающих примерно равный уровень потерь поступающих потоков (левая часть) и создающих преимущество для обслуживания файлов (правая часть)

Характеристики обслуживания трафика сервисов реального времени и файлов можно рассчитывать независимо. В первом случае отсутствует поток файлов, поэтому для расчета характеристик можно воспользоваться частным случаем построенной модели, который исследован в подразделе 3.3.4. В анализируемой ситуации численный анализ модели сводится к использованию рекурсивной зависимости (3.13). Во втором случае отсутствуют потоки трафика сервисов реального времени, поэтому для расчета характеристик можно воспользоваться частным случаем построенной модели, который исследован в подразделе 3.3.5. Определения характеристик сформулированы в виде соотношений (3.17). Для расчетов использовалась рекурсия (3.20). В левой части рисунка 4.9 показаны результаты оценки объема 1-го и 2-го слайсов, обеспечивающие примерно равные значения потерь

заявок на передачу трафика сервисов реального времени и данных. Величины слайсов определяются из соотношений $v_1 = 70$ в.к., соответственно $v_2 = 30$ в.к. Потери заявок находятся примерно на уровне 0,3.

Можно изменить поставку задачи и найти такое разбиение имеющегося ресурса узла доступа на слайсы, которое бы обеспечило преимущество в обслуживании файлов, например, обеспечило бы потери файлов на уровне $\approx 0,02$. Решение задачи рассмотрено в правой части рисунка 4.9. Величины слайсов определяются из соотношений $v_1 = 40$ в.к., соответственно $v_2 = 60$ в.к. В рассматриваемой ситуации $\pi_d \approx 0,02$, однако потери заявок на обслуживание сервисов реального времени превышают 60%. Понятно, что такая положение дел приводит к изменению постановки задачи. Теперь она будет звучать так. Требуется найти общий объем ресурса и его разбиение на слайсы, которое бы обеспечило заданный уровень потерь поступающих информационных потоков. Пусть желаемый уровень потерь составляет 5%. На рисунке 4.10 показаны результаты подбора минимальной пропускной способности слайса, достаточной для обслуживания трафика реального времени с потерями заявок, не превосходящими 0,05. Расчет характеристик передачи файлов осуществляется в точности так, как это было сделано для характеристик передачи трафика сервисов, представленных на рисунке 4.9. Результаты выбора минимальной величины 1-го слайса v_1 показаны на рисунке 4.10, где приведены зависимости $\pi_{r,1}$ и $\pi_{r,2}$ от v_1 . Решение задачи достигается при $v_1 = 145$ в.к.

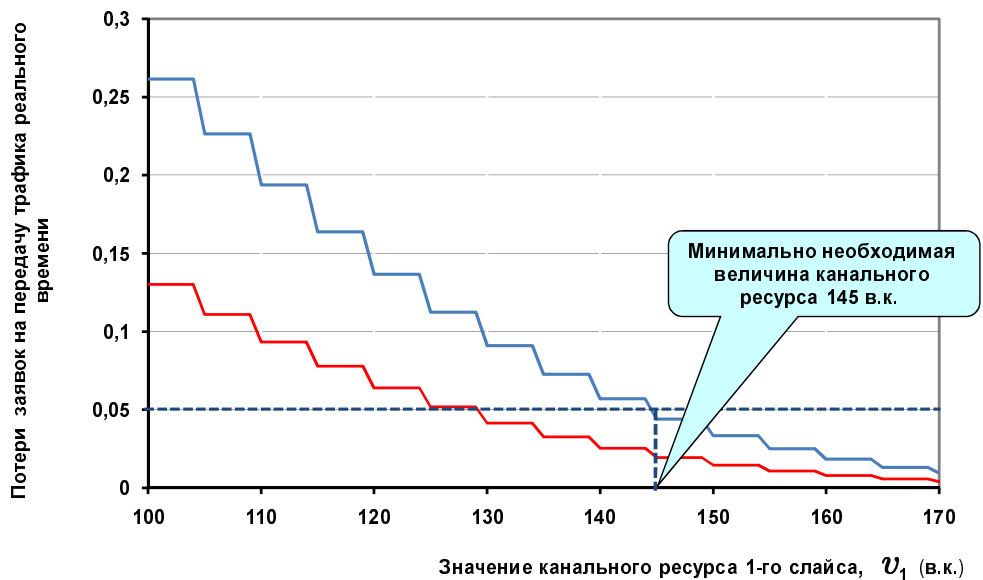


Рисунок 4.10 — Решение задачи оценки минимальной пропускной способности слайса, достаточной для обслуживания трафика online-услуг с потерями заявок, не превосходящими 0,05

На рисунке 4.11 представлены результаты определения минимальной пропускной способности слайса, необходимой для передачи эластичного трафика данных с допустимым уровнем потерь файлов не более 0,05. Методика расчета параметров передачи файлов идентична подходу, использованному для анализа характеристик трафика данных на рисунке 4.9. Оптимальное значение пропускной способности составляет $v_2 = 53$ в.к. Общий объем ресурса v , обеспечивающий дифференцированное обслуживание рассматриваемых потоков данных, вычисляется по формуле $v = 198$ в.к.

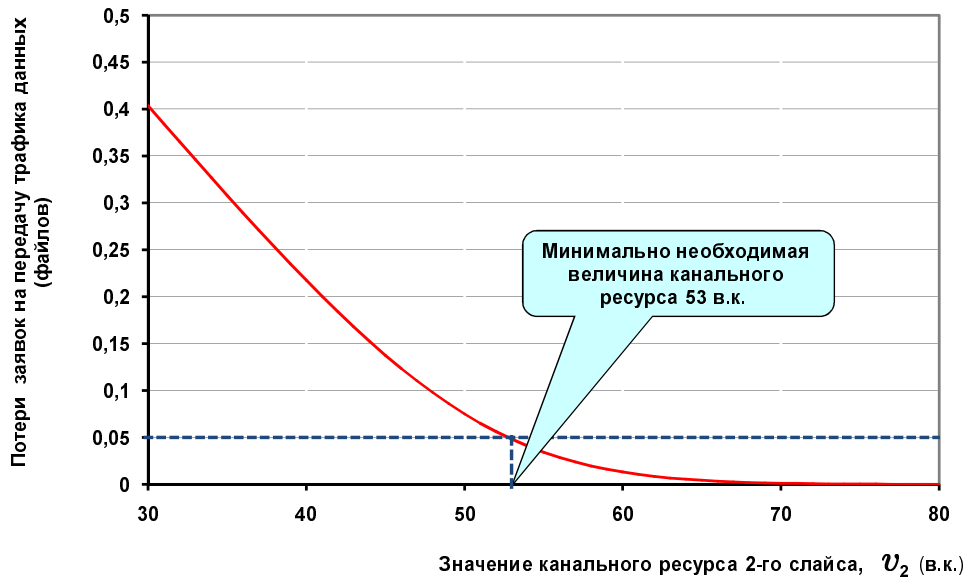


Рисунок 4.11 — Результаты подбора минимальной пропускной способности слайса, достаточной для обслуживания эластичного трафика данных с потерями файлов, не превосходящими 0,05

Заметим, что в исследуемом случае дифференцированное обслуживание требует больше каналов, чем совместная передача этих потоков на общем ресурсе. Покажем этот результат на численном примере. Рассмотрим задачу выбора пропускной способности узла v , обеспечивающую потери заявок потоков, характеристики которых были представлены на рисунках 4.9–4.11, на уровне 0,05. Результаты вычисления $\pi_{r,1}$, $\pi_{r,2}$, π_d в зависимости от числа виртуальных каналов v , определяющих пропускную способность узла доступа, показаны на рисунке 4.12.

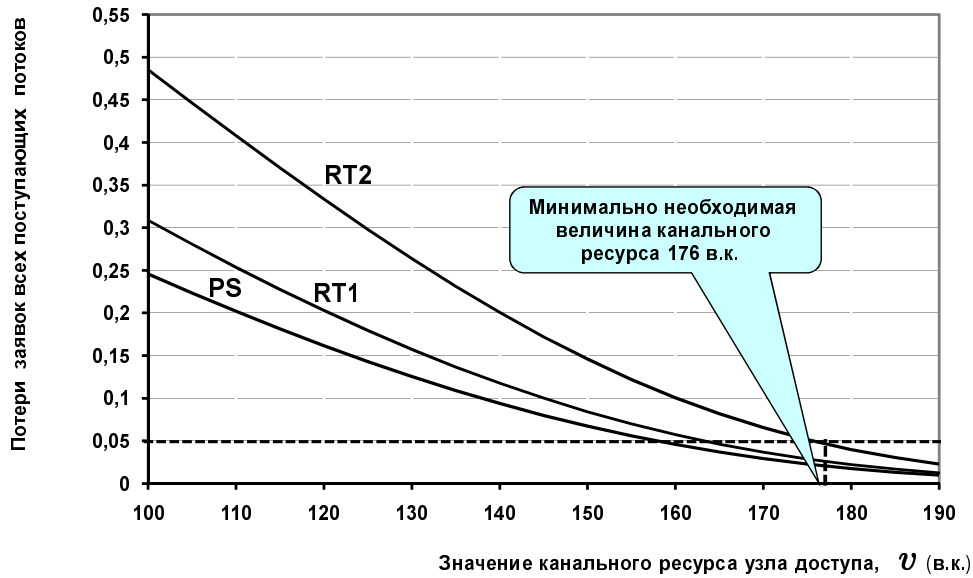


Рисунок 4.12 — Результаты подбора минимальной пропускной способности слайса, достаточной для обслуживания трафика сервисов реального времени и эластичного трафика данных с потерями, не превосходящими 0,05

Для решения сформулированной задачи достаточно наличия 176 виртуальных каналов. Таким образом совместное обслуживание требует на 20 каналов меньшую пропускную способность для обеспечения потерь для всех потоков менее, чем 0,05. Столь существенное уменьшение пропускной способности связано с использованием дисциплины Processor Sharing при передаче эластичных данных. Файлы передаются на всех свободных каналах, что приводит к уменьшению требуемой пропускной способности узла. Этот факт можно показать примере модели узла, рассмотренной в подразделе 3.3.1, где файлы передаются с использованием возможностей только одного канала. На рисунке 4.12 показана зависимость $\pi_{r,1}$, $\pi_{r,2}$, π_d от u для величин параметров информационных потоков, что применялись при расчете диаграмм, представленных на рисунке 4.12. Значения характеристик определены в подразделе 3.3.1. Они рассчитывались с помощью решения (3.7) методом последовательных замещений Гаусса-Зейделя. Для обеспечения потерь заявок на уровне, не превосходящем 0,05, требуется пропускная способность на 22 канала больше, чем в ситуации, когда файлы обслуживаются с использованием дисциплины Processor Sharing.

4.3.2 Использование ожидания при передаче файлов

Условия по дифференцированному обслуживанию поступающих заявок можно создать, если воспользоваться их свойствами по отношению к процессу обслуживания. На рисунке 4.12 показано, что свойство эластичности позволяет значительно ускорить передачу файлов и уменьшить требование к пропускной способности узла, обеспечивающей задан-

ное качество обслуживания. Имеется еще одна возможность, которая позволяет варьировать показатели обслуживания файлов. Речь идет об использовании ожидания. Интуитивно понятно, что показатели качества обслуживания файлов можно улучшить без увеличения пропускной способности узла доступа, если предоставить возможность ожидания начала обслуживания для заблокированных файлов. Техническая организация возможности ожидания экономически более выгодна, чем увеличение пропускной способности узла доступа. Единственное, что требуется контролировать это влияние ожидания на качество предоставляемой услуги.

Рассмотрим модель узла с параметрами: $v = 60$ в.к.; $b_1 = 1$ в.к.; $b_2 = 2$ в.к. Примем 1 в.к.=1 Мбит/с. Характеристики обслуживания заявок $\alpha_{r,1} = 0,1$, $\alpha_{r,2} = 0,1$; $F = 1$ Мбит; $\alpha_d = 1$. Параметры группового поступления файлов определяются из выражений: $f_s = 1/20$, $s = 1, 2, \dots, 20$ ($\bar{b} = 10,5$); $w = 0$; $\sigma = 0,1$. Файлы обслуживаются в соответствии с положениями дисциплины Processor Sharing. Интенсивности потоков определяются из (4.7) для $\rho = 1,5$ и принимают значения $\lambda_{r,1} = 3$; $\lambda_{r,2} = 1,5$; $\lambda_d = 2,8571$. Для выбранных значений параметров доли потерянных заявок $\pi_{r,1} = 0,159831$; $\pi_{r,2} = 0,286629$; $\pi_d = 0,606339$. Сократим потери до уровня примерно 5%. Решение с использованием двух отдельных слайсов для трафика реального времени и файлов требует соответственно 71 в.к.+44 в.к.=115 в.к. Результаты решения этой же задачи с использованием ожидания на общем ресурсе $v = 100$ в.к. показаны на рисунке 4.13. Таким образом, использование $w = 6$ мест ожидания позволяет снизить требование к пропускной способности примерно на 15% по сравнению с использованием процедуры Network Slicing.

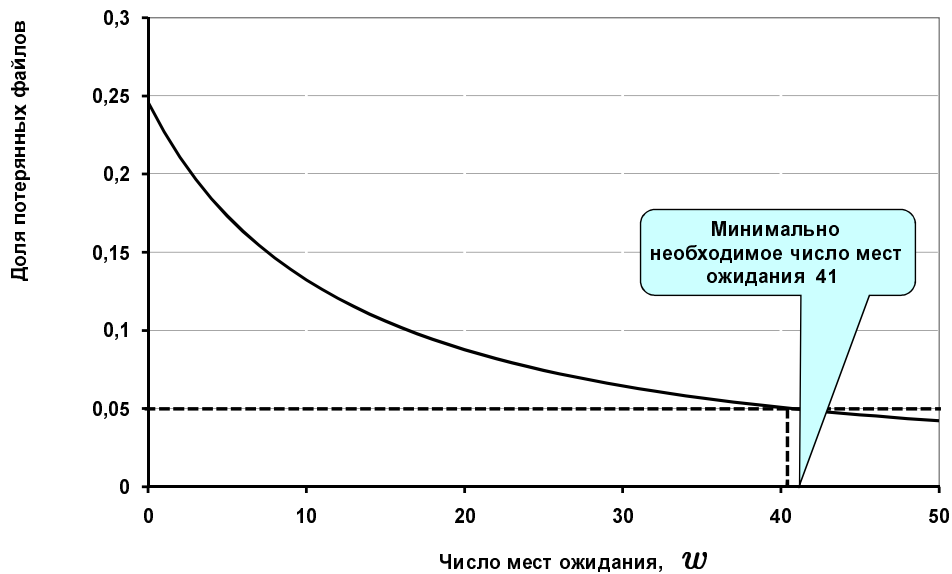


Рисунок 4.13 — Результаты подбора минимального числа ожидающих файлов, достаточной для обслуживания эластичного трафика данных с потерями, не превосходящими 0,05

4.3.3 Использование выгрузки файлов

Теперь рассмотрим способ создания условий по дифференцированному обслуживанию поступающих файлов, который уменьшает вероятности потерь для всех поступающих потоков. Он заключается в просеивании входного потока файлов и перемаршрутизации части выгруженных файлов на другие узлы доступа, обеспеченные аналогичным функционалом обслуживания. Обозначим через p_o вероятность выгрузки поступившей группы файлов. С дополнительной вероятностью поступившая группа обслуживается в анализируемом узле доступа. Процесс обслуживания этой группы и всех других поступающих заявок проводится в точности так как это было сделано в обобщенной модели, рассмотренной в подразделе 2.2. Параметры модифицированной математической модели узла доступа с использованием процедуры просеивания для потока файлов показаны на рисунке 4.14.

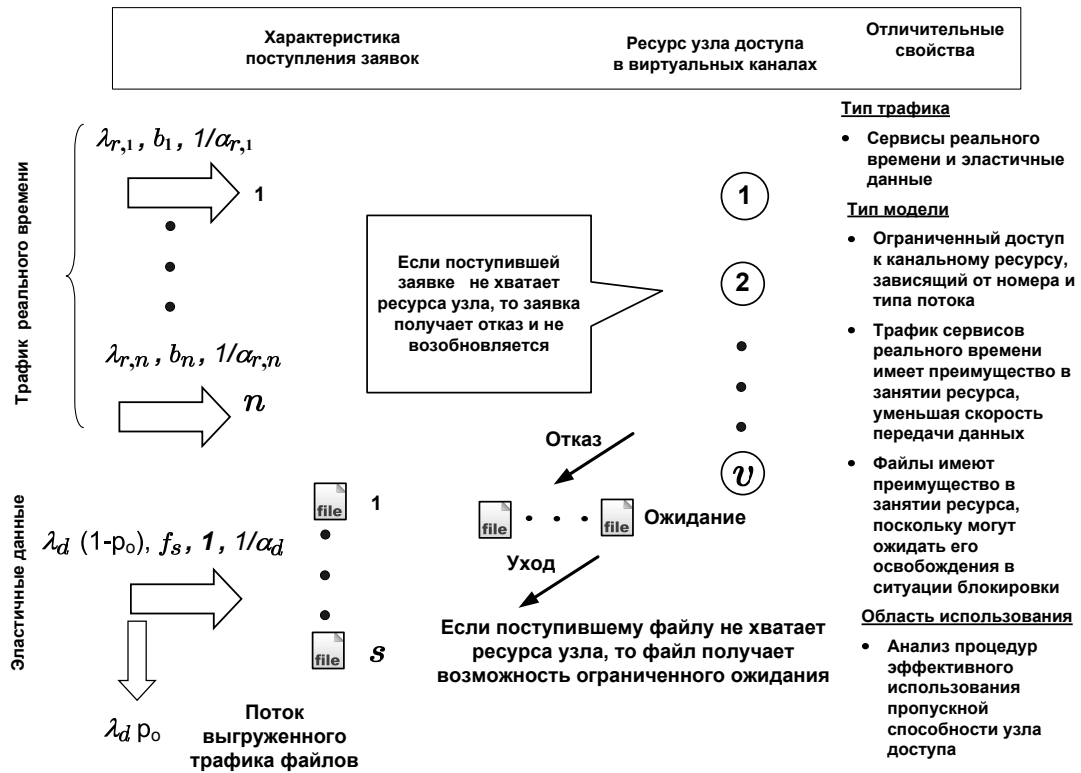


Рисунок 4.14 — Математическая модель узла доступа с функцией просеивания для поступающих потоков файлов

На рисунке 4.15 показана зависимость $\pi_{r,1}$, $\pi_{r,2}$, π_d от v для величин входных параметров, которые применялись при расчете диаграмм, представленных на рисунке 4.12 за исключением величины ρ , которая теперь определяется из соотношения $\rho = 0,9$. Значения характеристик определены в подразделе 3.3.1. Их численные значения получены в результате решения системы уравнений (3.7) итерационным методом Гаусса-Зейделя. Для обеспечения потерь заявок на уровне, не превосходящем 0,05, требуется выгрузить 81 процент поступивших групп файлов. Столь существенный процент выгрузки связан с

тем, что файлы поступают группами, что увеличивает потребность в пропускной способности узлов. Этот результат будет более подробно исследован в следующем подразделе диссертационного исследования.

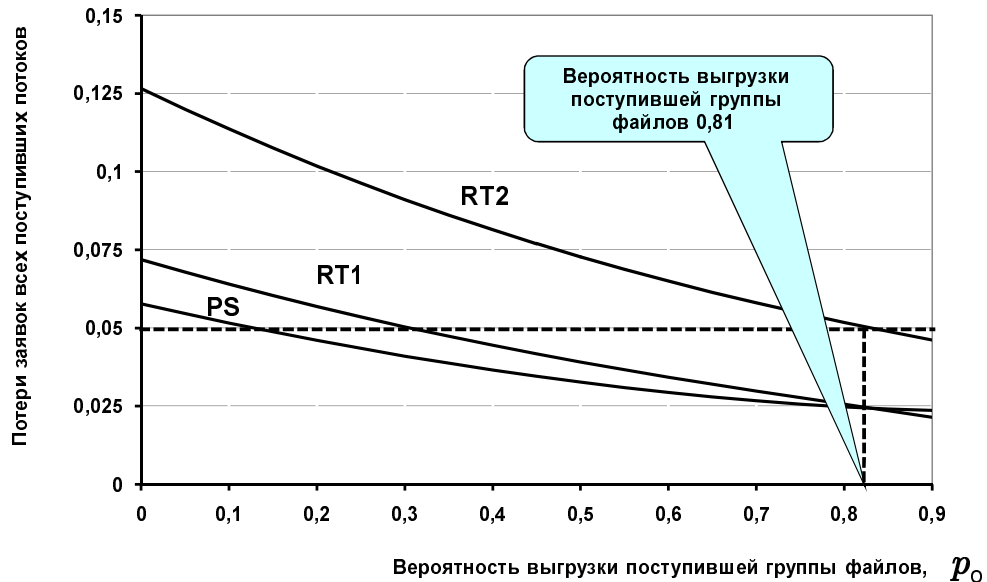


Рисунок 4.15 — Результаты подбора вероятности выгрузки поступившей группы файлов, обеспечивающей обслуживание трафика онлайн-услуг и трафика файлов, допускающих переменную скорость передачи, с потерями, не превосходящими 0,05

4.4 Особенности планирования пропускной способности узлов доступа

4.4.1 Оценка влияния группового поступления поступления файлов

Построенная в диссертации модель учитывает ряд особенностей формирования и обслуживания трафика данных, формируемого в процессе реализации концепции Интернета Вещей (IoT). Исследуем влияние указанных особенностей на расчёт ресурсов, требуемых для обслуживания IoT-трафика с заданным качеством сервиса (QoS). К ключевым особенностям относятся: импульсный характер поступления файлов, возможность адаптации скорости передачи файлов под имеющийся свободный ресурс узла доступа, возможная утрата ценности информации с течением времени, возможность задержки информационного потока началом обслуживания. Как было получено в предыдущих разделах диссертации, наибольший вклад в оценку необходимого объёма ресурса вносят: групповой характер поступления файлов, эластичность их обслуживания. Анализ влияния задержки

на обслуживание представлен в подразделе 4.3.2. Для упрощения моделирования зависимостей характеристик далее рассматривается только групповой поток файлов.

Параметры поступления и обслуживания файлов определим из соотношений:

- Качество обслуживания файлов зададим значением $\pi_d = 0,03$.
- Размер передаваемого файла подчиняется экспоненциальному распределению со средним значением $F = 100$ байт.
- Минимальная скорость передачи одного файла задана как $c = 100$ кбит/с.
- Параметры группового поступления файлов определяются соотношениями: $f_i = 1/5$, где $i = 1, 2, 3, 4, 5$, что соответствует среднему значению $\bar{b} = 3$.
- Предполагается, что один датчик отправляет файлы с интенсивностью $\gamma = 1$ файл/мин. При общем числе IoT-устройств $N_d = 37500$ групповая интенсивность поступления файлов составляет $\lambda_d = \frac{\gamma N_d}{\bar{b}} = 208,34$ групп файлов в секунду, а среднее время передачи файла на минимальной скорости определяется из выражения $\frac{1}{\mu_d} = \frac{F}{c} = 0,008$ с = 8 мс.

На рисунке 4.16 представлены результаты расчётов для указанных параметров. Сравниваются две модели: эквивалентная модель Эрланга с интенсивностью предложенного трафика $a = \lambda_d \cdot 5 \cdot \bar{b} \cdot \frac{1}{\mu_d} = 5$ Эрланг и модель с импульсным характером поступления файлов, в которой чтобы упростить анализ не учитывались эластичность трафика, очередь ожидания и старение данных — только групповой характер поступления. Результаты вычислений, представленные на рисунке 4.16, показывают, что отказ от учета импульсных свойств поступления файлов приводит к существенной (порядка одной трети от требуемого значения) погрешности в определении необходимого объема пропускной способности узла доступа.

Учет эластичности трафика в модели передачи файлов помогает уменьшить требуемый объем ресурсов для передачи данных и частично компенсировать негативное влияние импульсного характера генерации трафика. Для демонстрации этого эффекта приведен численный пример. Результаты расчетов представлены на рисунке 4.17 при тех же исходных параметрах, что и на рисунке 4.16, где анализировались потери при групповой передаче файлов. В модели не учитывались механизмы ожидания и устаревания информации.

Потребность в ресурсе снизилась до значения $C = 1,1$ Мбит/с, практически сравнявшись с результатом, полученным для модели Эрланга (см. рисунок 4.16). Этот факт подчеркивает эффективность использования дисциплины PS при передаче эластичного трафика.

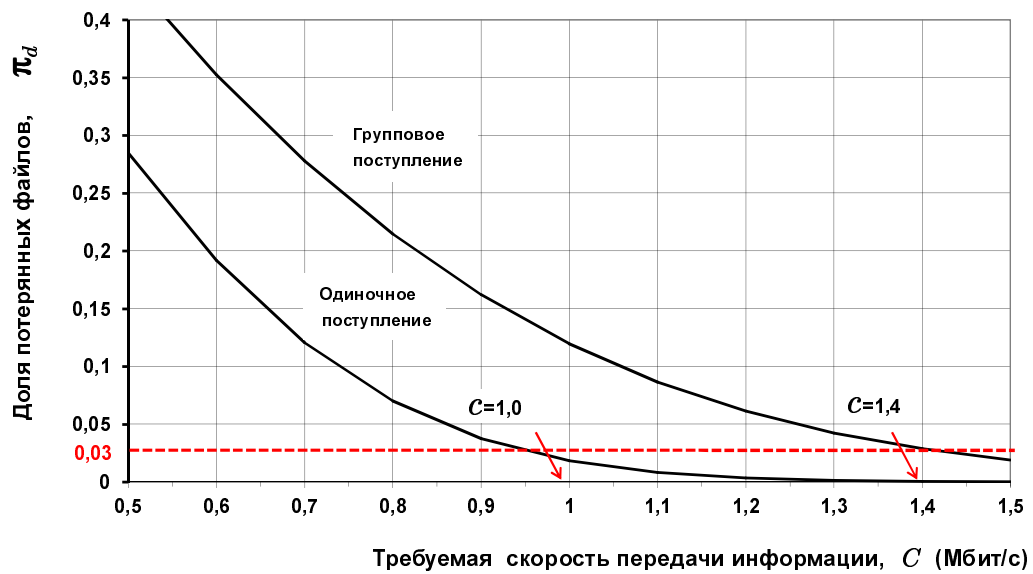


Рисунок 4.16 — Зависимость оценки необходимого объема ресурса передачи от учета импульсного характера поступления данных

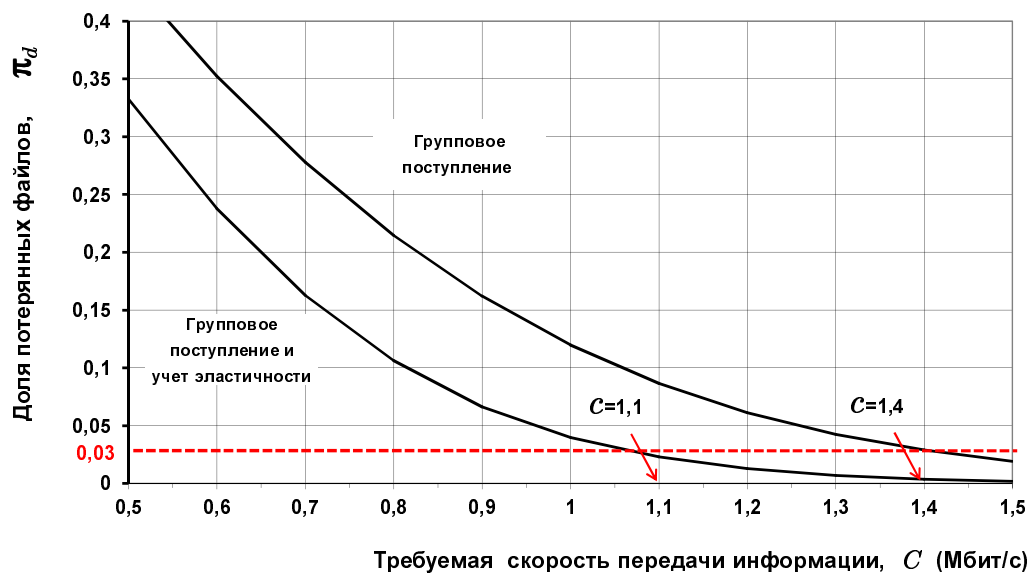


Рисунок 4.17 — Зависимость оценки необходимого объема ресурса передачи от учета импульсного характера поступления данных и использования дисциплины PS при разделении ресурса

4.4.2 Точные и приближенные алгоритмы оценки ресурса передачи информации

Оценка необходимого объема ресурса осуществляется методом перебора, начиная с некоторого минимального значения, которое определяется после анализа входных параметров поступающих информационных потоков. Далее с использованием точного или приближенного алгоритма оценивается значение применяемого в расчетах функционала качества обслуживания поступающих заявок и реализуется процесс выбора требуемого значения пропускной способности узла доступа. Следует отметить, что процедура планирования достаточно сложна, требует учета многих факторов и здесь будет рассмотрена в упрощенном виде, в частности, без учета экономических аспектов организации процесса передачи информации.

Перечень исходных данных для моделирования:

1. Параметры входных потоков и условий обслуживания. Данные определяются на основе маркетинговых исследований и характеристик используемых технологий передачи информации. Включают:
 - Количество потоков заявок на сервисы реального времени n и скорость одного виртуального канала c . Для каждого потока k ($k = 1, \dots, n$) задаются:
 - частота появления заявок $\lambda_{r,k}$;
 - математическое ожидание времени обслуживания заявки $h_{r,k} = \frac{1}{\alpha_{r,k}}$;
 - необходимая по условиям пересылки скорость передачи c_k ;
 - количество виртуальных каналов b_k , предоставляемых на обслуживание поступившего запроса k -го потока.
 - Характеристики поступления файлов:
 - интенсивность поступления групп файлов с эластичным трафиком λ_d ;
 - средний объем файла F ;
 - вероятность f_s того, что в поступившей группе содержится s файлов;
 - среднее количество файлов в группе $\bar{b}$;
 - w число возможностей для заблокированных файлов ожидать начала обслуживания;
 - $\frac{1}{\alpha_d}$ среднее время передачи файла на одном канале;
 - $\frac{1}{\sigma}$ ограничение на время ожидания начала обслуживания.
 - Интенсивности поступления заявок и средние времена обслуживания выражены в единицу времени, в качестве которой обычно выбирается среднее время обслуживания одного из поступающих запросов. Все перечисленные параметры фиксированы на время решения задачи оценки ресурса.
 - Формулировка функционала, определяющего качество обслуживания. В контексте данного исследования таким функционалом будет величина максимальных потерь у поступающих заявок. В зависимости от постановки задачи могут

быть выбраны и другие функционалы. Задается требуемое оператором нормативное значение функционала.

2. Задается начальное значение ресурса r_0 , выраженное в виртуальных каналах. Обычно в качестве такового берется целая часть от среднего числа потенциально занятых виртуальных каналов

$$r_0 = \lfloor \frac{\lambda_{r,1}}{\alpha_{r,1}} b_1 + \dots + \frac{\lambda_{r,n}}{\alpha_{r,n}} b_n + \frac{\lambda_d}{\alpha_d} \bar{b} \rfloor.$$

3. Производится расчёт значения функционала с использованием численных методов, развитых в 3-м разделе диссертации.
4. Значение функционала сравнивается с нормативной величиной. Оценка требуемого числа каналов v заканчивается или продолжается.
5. Трудоемкость оценки необходимого числа каналов определяется числом шагов перебора потенциальных вариантов решения сформулированной задачи.

Пример реализации сформулированного подхода рассмотрен на рисунке 4.12. Оценка характеристик выполняется решением СУР (2.8) итерационным методом Гаусса-Зейделя. Это не всегда удобно при решении задач планирования ресурса узла доступа поскольку число шагов реализации итерационного алгоритма может быть велико. Эту же задачу можно решить и приближенным алгоритмом за существенно меньшее время. Изложение соответствующего алгоритма приведено в подразделе 3.4. Рассмотрим эффективность и погрешность решения анализируемой задачи указанным методом.

Упрощенная процедура планирования пропускной способности узла доступа состоит из двух этапов. На первом этапе, используя результаты подразделов 3.3.1 и 3.3.2, где численно показано, что изменение процедуры распределения ресурса при обслуживании файлов с Processor Sharing на дисциплину обслуживания трафика сервисов реального времени (т.е. передача файлов ведется на минимальном ресурсе в один канал) при сохранении прочих параметров потоков приводит к оценкам сверху для величин потерь. Эти оценки могут быть рассчитаны с использованием рекурсии (3.13). При этом групповой поток файлов меняется на пуассоновский ординарный поток файлов интенсивности $\lambda_d \bar{b}$. Результаты реализации этой процедуры показаны на рисунке 3.8 (правая часть).

На рисунке 4.18 рассмотрена реализация этого подхода для модели узла, представленной на рисунке 4.12. Расширена область изменения пропускной способности v и добавлена кривая зависимости максимальных потерь запросов, полученная с использованием сформулированной выше аппроксимации. Из приведенной оценки следует, что для оценки пропускной способности узла с уровнем максимальных потерь, не превосходящем 0,01, достаточно выбрать ресурс узла из соотношения $v = 223$ в.к. Точные значения потерь поступающих запросов, представленные на рисунке 4.18, подтверждают этот результат. В этой области значений входных параметров величина нагрузки на канал $\rho = \frac{150}{223} = 0,67$. Следовательно можно применить приближенную процедуру оценки параметров обслуживания файлов на числе каналов, которые в соответствии с рекомендациями подраздела 3.4

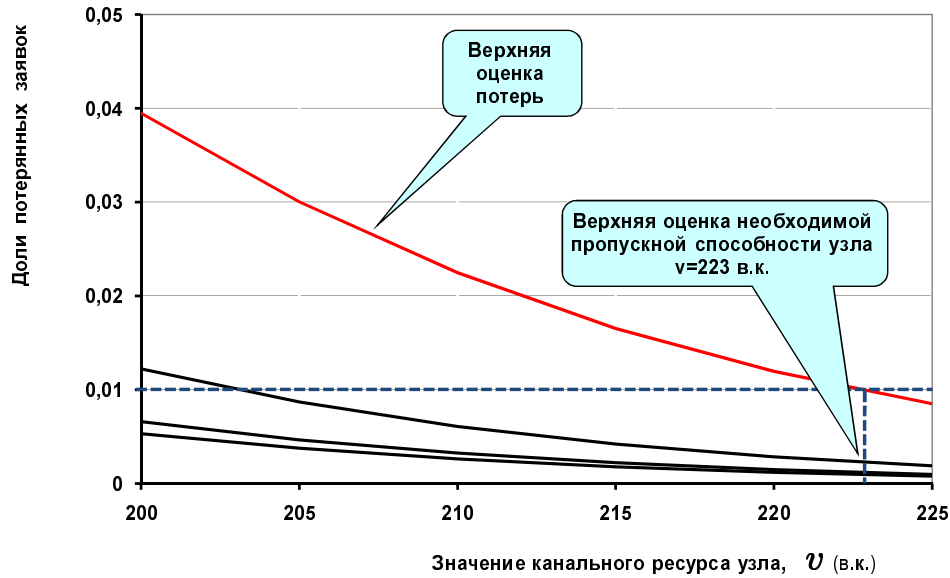


Рисунок 4.18 — Верхняя оценка необходимой пропускной способности узла доступа для обеспечения уровня потерь, не превосходящего 0,01

находятся из соотношения $223 - 99 = 124$ в.к., где 99 — это среднее число каналов, занятых на обслуживание трафика реального времени. Эта величина найдена с использованием рекурсии (3.13). Далее реализуется второй этап — оценка параметров обслуживания файлов. При проведении вычислений используются материалы подраздела 3.3.5, в частности определения характеристик (3.17) и алгоритм их вычисления (3.21). Приближенное значение $h_d \approx 0,055$, точное, найденное с использованием итерационного метода $h_d = 0,0689$. Погрешность примерно 20%, что приемлемо для практических приложений, имея ввиду эффективность реализации приближенного алгоритма.

4.5 Выводы по результатам четвертого раздела

1. Выполнен анализ влияния неоднородности входных потоков на оценку пропускной способности узла доступа. Во всех рассмотренных случаях наличие свойства гетерогенности влияет на значения основных характеристик совместного обслуживания поступающих потоков заявок, а, следовательно, и на решение задачи планирования ресурса узла доступа, которая осуществляется с использованием этих характеристик. Это влияние зависит от нагрузки и от используемых сценариев формирования потоков заявок и распределения ресурса передачи информации, но во всех исследованных случаях легкий трафик вытесняет из обслуживания тяжелый трафик.
2. Из всех изученных в расчетах факторов наиболее сильно влияют на значения характеристик наличие ожидания (см. рисунки 4.2, 4.6) и групповой характер поступления

файлов (см. рисунки 4.3, 4.7). Здесь, в отличие от основной модели, наблюдается рост доли потерянных файлов, что частично снижает потери запросов на передачу трафика реального времени и повышает степень загрузки канала данным типом трафика.

3. Использование дисциплины Processor Sharing уменьшает долю файлов, находящихся на обслуживании. Это по сравнению с основной моделью увеличивает шансы трафика сервисов реального времени попасть на обслуживание. В результате у подобных заявок уменьшаются потери и увеличивается коэффициент использования канала. Это ведет к увеличению потерь заявок на передачу файлов и уменьшению коэффициента использования канала.
4. В целом давая оценку анализируемым факторам отметим, что наличие неоднородности в совместно обслуживаемых потоках трафика и групповой характер поступления файлов увеличивают требование к ресурсу передачи достаточному для обслуживания поступающих потоков заявок с заданным качеством. В тоже время использование ожидания и дисциплины распределения ресурса Processor Sharing уменьшают требование к ресурсу.
5. Условия по дифференцированному обслуживанию информационных потоков в узле доступа можно создать разными способами. Наиболее простой сценарий — это воспользоваться возможностями, которые предоставляет архитектура Network Slicing. Положительные моменты этого решения — это использование относительно простых моделей и алгоритмов расчета характеристик обслуживания трафика в отдельных слайсах. Отрицательные — избыточные объемы ресурса передачи информации и, как следствие, его неэффективное использование.
6. Другие возможности по дифференцированному обслуживанию трафика заключаются в контроле за доступом поступающих заявок. Контроль может быть организован за счет реализации возможности ожидания и осуществления выгрузки отдельных видов заявок в другие узлы доступа или в другие более доступные диапазоны частот. Оценка условий реализации перечисленных возможностей может быть осуществлена с помощью построенных моделей и разработанных на их основе алгоритмов оценки показателей качества обслуживания поступающих информационных потоков.
7. Сформулирована процедура оценки необходимой по нагрузке и качеству обслуживания величины пропускной способности узла доступа. Оценка необходимого объема ресурса осуществляется методом перебора, начиная с некоторого минимального значения, которое определяется после анализа входных параметров поступающих информационных потоков. Далее с использованием точного или приближенного алгоритма оценивается значение применяемого в расчетах функционала качества обслуживания поступающих заявок и реализуется процесс выбора требуемого значения пропускной способности узла доступа. Следует отметить, что процедура планирования достаточно сложна, требует учета многих факторов и здесь будет рассмотрена

в упрощенном виде, в частности, без учета экономических аспектов организации процесса передачи информации. Для точного решения сформулированной задачи используется метод основанный на решении системы уравнений равновесия. Приближенный подход основан на частных случаях построенной модели узла доступа. С вычислительной точки зрения он значительно эффективнее. Погрешность оценки составляет примерно 20%.

Заключение

Основные результаты работы состоят в следующем:

1. Разработана и исследована комплексная модель совместного обслуживания разнородного трафика в беспроводном узле доступа. В отличие от существующих аналогов, данная модель позволяет учитывать ключевые особенности формирования и передачи данных, характерные для современных и перспективных коммуникационных сервисов. К таким особенностям относятся: зависимость скорости передачи от типа запроса, приоритетность нагрузки онлайн-услуг, возможность поступления файлов группами случайного размера, наличие или отсутствие у файлов эластичных свойств, присутствие у файлов возможности ожидания начала передачи и утраты ценности передаваемой информации с течением времени.
2. С помощью построенной математической модели узла получены выражения для оценки основных характеристик совместной передачи поступающих информационных потоков. Это доля запросов, получивших отказ, средний объем использованного ресурса передачи информации, среднее время передачи файла, среднее число файлов, ожидающих начала передачи. В отличие от более ранних исследований полученные выражения позволяют изучить действие учтенных в модели особенностей формирования потоков заявок и их обслуживания на решение задачи планирования ресурса и создание условий по их дифференцированному обслуживанию.
3. Стационарные вероятности модели и связанные с ними характеристики можно определить, решая систему уравнений баланса. В отличие от стандартных методов решения этой задачи, разработанный в диссертации алгоритм на основе метода последовательных замещений Гаусса-Зейделя позволяет рассчитывать характеристики моделей, включающих до нескольких миллионов состояний.
4. Для приближенной оценки характеристик узла предложена процедура декомпозиции модели на две составляющие. В каждой из них рассматривается по отдельности процесс обслуживания трафика сервисов реального времени и пересылка эластичного трафика файлов. В отличие от предыдущих исследований, расчет параметров обслуживания входящих запросов можно проводить с использованием несложных рекуррентных формул, применимых при любых значениях структурных параметров. Алгоритм ориентирован на реализацию в области малых потерь, где решается задача планирования требуемой пропускной способности узла доступа. Показано, что для потерь порядка 0,01–0,05 погрешность оценки составляет 10%–20%. Построена и численно исследована процедура оценки нагрузки, где можно использо-

вать рассмотренную процедуру декомпозиции. Для этого применяется модель узла, обеспечивающая верхний характер оценок потерь заявок.

5. Исследовано влияние учитываемых в модели особенностей формирования и обслуживания поступающих заявок на расчет величины требуемой пропускной способности узла доступа. В целом, давая оценку анализируемым факторам отметим, что наличие неоднородности в совместно обслуживаемых потоках трафика и групповой характер поступления файлов увеличивают требование к необходимой величине пропускной способности узла доступа достаточной для пересылки поступающих информационных потоков заданным качеством. В тоже время применение возможности ожидания для заблокированных заявок и использование дисциплины равномерного распределения ресурса Processor Sharing уменьшают требование к ресурсу. Отмеченные изменения в значениях характеристик можно оценить с использованием построенной модели и разработанных на ее основе алгоритмов. Так установлено, что игнорирование импульсного характера поступления файлов приводит к существенной (порядка одной трети от требуемого значения) ошибке в оценке необходимого объема ресурса.
6. Показано, что в ситуации перегрузки условия по дифференцированному обслуживанию эластичного трафика можно создать на общем ресурсе без увеличения пропускной способности узла доступа, если воспользоваться возможностью ожидания файлов или их выгрузки на другие узлы. Оценка условий реализации перечисленных возможностей может быть осуществлена с помощью построенных моделей и разработанных на их основе алгоритмов оценки показателей качества обслуживания поступающих информационных потоков.

Таким образом, в результате проведенных в диссертационной работе исследований, решена научная задача, имеющая значение для развития сетей связи, — построена и проанализирована комплексная модель мультисервисного беспроводного узла доступа, которая дает возможность решить задачу планирования необходимой по нагрузке пропускной способности узла, а также разработать и исследовать процедуры дифференцированного обслуживания поступающих информационных потоков с учетом их гетерогенных свойств и импульсного характера формирования трафика данных. Тем самым, цель диссертационного исследования достигнута.

Список литературы

1. Анисимов А. Калькулятор для расчета пропускной способности LTE [Электронный ресурс] / А. Анисимов // anisimoff / Режим доступа: <https://anisimoff.org/lte>
2. Башарин Г.П. Лекции по математической теории телетрафика / Г.П. Башарин. — М.:РУДН, 2009. — 342 с.
3. Бакулин М.Г. Технология OFDM: учебное пособие для вузов / М.Г. Бакулин, В.Б. Крейнделин, А.М. Шлома, А.П. Шумов. — М.: Горячая линия–Телеком, 2015. — 360 с.
4. Бегишев В.О. Стратегия распределения радиоресурсов в гетерогенных сетях с трафиком Narrow-Band IoT / В.О. Бегишев, А.К. Самуйлов, Д.А. Молчанов, К.Е. Самуйлов // Системы и средства информатики. — 2017. — Т. 27, №. 4. — С. 64–79.
5. Васин Н.Н., Вострикова В.А., Диязитдинов Р.Р. Основы построения инфокоммуникационных систем и сетей: учебник для вузов. — Самара: ПБГУТИ, 2017. — 220 с.
6. Власкина А.С. Управляемая система массового обслуживания для анализа динамической нарезки радиоресурсов в сети 5G // В сборнике: Информационно - телекоммуникационные технологии и математическое моделирование высокотехнологичных систем. Материалы Всероссийской конференции с международным участием. — Москва. — 2022. — С. 34–38.
7. Вэнь Т., Пейин Ч. Сети 6G. Путь от 5G к 6G глазами разработчиков. — Москва: ДМК пресс, 2021. — 632 с.
8. Гельгор А.Л. Технология LTE мобильной передачи данных: учеб. пособие / А.Л. Гельгор, А. Попов. — СПб.: Изд-во Политехн. ун-та, 2011. — 204 с.
9. Долгун В. Переход от LTE-Advanced к 5G / В. Долгун, О. Козлова // Беспроводные технологии. — 2016. — № 2. — С. 12–14.
10. Дроздова В.Г. Алгоритмы формирования и приема данных канала RVCH сетей 5G / В.Г. Дроздова, Д.В. Завьялова // Вестник СибГУТИ. — 2021. — Т. 2. — № 54. — С. 3–13.
11. Канищева М.Г. Планирование ресурсов системы наблюдения при наличии контроля доступа на основе объема занимаемых ресурсов по транспортным потокам / Канищева М.Г., Степанов С.Н. // Технологии информационного общества. Сборник трудов XXX Международной отраслевой научно-технической конференции «Технологии информа-

- ционного общества». 20-21 марта 2022 г. Москва, МТУСИ. — М.: ИД МБА. — 2022. — С. 107–113.
12. Клейнрок Л. Теория массового обслуживания. — М.: Машиностроение, 1979. — 452 с.
 13. Корнышев Ю.Н. Теория телетрафика: учебник для вузов / Ю.Н. Корнышев, А.П. Пшеничников, А.Д. Харкевич. — М.: Радио и связь, 1996. — 272 с.
 14. Колмогорова С.С. Технологии стандарта 5G NR: учебное пособие / С.С. Колмогорова, А.К. Бойцов, В.С. Павлов. — Санкт-Петербург: Реноме, 2022. — 72 с.
 15. Колмогорова С.С. Интернет вещей. Общие принципы: учебное пособие для студентов высших учебных заведений / С.С. Колмогорова, С.В. Бирюков, А.К. Бойцов. Санкт-Петербург: Реноме, 2022. — 108 с.
 16. Лагутин В.С. Телетрафик мультисервисных сетей связи / В.С. Лагутин, С.Н. Степанов // М.: Радио и связь, 2000. — 320 с.
 17. Лапцевич А.А. Проектирование системы мобильной связи стандарта LTE при создании эталонной сети сотовой подвижной электросвязи на территории Республики Беларусь / А.А. Лапцевич, С.И. Половения // Проблемы инфокоммуникаций. — 2019. — Т. 1. — № 9. — С. 28–35.
 18. Лохвицкий М.С. Сотовая связь: от поколения к поколению / М.С. Лохвицкий, Н.С. Мардер. — М.: издательство ИКАР, 2014. — 236 с.
 19. Молчанов Д.А. Сети 5G/6G: Архитектура, технологии, методы анализа и расчета / Д.А. Молчанов, В.О. Бегипшев, К.Е. Самуйлов, Е.А. Кучерявый. — Москва: РУДН, 2022. — 516 с.
 20. Наумов В.А. Теория телетрафика мультисервисных сетей / В.А. Наумов, К.Е. Самуйлов, Н.В. Яркина // — М.: Изд-во РУДН, 2007. — 191 с.
 21. Попов В.И. Основы проектирования сотовых сетей мобильной связи / В.И. Попов, В.А. Скуднов // — Москва: Горячая линия-Телеком, 2017. — 398 с.
 22. Пшеничников А.П. Концептуальные основы будущих сетей / А.П. Пшеничников, И.М. Даудов // REDS: Телекоммуникационные устройства и системы. — 2022. — Т. 12. — № 2. — С. 24–28.
 23. Приказ Минкомсвязи России № 923 «Об утверждении Концепции создания и развития сетей 5G/IMT-2020 в Российской Федерации» [Электронный ресурс] // Минцифры / Режим доступа: <https://digital.gov.ru/ru/documents/6990> (дата обращения: 27.03.2025).
 24. Росляков А.В. Сети фиксированной связи пятого поколения: учебное пособие / А.В. Росляков. — М.: Колос-с ИКЦ, 2024. — 232 с.

25. Рыков В.В. Управляемые системы массового обслуживания. В кн.: Теория вероятностей. Математическая статистика. Теоретическая кибернетика. Итоги науки и техники. М.:ВИНИТИАН, 1975. — Т. 12. — С. 43–153.
26. Рыжков А.Е. Системы и сети радиодоступа 4G: LTE, WiMax / А.Е. Рыжков, М.А. Сиверс, В.О. Воробьев и др. СПб: Линк, 2012. — 226 с.
27. Свидетельство о государственной регистрации программы для ЭВМ № 2020665367. Программа для оценки характеристик обслуживания гетерогенного трафика Интернета Вещей с контролем доступа. / С.Н. Степанов, М.С. Степанов, М.О. Шишкин, М.Г. Канищева; правообладатель: Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования // Федеральная служба по интеллектуальной собственности (РОСПАТЕНТ). — 2020. Заявка № 2020664744 от 17.11.2020.
28. Свидетельство о государственной регистрации программы для ЭВМ № 2023610805. Программа для оценки характеристик обслуживания гетерогенного трафика / М.Г. Канищева, М.С. Степанов, С.Н. Степанов, М.О. Шишкин; правообладатель: Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования // Федеральная служба по интеллектуальной собственности (РОСПАТЕНТ). — 2023. Заявка № 2022686658 от 31.12.2022.
29. Свидетельство о государственной регистрации программы для ЭВМ № 2023688868. Программа для исследования процедур маршрутизации запросов на информационное обслуживание в справочных и ситуационных центрах, работающих в условиях перегрузки / С.Н. Степанов, М.С. Степанов, М.О. Шишкин, М.Г. Канищева; правообладатель: Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования // Федеральная служба по интеллектуальной собственности (РОСПАТЕНТ). — 2023. Заявка № 2023688131 от 15.12.2023.
30. Свидетельство о государственной регистрации программы для ЭВМ № 2024667400. Программа для оценки пропускной способности мультисервисного узла доступа / С.Н. Степанов, М.Г. Канищева, М.С. Степанов, М.О. Шишкин; правообладатель: Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования // Федеральная служба по интеллектуальной собственности (РОСПАТЕНТ). — 2024. Заявка № 2024666386 от 15.07.2024.
31. Степанов С.Н. Построение и анализ модели группового поступления данных IoT в беспроводном узле доступа / С.Н. Степанов, М.Г. Канищева // Труды международной научно-технической конференции «Телекоммуникационные и вычислительные системы – 2023» — М.: Горячая линия – Телеком. — 2023. — С. 84–87.
32. Степанов С.Н. Эффективный алгоритм оценки требуемого объема ресурса беспроводных систем связи при совместном обслуживании гетерогенного трафика устройств интернета вещей / С.Н. Степанов, М.С. Степанов // Автоматика и телемеханика. —

2019. — № 11. — С. 108–126.
33. Степанов С.Н. Планирование ресурса передачи при совместном обслуживании мультисервисного трафика реального времени и эластичного трафика данных / С.Н. Степанов, М.С. Степанов // Автоматика и телемеханика. — 2017. — № 11. — С. 79–93.
 34. Степанов С.Н. Теория телетрафика: концепции, модели, приложения / С.Н. Степанов. М.: Горячая линия – Телеком, 2015. — 868 с.
 35. Степанов С.Н. Методы оценки необходимого объема ресурса мультисервисных узлов доступа / С.Н. Степанов, М.С. Степанов // Автоматика и телемеханика. — 2020. — № 12. — С. 129–152.
 36. Степанов С.Н. Модель совместного обслуживания трафика сервисов реального времени и трафика данных / С.Н. Степанов // Автоматика и телемеханика. — 2011. — № 4. — С. 121–132.
 37. Тихвинский В.О. Сети мобильной связи LTE/LTE Advanced / В.О. Тихвинский, С.В. Терентьев, В.П. Высочин. Москва: Медиа Паблишер, 2014. — 384 с.
 38. Тихвинский В.О. Развитие сетей мобильной связи от 5G Advanced к 6G; проекты, технологии, архитектура / В.О. Тихвинский, С.В. Терентьев, В.А. Коваль, Е.Е. Девяткин. Москва: ТЕХНОСФЕРА, 2023. — 528 с.
 39. Яшков С.Ф. The M/D/1 processor sharing queue revisited / С.Ф. Яшков // Информационные процессы. — 2009. — Т. 9. — № 3. — С. 216–223.
 40. Abu-Ali N. Uplink Scheduling in LTE and LTE-Advanced: Tutorial, Survey and Evaluation Framework / N. Abu-Ali, A.M. Taha, M. Salah, H. Hassanein // IEEE Communications Surveys and Tutorials. — 2014. — V. 16. — No. 3. — P. 1239–1265.
 41. Afolabi I. Network Slicing and Softwarization: A Survey on Principles, Enabling Technologies and Solutions / I. Afolabi, T. Taleb, K. Samdanis, [et. all] // IEEE Communications Surveys and Tutorials. — 2018. — V. 20. — No. 3. — P. 2429–2453.
 42. Ang L. Wireless Multimedia Sensor Networks on Reconfigurable Hardware: Information Reduction Techniques / L. Ang, K.P. Seng, L.S. Yeong, W.C. Chia. Berlin: Springer Science + Business Media, 2013. — 383 p.
 43. Andrabi U.M. Cellular network resource distribution methods for the joint servicing of real-time multiservice traffic and grouped IoT traffic / U.M. Andrabi, S.N. Stepanov, J. Ndayikunda, M.G. Kanishcheva // T-Comm: Телекоммуникации и Транспорт. — 2020. — V. 14. — No. 10. — P. 61–69.
 44. Andrabi U.M. The model of conjoint servicing of real time and elastic traffic streams through processor sharing (PS) discipline with access control / Andrabi U.M., Stepanov S.N., Stepanov M.S., [et. all.] // 2021 International Conference Engineering and Telecommunication. Moscow, Russia. — 2021. — P. 1–5.

45. Andrabi U.M. The Model of Conjoint Servicing of Real-Time Traffic of Surveillance Cameras and Elastic Traffic Devices with Access Control / U. M. Andrabi, S. N. Stepanov // 2nd International Informatics and Software Engineering Conference (IISEC). Moscow, Russia. — 2021. — P. 1–6.
46. Andrabi U.M. Observation system resource planning in presence of access control based on volume of resource occupied by traffic flows / U.M. Andrabi, M.G. Kanishcheva, S.N. Stepanov // T-Comm: Телекоммуникации и Транспорт. — 2022. — V. 16. — No. 8. — P. 54–62.
47. Akyildiz I.F. A flexible network architecture for QoE-aware communications in 5G systems / I.F. Akyildiz, A. Kak, E. Khorov, [et. all] // Computer Networks, — 2018. — No. 147. — P. 262–279.
48. Ameigeiras P. Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges / P. Ameigeiras, J. Ordonez-Lucena, D. Lopez, [et. all] // IEEE Communications Magazine. — 2017. — V. 55. — No. 5. — P. 80–87.
49. Adou K.Y. Methods for Analyzing Slicing Technology in 5G Wireless Network Described as Queueing System with Unlimited Buffer and Retrial Group / K.Y. Adou, E.V. Markova // Discrete and Continuous Models and Applied Computational Science. — 2021. — P. 264–278.
50. Ageev K. Modelling of Virtual Radio Resources Slicing in 5G Networks / K. Ageev, A. Garibyan, A. Golskaya, [et. all] // Communications in Computer and Information Science. — 2019. — V. 1109. — P. 150–161.
51. Ageev K. Simulation of the Limited Resources Queueing System with Signals / K. Ageev, Ed. Sopin, K. Samouylov // 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). — 2018. — P. 1–6.
52. Ageev K.A. Analysis of the simplified Network Slicing model / K.A. Ageev, E.S. Sopin // Информационные Технологии И Математическое Моделирование (ИТММ-2020). — 2020. — P. 11–16.
53. Basharin G.P. Mathematical theory of teletraffic and its application to the analysis of multiservice communication of next generation networks / G.P. Basharin, Yu.V. Gaidamaka, K.E. Samouylov // Automatic Control and Computer Sciences. — 2013. — V. 47. — No. 2. — P. 62–69.
54. Baena C. Estimation of video streaming KQIs for radio access negotiation in network slicing scenarios / Baena C., Fortes S., Baena E., Barco R. // IEEE Communications Letters. — 2020. — V. 24. — No. 6. — P. 1304–1307.
55. Begishev V. Resource Allocation and Sharing for Heterogeneous Data Collection over Conventional 3GPP LTE and Emerging NB-IoT Technologies / V. Begishev, V. Petrov, A. Samouylov, [et. all.] // Computer Communications. — 2018. — V. 120. — No. 2. — P. 93–101.

56. Blanco B. Technology Pillars in The Architecture of Future 5G Mobile Networks: NFV, MEC and SDN / B. Blanco, J.O. Fajardo, I. Giannoulakis, [et al.] // Computer Standards and Interfaces archive. — 2017. — V.54. — P. 216–228.
57. Bonald T. A recursive formula for multirate systems with elastic traffic / T. Bonald, J. Virtamo // IEEE Communications Letters. — 2005. — V. 9. — No. 8. — P. 753–755.
58. Bonald T. Insensitive traffic models for communication networks / T. Bonald // Discrete Event Dynamic Systems. — 2007. — V. 17. — No. 3. — P. 405–421.
59. Brown M.A. Traffic Control HOWTO [electronic resource] /M.A. Brown // Retrieved from <https://tldp.org/HOWTO/>
60. Benzekki K. Software-defined networking (SDN): A survey / K. Benzekki, A. El Fergougui, A. E. Elalaoui // Security and Communication Networks. — 2016. — V. 9. — No. 18. — P. 5803–5833.
61. Chen J. Resource Allocation and Computation Offloading for Multi-Access Edge Computing With Fronthaul and Backhaul Constraints / J. Chen, Z. Chang, X. Guo, [et. all] // IEEE Transactions on Vehicular Technology. — 2021. — V. 70. — No. 8. — P. 8037–8049.
62. Dawood T.Z. The analysis of heterogeneous traffic servicing in the Cellular Internet of Things access node / Dawood T.Z., Stepanov M.S., Stepanov S.N., Kanishcheva M.G., Shishkin M.O. // 2024 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO). Moscow, Russia. — 2024. — P. 1–5.
63. Foresti G.L. Multimedia Video-Based Surveillance Systems: Requirements, Issues and Solutions / G.L. Foresti, P. Mahonen, C.S. Regazzoni. – US: The Springer International Series in Engineering and Computer Science, 2012. — 289 p.
64. Gomes R.L. Reliability-Aware Network Slicing in Elastic Demand Scenarios / R.L. Gomes, L.F. Bittencourt, E.R.M. Madeira // IEEE Communications Magazine. — 2020. — V. 58. — No. 10. — P. 29–34.
65. Galiaskarov D.F. Dynamic load balancing in data center / D.F. Galiaskarov, S.N. Stepanov, N.N. Nikolaychuk, [et. all] // 2024 Systems of Signals Generating and Processing in the Field of on Board Communications. Moscow, Russia. — 2024. — P. 1–5.
66. Howard R.A. Dynamic programming and markov processes / Howard R.A. // Cambridge: Technology Press of Massachusetts Institute of Technology, 1960. — 136 p.
67. Holma H. LTE for UMTS - OFDMA and SC-FDMA Based Radio Access / H. Holma, A. Toskala. New Jersey: Wiley and sons, 2009. — 450 p.
68. Jean Thierry Stephen Avocanh. An enhanced two level scheduler to increase multimedia services performance in LTE networks / J.T.S. Avocanh, M. Abdennebi, J. Ben-Othman // IEEE International Conference on Communications (ICC). — 2014. — P. 2351–2356

69. Kaloxylos A. View on 5G Architecture / A. Kaloxylos, A. Galis, P. Rost, [et. all]. 5G Public Private Partnership (5G PPP) Architecture Working Group, 2017. — 140 p.
70. Khan R. Survey on Security and Privacy of 5G Technologies: Potential Solutions, Recent Advancements, and Future Directions / R. Khan, P. Kumar, D.N.K. Jayakody, M.A. Liyanage // IEEE Communications Surveys and Tutorials. — 2020. — V. 22. — No. 1. — P. 196–248.
71. Khan S. Highly Accurate and Reliable Wireless Network Slicing in 5th Generation Networks: A Hybrid Deep Learning Approach / S. Khan, Y. Ali, Z. Ullah, [et. all] // Journal of Network and Systems Management. — 2022. — V. 30. — No. 2. — P. 29.
72. Khan H. Enhancing Video Streaming in Vehicular Networks via Resource Slicing / H. Khan, S. Samarakoon, M. Bennis // IEEE Transactions on Vehicular Technology. — 2020. — V. 69. — No. 4. — P. 3513–3522.
73. Kochetkova I. Analyzing the effectiveness of dynamic network slicing procedure in 5G network by queuing and simulation models / I. Kochetkova, A. Vlaskina, S. Burtseva, [et. all] // Lecture Notes in Computer Science (LNCCN). — 2020. — V. 12525. — P. 71–85.
74. Kochetkova I. Controllable Queuing System with Elastic Traffic and Signals for Resource Capacity Planning in 5G Network Slicing / I. Kochetkova, K. Leonteva, I. Ghebrial, [et. all.] // Future Internet. — 2024. — V. 16. — No. 8. — P. 18.
75. Kitaev M.Yu. Controlled queueing systems / M.Yu. Kitaev, V.V. Rykov. N. Y.: CRC Press, 1995. — 304 p.
76. Li X. Network Slicing for 5G: Challenges and Opportunities / X. Li, M. Samaka, H.A. Chan, [et. all] // IEEE Internet Computing. — 2017. — V. 21. — No. 5. — P. 20–27.
77. Li Y. Survey of Spatio-Temporal Interest Point Detection Algorithms in Video / Y. Li, R. Xia, Q. Huang, [et. all] // IEEE Access. — 2017. — V. 5. — P. 10323–10331.
78. Lieto A. Enabling Dynamic Resource Sharing for Slice Customization in 5G Networks / A. Lieto, I. Malanchini, A. Capone // IEEE Global Communications Conference. — 2018. — P. 1–7.
79. Malik H. Radio Resource Management Scheme in NB-IoT Systems / H. Malik, H. Pervaiz, M. Mahtab Alam, [et. all] // IEEE Access. — 2018. — V. 6. — P. 15051–15064.
80. Marochkina A. Ultra-Dense Internet of Things Model Network / A. Marochkina, A. Paramonov, T.M. Tatarnikova // Communications in Computer and Information Science. — 2022. — V. 1552. — P. 111–122.
81. Moltchanov D. Tutorial on Mathematical Modeling of 5G/6G Millimeter Wave and Terahertz Cellular Systems / D. Moltchanov, E. Sopin, V. Begishev, [et. all] // IEEE Communications Surveys and Tutorials. — 2022. — V. 24. — No. 2. — P. 1072–1116.

82. Moskaleva F. Resource Queueing System for Analysis of Network Slicing Performance with QoS-Based Isolation / F. Moskaleva, E. Lisovskaya, Y. Gaidamaka // Communications in Computer and Information Science. — 2021. — V. 1391. — P. 198–211.
83. Muhizi S. A novel slice-oriented network model / S. Muhizi, A.A. Ateya, A. Muthanna, [et. all] // Communications in Computer and Information Science. — 2018. — V. 919. — P. 421–431.
84. Naumov V. Matrix and Analytical Methods for Performance Analysis of Telecommunication Systems / V. Naumov, Yu. Gaidamaka, N. Yarkina, K. Samouylov. Switzerland: Springer Nature, 2021. — 308 p.
85. Parvez I. A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions / I. Parvez, A. Rahmati, I. Gcvcnc, [et. all] // IEEE Communications Surveys and Tutorials. — 2018. — V. 20. — No. 4. — P. 3098–3130.
86. Popovski P. 5G wireless network slicing for eMBB, URLLC, and mMTC: A communication - theoretic view / P. Popovski, K.F. Trillingsgaard, O. Simeone, G. Durisi // IEEE Access. — 2018. — V. 6. — P. 55765–55779.
87. Release description:15 / 3GPP // TS 36.306. Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio access capabilities — Rel. 15. — 2018. — V. 15.1.0.
88. Release description:7 / 3GPP // TS 23.203. Policy and charging control architecture — Rel. 7. — 2024. — V. 19.1.0.
89. Release description:16 / 3GPP // TS 23.501. System architecture for the 5G System. — Rel. 16. — 2018. — V. 16.12.0.
90. Release description:15 / 3GPP // TS 38.306. NR; User Equipment (UE) radio access capabilities. — Rel. 15. — 2023. — V. 15.2.0.
91. Release description:15 / 3GPP // TS 38.101-1. NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone. — Rel. 15. — 2018. — V. 15.2.0.
92. Release description:13 / 3GPP // TR 45.820. Cellular system support for ultra-low complexity and low throughput Internet of Things. — Rel. 13. — 2015. — V. 0.3.0.
93. Release description:15 / 3GPP // TS 36.213. Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures. — Rel. 15. — 2018. — V. 15.2.0.
94. Release description:14 / 3GPP // TR 38.913. Study on scenarios and requirements for next generation access technologies. — Rel. 14. — 2022. — V. 17.0.0.
95. Rost P. Network slicing to enable scalability and flexibility in 5G mobile networks / P. Rost, C. Mannweiler, D.S. Michalopoulos, [et. all] // IEEE Communications Magazine. — 2017. — V. 55. — No. 5. — P. 72–79.

96. Ruby R. Uplink Scheduler for SC-FDMA-Based Heterogeneous Traffic Networks with QoS Assurance and Guaranteed Resource Utilization / R. Ruby, V.C.M. Leung, D.G. Michelson // IEEE Transactions on Vehicular Technology. — 2015. — V. 64. — No. 10. — P. 4780–4796.
97. Sauter M. From GSM to LTE-Advanced an introduction to mobile networks and mobile broadband. United Kingdom: John Wiley and Sons, 2014. — 458 p.
98. Sayadi B. Network Applications: Opening up 5G and Beyond networks / B. Sayadi, C. Chang, C. Tranoris, [et. all]. 5G-PPP Software Network Working Group, 2023. — 78 p.
99. Stepanov S.N. The Analysis of Resource Sharing for Heterogenous Traffic Streams over 3GPP LTE with NB-IoT Functionality / S.N. Stepanov, M.S. Stepanov, U.M. Andrabhi, J. Ndayikunda // Distributed Computing and Computer Networks. DCCN 2020. Lecture Notes Computer Science, Springer, Cham. — 2020. — V. 12563. — P. 422–435.
100. Stepanov M. S. The Development and Analysis of a Service Model for the Traffic of a Surveillance System Operator by a Dedicated Resource of an LTE Cell / M. S. Stepanov, M. G. Kanishcheva, E. E. Malikova, J. Ndayikunda, N. G. Kaigorodov // 2022 Systems of Signals Generating and Processing in the Field of on Board Communications. Moscow, Russia. — 2022. — P. 1–5.
101. Stepanov S.N. Resource allocation model for LTE technology with functionality of NB-IoT and reservation / S.N. Stepanov, J. Ndayikunda, M.G. Kanishcheva // T-Comm: Телекоммуникации и транспорт. — 2021. — V. 15. — No. 11. — P. 69–76.
102. Stepanov S.N. The analysis of traffic balancing for data centers serving requests of LEO mobile satellite systems / S.N. Stepanov, A.V. Korobkina, A.O. Volkov, E.E. Malikova, A.E. Panov // 2021 Systems of Signals Generating and Processing in the Field of on Board Communications. Moscow, Russia. — 2021. — P. 1–6.
103. Stepanov S.N. The construction and analysis of generalized model of resource sharing for LTE technology with functionality of NB-IoT / S. N. Stepanov, M.S. Stepanov, E.E. Malikova,[et. all] // T-Comm: Телекоммуникации и Транспорт. — 2018. — V. 12. — No. 12. — P. 71–77.
104. Stepanov S.N. Construction and analysis of a model of group receipt of IoT elastic data in a wireless access node / S.N. Stepanov, M.G. Kanischeva, V.I. Korotkova, A.P. Pshenichnikov // 2023 Systems of Signals Generating and Processing in the Field of on Board Communications. Moscow, Russia. — 2023. — P. 1–5.
105. Stepanov M.S. Analysis of Procedures for Joint Servicing of Multiservice Traffic in Access Nodes / M.S. Stepanov, S.N. Stepanov, M.G. Kanischeva, F.S. Kroshin // Distributed computer and communication networks: control, computation, communications. Moscow, Russia. — 2022. — P. 1–5.
106. Stepanov M.S. Analysis of Procedures to Ensure the Required QoS Indicators in Multiservice Access Nodes / M.S. Stepanov, S.N. Stepanov, M.G. Kanischeva, F.S. Kroshin //

- Distributed computer and communication networks: control, computation, communications. Moscow, Russia. — 2023. — P. 1–5.
107. Stepanov S.N. Methods for Estimating the Required Volume of Resource for Multiservice Access Nodes / S.N. Stepanov, M.S. Stepanov // Automation and Remote Control. — 2020. — V. 81. — No. 12. — P. 2244–2261.
 108. Stepanov S.N. Markov models with retrials: The calculation of stationary performance measures based on the concept of truncation / S.N. Stepanov // Mathematical and Computer Modelling. — 1999. — No. 30. — P. 207–228.
 109. Stepanov S. Estimation of the Performance Measures of a Group of Servers Taking into Account Blocking and Call Repetition before and after Server Occupation / S. Stepanov, M. Stepanov // Mathematics. — 2021. — V. 9. — No. 21. — P. 1–24.
 110. Stepanov S.N. Reservation Based Joint Servicing of Real Time and Batched Traffic in Inter Satellite Link / Stepanov S.N., Andrabi U.M., Stepanov M.S., Ndayikunda J. // Proc. of 2020 Systems of Signals Generating and Processing in the Field of on Board Communications. Moscow, Russia. — 2020. — P. 1–5.
 111. Stepanov D. The Concept of Video Surveillance System Based on the Principles of Stereo Vision / D. Stepanov, I. Tishchenko // Conference of Open Innovations Association FRUCT. Moscow, Russia. — 2014. — P. 328–334.
 112. Samouylov K.E. Recursive computation for a multi-rate model with elastic traffic and minimum rate guarantees / K.E. Samouylov, I.A. Gudkova // 2010 International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). Moscow, Russia. — 2010. — P. 1065–1072.
 113. Samouylov K.E. Analysis of an admission model in a fourth generation mobile network with triple play traffic / Samouylov K.E., Gudkova I.A. // Automatic Control and Computer Sciences. — 2013. — V. 47. — No. 4. — P. 202–210.
 114. SanMiguel J.C. Self-Reconfigurable Smart Camera Networks / J.C. SanMiguel, C. Micheloni, K. Shoop, [et. all] // Computer. — 2014. — V. 47. — No. 5. — P. 67–73.
 115. Sevastianov L.A. Telecommunication market model and optimal pricing scheme of 5G services / L.A. Sevastianov, S.A. Vasilyev // International Congress on Ultra Modern Telecommunications and Control Systems and Workshops. — 2019. — P. 1–5.
 116. Shorov A. 5G testbed development for network slicing evaluation / Shorov A. // Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). — 2019. — P. 39–44.
 117. Tang Y. Intelligent Video Surveillance System for Elderly People Living Alone Based on ODVS / Y. Tang, B. Ma, H. Yan // Advances in Internet of Things. — 2013. — V. 3. — No. 2. — P. 44–52.

118. Teppei N. What is 5G? What is Network Slicing? [Electronic resource] / teppei log 2021// Access mode: <https://teppeilog.com/whatisnwslicing>
119. Tikhvinskiy V.O. Prospects and QoS requirements in 5G networks / V.O. Tikhvinskiy, G. Bochechka // Journal of Telecommunications and Information Technology. — 2015. — P. 23–26.
120. Umer M.A Cellular network resource distribution methods for the joint servicing of real-time multiservice traffic and grouped IoT traffic / M.A. Umer, S.N. Stepanov, J. Ndayikunda, M.G. Kanishcheva // T-Comm: Телекоммуникации и Транспорт. — 2020. — Т. 14. — No. 10. — P. 61–69.
121. Umer M.A. Investigation of Cellular Network Resource Division Procedures for the Joint Servicing of Real-Time Multiservice Traffic and Elastic IoT Traffic / M.A. Umer, S.N. Stepanov // Institute of Radio and Information Systems (IRIS): Synchroinfo Journal. — 2020. — V. 6. — No. 1. — P. 7–10.
122. Volkov A. O. Development of Model and Algorithms for Servicing Real-Time and Data Traffic in a Cloud Computing System / A.O. Volkov, A.V. Korobkina, S.N. Stepanov // Proc. of 2022 Systems of Signals Generating and Processing in the Field of on Board Communications. Moscow. Russia. — 2022. — P. 1–6.
123. Vikhrova O. Enhanced Radio Access Procedure in Sliced 5G Networks / O. Vikhrova, C. Suraci, A. Tropeano, [et. all] // International Congress on Ultra Modern Telecommunications and Control Systems and Workshops. — 2019. — P. 1–6.
124. Verizon and AT&T want to virtualize the 5G Network Core and use Mobile Edge Computing [Electronic Resource] / Techblog// Access mode: <https://techblog.comsoc.org/>
125. Wang Y. Adaptive Power and Rate Control for Real-Time Status Updating over Fading Channels / Y. Wang, W. Chen // IEEE Transactions on Wireless Communications. — 2021. — V. 20. — No. 5. — P. 3095–3106.
126. You X. Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts /X. You, Ch. Wang, J. Huang, [et. all] // Science China Information Sciences. — 2021. — V. 64. — No. 1. — 74 p.
127. Yarkina N. An analytical model for 5G network resource sharing with flexible SLA-oriented slice isolation / N. Yarkina, Y. Gaidamaka, L.M. Correia, K. Samouylov // Mathematics. — 2020. — V. 8. — No. 7. — P. 1177.
128. Yu Y. Uplink Resource Allocation for Narrowband Internet of Things (NB-IoT) Cellular Networks / Y. Yu, J. Wang // Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. — 2018. — P. 446–471.

Приложение

1. Акт об использовании результатов диссертации в ООО «ЕПСКОМ»
2. Акт об использовании результатов диссертации в ЗАО «Испытательный центр МирТелеТест»
3. Акт об использовании результатов диссертации в учебном процессе МТУСИ

Приложение 1



Общество с ограниченной ответственностью «ЕПСКом»
111024, г. Москва, ул. Авиамоторная, д. 8
ОГРН 1057747111950
ИНН 772547148



УТВЕРЖДАЮ
Генеральный директор ООО «ЕПСКом»
Жвания А.Г.

МП

07.04.25

АКТ

об использовании результатов диссертационной работы Канищевой М.Г. «Разработка модели и процедур эффективного распределения ресурса мультисервисных узлов доступа при обслуживании гетерогенного трафика современных коммуникационных приложений», в ООО «ЕПСКОМ»

Комиссия в составе:

председателя комиссии, Генерального директора Общества с ограниченной ответственностью "ЕПСКом" Жвания А.Г.

и членов комиссии, заместителя Генерального директора, ведущего инженера Митронова И.Ю. и ведущего инженера-разработчика Кузнецова В.Л.

составили настоящий акт в том, что:

1. При анализе влияния импульсного характера формирования входных потоков заявок на передачу и обслуживание эластичных данных в мультисервисном беспроводном узле доступа были использованы следующие результаты и методики, полученные в диссертационной работе Канищевой М.Г. «Разработка модели и процедур эффективного распределения ресурса мультисервисных узлов доступа при обслуживании гетерогенного трафика современных коммуникационных приложений», представленной на соискание ученой степени кандидата технических наук:
 - математическая модель формирования и обслуживания в мультисервисном беспроводном узле доступа гетерогенного трафика современных коммуникационных приложений, учитывающая возможность поступления файлов группами случайного размера;
 - методика оценки показателей обслуживания данных, основанная на составлении и решении системы уравнений равновесия при заданных значениях входных параметров.
2. Реализация разработанных Канищевой М.Г. методов и алгоритмов позволила оценить необходимую пропускную способность узла доступа при обслуживании мультимедийных сообщений и уточнить методику оценки объемов и качества передаваемых данных.

Председатель комиссии

Жвания А.Г.

Члены комиссии

Митронова И.Ю.

Кузнецова В.Л.

Приложение 2

**Закрытое акционерное общество
«Испытательный центр МирТелеТест»
(ЗАО ИЦ МТТ)**

JSC «MirTeleTest» (MTT)

Юридический адрес: 117334, Россия, г. Москва, Ленинский проспект, д. 32А, корп. Г
ИНН: 7736193957
КПП: 773601001
ОГРН: 1027700120722

Тел: +7 (495) 957-70-25/36
e-mail: info-mtt@mirtelest.ru
www.mirtelest.ru

**Испытательный центр Закрытого акционерного общества
«Испытательный центр МирТелеТест» (ЗАО ИЦ МТТ)**

Адрес места осуществления деятельности: 111024, Россия, г. Москва, Авиамоторная ул., д. 8, стр.1, этаж 3, комн. 4, 5, 12
Уникальный номер записи об аккредитации в реестре аккредитованных лиц RA.RU.21AM76

Утверждаю
Генеральный директор ЗАО ИЦ МТТ
Ежков Д.А.
«27» марта 2025 года


МП
МирТелеТест

Акт
о внедрении результатов диссертационной работы Канищевой М.Г.
на соискание учёной степени кандидата технических наук
по научной специальности 2.2.15 – Системы, сети и устройства телекоммуникаций на тему:
«Разработка модели и процедур эффективного распределения ресурса в мультисервисных узлах доступа
при обслуживании гетерогенного трафика современных коммуникационных приложений»

Комиссия в составе:
Генерального директора Ежкова Д.А.
Руководителя Испытательного центра Силкина А.А.
Инженера-испытателя Казарновского А.В.

составила настоящий Акт о том, что научные результаты, полученные Канищевой Маргаритой Геннадьевной в диссертационной работе на соискание учёной степени кандидата технических наук, а именно: «Метод и программа расчёта (регистрационный № 2023610805 от 13.01.2023) пропускной способности мультисервисного узла доступа», используются при проведении сертификационных испытаний инфокоммуникационного оборудования на соответствие «Правил применения оборудования транзитных, оконечно-транзитных и оконечных узлов связи. Часть XV. Правила применения комбинированных телефонных станций, использующих технологии мультисервисных сетей», утв. Приказом Минкомсвязи России от 21.03.2017 № 129, при определении параметров поступающей нагрузки для создания имитации сетевого окружения для испытуемого образца системы коммутации.

Члены комиссии:

Генеральный директор ЗАО ИЦ МТТ, к.т.н., доцент Ежков Д.А.
Руководитель Испытательного центра ЗАО ИЦ МТТ Силкин А.А.
Инженер-испытатель ЗАО ИЦ МТТ, к.т.н., доцент Казарновский А.В.

Приложение 3

МИНИСТЕРСТВО ЦИФРОВОГО
РАЗВИТИЯ, СВЯЗИ И МАССОВЫХ
КОММУНИКАЦИЙ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Ордена Трудового Красного Знамени
федеральное государственное
бюджетное образовательное
учреждение высшего образования

«МОСКОВСКИЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ СВЯЗИ И
ИНФОРМАТИКИ»
(МТУСИ)



ул. Авиамоторная, д. 8а, Москва, 111024,
www.mtuci.ru; mtuci.pf; e-mail: kanc@mtuci.ru
Телефон (495) 957-77-31; факс (495) 957-77-36

ОГРН 1027700117191; ИНН/КПП 7722000820/772201001; ОКПО 01179952;
ОКВЭД 85.22, 46.19, 58.19, 61.10, 68.32, 72.19, 85.21, 85.23, 85.42.9, 71.20, 33.13, 26.60 ; ОКТМО 45388000

MINISTRY OF DIGITAL
DEVELOPMENT,
COMMUNICATIONS
AND MASS MEDIA OF
THE RUSSIAN FEDERATION

MOSCOW TECHNICAL
UNIVERSITY
OF COMMUNICATIONS
AND INFORMATICS
(MTUCI)

24.03 2025 № _____
На № _____ от _____



УТВЕРЖДАЮ
Проректор МТУСИ по научной работе
Доктор технических наук, профессор

_____ Леохин Ю.Л.

АКТ

об использовании результатов диссертационной работы

Канищевой Маргариты Геннадьевны

«Разработка модели и процедур эффективного распределения ресурса мультисервисных
узлов доступа при обслуживании гетерогенного трафика современных коммуникационных
приложений»,

представленной на соискание ученой степени кандидата технических наук, в учебном процессе
по кафедре «Сети связи и системы коммутации» МТУСИ

Комиссия в составе: председатель – зам. зав. кафедрой СС и СК, доцент Маликова Е.Е. члены
доцент Степанова И.В. и профессор Пшеничников А.П. составили настоящий акт в том, что результаты
диссертационной работы:

- обобщенная модель совместного обслуживания гетерогенного трафика в беспроводном узле доступа, в которой учитывается зависимость требуемой скорости передачи от типа запроса, наличие приоритета у трафика сервисов реального времени, возможность поступления файлов группами случайного размера, наличие или отсутствие у файлов эластичных свойств, возможность у файлов ожидания начала передачи и старения передаваемой информации;
- методики точной и приближенной оценки характеристик обслуживания заявок в мультисервисном узле доступа, основанные на решении системы линейных уравнений равновесия метода Гаусса-Зейделя и реализации принципа декомпозиции модели;
- программы расчета характеристик пропускной способности мультисервисного узла доступа использованы в учебно-исследовательской работе бакалавров и магистров кафедры СС и СК, при курсовом проектировании по дисциплине «Основы Интернета Вещей».

Председатель комиссии:

_____ Маликова Е.Е.

Члены комиссии:

_____ Степанова И.В.

_____ Пшеничников А.П.