

Федеральное государственное автономное образовательное учреждение
высшего образования «Санкт-Петербургский государственный университет
аэрокосмического приборостроения»

На правах рукописи



Янковский Никита Андреевич

**МОДЕЛИ И МЕТОДЫ ДИНАМИЧЕСКОГО РАСПРЕДЕЛЕНИЯ
РЕСУРСОВ В СЕТЯХ 5G**

Специальность 2.2.15 – Системы, сети и устройства телекоммуникаций

Диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:
Татарникова Татьяна Михайловна
доктор технических наук, профессор

Санкт-Петербург – 2026

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
1. АНАЛИЗ ТРЕБОВАНИЙ К СОВМЕСТНОМУ ИСПОЛЬЗОВАНИЮ СЦЕНАРИЕВ СЕТЕЙ 5G	11
1.1. Концептуальное представление сетей 5G.....	11
1.1.1. Общая характеристика сетей 5G.....	11
1.1.2. Основные компоненты сети 5G	17
1.1.3. Концепция сетевых сегментов в 5G	20
1.2. Описание сценариев использования сетей 5G	25
1.2.1. Сценарий расширенной мобильной широкополосной связи	25
1.2.2. Сценарий сверхнадежной связи с низкой задержкой.....	28
1.2.3. Сценарий массовых коммуникаций машинного типа.....	33
1.3. Структура фрейма сети 5G	37
1.4. Постановка задачи исследования.....	40
Выводы по первой главе.....	43
2. РАСПРЕДЕЛЕНИЕ РЕСУРСОВ В ВОСХОДЯЩЕМ КАНАЛЕ СЕТИ 5G	45
2.1. Методы случайного доступа к каналу	45
2.2. Модель трафика восходящего канала	50
2.3. Описание процесса передачи данных в восходящем канале	51
2.4. Обеспечение надежности передачи данных в межмашинном взаимодействии	54
2.4.1. Метод K -повторений.....	54
2.4.2. Алгоритм скользящего сетевого кодирования.....	56
2.4.3. Результаты сравнения методов скользящего сетевого кодирования и K -повторений.....	60

2.5. Метод динамического изменения числа преамбул	62
2.6. Модель восходящего канала связи сети 5G	67
2.6.1. Описание объектов и процессов модели восходящего канала	68
2.6.2. Результаты моделирования	72
Выводы по второй главе	74
3. МЕТОД АДАПТИВНОГО МУЛЬТИПЛЕКСИРОВАНИЯ БИТОВЫХ ПОТОКОВ В НИСХОДЯЩЕМ КАНАЛЕ СВЯЗИ	75
3.1. Анализ существующих методов мультиплексирования	75
3.2. Формулировка задачи мультиплексирования гетерогенного трафика в нисходящем канале сети 5G	78
3.3. Описание схемы адаптивного мультиплексирования потоков в нисходящем канале	80
3.4. Модель нисходящего канала	83
3.4.1. Описание объектов и процессов модели нисходящего канала	83
3.4.2. Результаты моделирования	89
Выводы по третьей главе	93
4. МЕТОД ДИНАМИЧЕСКОЙ АССОЦИАЦИИ ПОЛЬЗОВАТЕЛЕЙ К БАЗОВЫМ СТАНЦИЯМ	94
4.1. Анализ существующих решений ассоциации пользователей к базовым станциям	94
4.2. Модель ассоциации пользователей к базовым станциям	100
4.3. Асимптотическая нижняя граница средней задержки	103
4.4. Метод динамической ассоциации пользователей с базовыми станциями	107
4.5. Результаты моделирования	110
Выводы по четвертой главе	113

ЗАКЛЮЧЕНИЕ	115
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ	117
СПИСОК ЛИТЕРАТУРЫ.....	118
ПРИЛОЖЕНИЕ А	128
ПРИЛОЖЕНИЕ Б.....	129
ПРИЛОЖЕНИЕ В	131

ВВЕДЕНИЕ

Актуальность темы исследования. В Стратегии развития отрасли связи Российской Федерации на период до 2035 г., подготовленный Министерством цифрового развития России и утвержденной распоряжением правительства РФ от 23.11.2023 г. представлены 19 стратегических целей в области связи, среди которых обеспечение малочисленных населенных пунктов и труднодоступных районов услугами связи 5G на отечественном оборудовании к 2030 году и создание гибридной сети связи на основе отечественной низкоорбитальной спутниковой группировки, сегментов мобильной и фиксированной связи к 2035 году. Эта сеть, по задумке Минцифры, покроет всю территорию страны, включая федеральные и региональные автомобильные дороги, Арктическую зону и трассу Северного морского пути, а также обеспечит технологические возможности управления беспилотными авиационными системами и глобальное покрытие с возможностью экспорта услуг связи в зарубежные страны.

Сценарии использования сотовых сетей связи пятого поколения или сетей 5G: eMBB (enhanced Mobile Broadband) – сверхширокополосная мобильная связь, URLLC (Ultra-Reliable Low Latency Communication) – сверхнадежная связь с низкими задержками и mMTC (Massive Machine-Type Communications) – массовая межмашинная связь направлены на поддержку большого количества одновременно подключенных устройств, автономность, низкую задержку, эффективное распределение ресурсов, механизмов резервирования и приоритетное планирования данных. Для реализации гибридной сети связи предстоит преодолеть ряд противоречий, связанных прежде всего с гетерогенностью трафика и, соответственно, разных требований к качеству его обслуживания (Quality of Service, QoS).

QoS в сетях 5G представляет собой набор механизмов для управления и приоритизации разного типа трафика таким образом, чтобы обеспечить соответствующий уровень производительности и надежности. Каждый

сценарий имеет связанный профиль QoS, который указывает требования к задержке, надежности и пропускной способности. Например, к URLLC трафику, характерному для приложений промышленной автоматизации, здравоохранения, автономных автомобилей предъявляются строгие требования QoS: надежность передачи – не менее 99,999% с задержкой не более 1 мс. В то же время для mMTC трафика, генерируемого такими устройствами как счетчики воды, газа, электричества, контроллеры уличного освещения, датчики парковочного места, различные сенсоры дыма/огня, «умные» мусорные баки и прочих устройств интернета вещей не важны высокая скорость и сверхнизкие задержки, но очень важны автономность и огромное число подключений в сети. Для eMBB трафика, например, потокового видео или игр виртуальной/дополненной реальности важна массовость подключений при малом объеме данных.

Для разрешения сложившихся противоречий необходимы новые модели и методы, обеспечивающие обслуживание гетерогенного трафика в сетях 5G согласно требованиям QoS.

Степень разработанности темы диссертации. Научные работы в области моделирования гетерогенных телекоммуникационных сетей, в том числе сотовых сетей связи, посвящены особенностям применения теории телетрафика, случайных процессов, теории очередей, имитационного моделирования при оценивании показателей QoS, распределении ресурсов сети, организации множественного доступа к разделяемому ресурсу.

Определяющий вклад в развитие математических моделей гетерогенных телекоммуникационных сетей внесли российские и зарубежные ученые, такие как Г.П. Башарин, В.М. Вишневский, А.Н. Дудин, В.Г. Карташевский, О.И. Кутузов, Б.С. Лившиц, Е.В. Морозов, А.Н. Назаров, Ю.Н. Орлов, А.П. Пшеничников, С.Н. Степанов, Н.А. Соколов, Т.М. Татарникова, Я.В. Фидлин, О.И. Шелухин, М.А. Шнепс, А.Д. Харкевич, Г.Г. Яновский, V.-B. Iversen, M. Gerla, F. Kelly, P.J. Kuhn, L. Юетгоск, K.W. Ross, W. Whitt и др. Предложенные в диссертации модели и методы распределения ресурсов сетей 5G в условиях

гетерогенного трафика и высокой плотности устройств основаны на результатах, полученных этими учеными.

Исследованию методов доступа в телекоммуникационных сетях посвящены работы А.С. Аджемова, Б.С. Гольдштейна, Е.А. Крука, А.Е. Кучерявого, К.Е. Самуйлова, В.О. Тихвинского, А.М. Тюрликова, J. Andrews, N. Himayat, F. Baccelli, S. Rangan, T. Rappoport, S. Singh, и др. Предложенный в диссертации метод распределения ресурсов в восходящем канале уточняет результаты вышеперечисленных ученых применительно к специфике обслуживания данных малого объема, динамического изменения числа используемых преамбул и приоритизации гетерогенного трафика.

Работы таких ученых как А.В. Росляков, А.И. Парамонов, Р.В. Киричек, M. Dohler, H. Yanikomeroglu, J. Nosek и др. посвящены одной из актуальных проблем в последние годы – анализу эффективности функционирования интернета вещей, являющегося источником mMTC трафика. Предложенные в работе модели сосуществования URLLC и eMBB трафика в нисходящем канале развиваюют идеи этих ученых при оценивании качества обслуживания устройств сетей 5G.

Объект и предмет исследования. Объектом исследования являются сети 5G.

Предметом исследования – применение моделей и методов распределения ресурсов в сетях 5G.

Цель и задачи исследования. Цель работы состоит в повышении эффективности распределения ресурсов в сотовых сетях 5G для удовлетворения показателям QoS.

Цель работы достигается последовательным решением следующих задач:

1. Разработка математических моделей восходящего канала и нисходящего канала сетей 5G.
2. Разработка метода доступа устройств M2M к разделяемому ресурсу в восходящем канале сетей 5G.

3. Разработка метода мультиплексирования битовых потоков URLLC и eMBB трафика в нисходящем канале сетей 5G

4. Разработка модели динамической ассоциации устройств с базовой станцией в сетях 5G.

Научная новизна диссертационной работы состоит в следующем:

1. Математические модели восходящего канала и нисходящего канала отличаются от известных учетом сценариев использования сетей 5G, что позволяет организовать эффективное совместное обслуживание гетерогенного трафика, удовлетворяющее QoS.

2. Метод доступа устройств M2M к ресурсам сетей 5G, отличается от известных учетом специфики обслуживания данных малого объема, динамическим изменением числа используемых преамбул, а также системой приоритизации трафика, что позволяет снизить задержку и улучшить энергоэффективность за счет сокращения повторных передач.

3. Метод мультиплексирования битовых потоков URLLC и eMBB трафика в нисходящем канале сетей 5G отличается от известных возможностью адаптивного изменения схемы распределения ресурсов между различными видами трафика, что позволяет для некоторых вариантов снизить среднюю задержку.

4. Математическая модель динамической ассоциации устройств с базовой станцией, отличается от известных учетом особенностей трафика устройств M2M, расположенных на границе соты, позволяющие снизить среднюю задержку передачи сообщений по нисходящему каналу связи.

Теоретическая значимость результатов диссертации заключается в предложении нового комплекса взаимосвязанных моделей и методов управления трафиком, учитывающих особенности современных сотовых сетей связи. Предложенные в работе модели и методы могут быть использованы при теоретическом анализе сотовых сетей пятого поколения.

Практическая значимость диссертационной работы заключается во внедрении моделей и методов эффективного управления гетерогенным трафиком сетей 5G.

Положения, выносимые на защиту.

1. Метод динамического изменения числа преамбул на основе комбинирования методов машинного обучения: регрессии опорных векторов SVR и рекуррентных нейронных сетей RNN позволяет снизить среднюю задержку передачи сообщений по каналу со случайным доступом.
2. Модель нисходящего канала и метод адаптивного мультиплексирования потоков eMBB и URLLC на основе комбинированного использования LDPC-кодов и решетчатой модуляции (TCM), позволяют обеспечить совместную передачу гетерогенного трафика в общем частотно-временном ресурсе с выполнением строгих требований URLLC по задержке (<1 мс) и надежности ($>99.999\%$) при минимальной деградации пропускной способности eMBB-трафика.
3. Модель и метод динамической ассоциации пользовательских устройств с базовыми станциями в гетерогенной сети 5G, основанные на прогнозировании нагрузки с помощью рекуррентных нейронных сетей (LSTM) позволяют снизить среднюю задержку передачи данных внутри соты по сравнению со статической ассоциацией.

Методы исследования, которые использовались в диссертационной работе: теория информации, теория связи, теория случайных процессов, теория вероятностей, теория Марковских цепей, методы математической оптимизации и методы имитационного моделирования.

Степень достоверности и апробация результатов. Результаты диссертационной работы получены на основе общепринятых математических методов, согласуются с ранее известными результатами и подтверждены результатами имитационного моделирования, публикациями в высокорейтинговых журналах и выступлениями на российских и международных конференциях.

Основные положения диссертационной работы были представлены и обсуждались на следующих конгрессах, конференциях и семинарах: международная научная конференция «Обработка, передача и защита информации в компьютерных системах», 11–15 апреля 2022; международная конференция «Волновая электроника и инфокоммуникационные системы», 2020, 2021, 2023 гг.; международная научно-практическая конференция «Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем», Москва, 2023, 2024 гг.; Санкт-Петербургская международная конференция «Региональная информатика и информационная безопасность», Санкт-Петербург, 25–27 октября 2023 год.

Соответствие работы паспорту научной специальности.

Диссертационная работа соответствует паспорту научной специальности 2.2.15 «Системы, сети и устройства телекоммуникаций» (пунктам 2, 6, 8 и 18).

Личный вклад автора. Все результаты, представленные в тексте диссертационной работы, получены автором лично.

Публикации. Материалы, отражающие основные результаты диссертации, опубликованы в 17 научных работах: 6 работ – в журналах из перечня ВАК, в том числе 1 из них без соавторов; 5 – в изданиях, индексируемых Scopus, и 5 – в сборниках конференций, индексируемых РИНЦ. По теме диссертации получено 1 свидетельство о государственной регистрации программы для ЭВМ.

Объем и структура работы.

Диссертация состоит из введения, четырех глав, заключения и списка литературы. Общий объем работы – 131 страниц сквозной нумерации, в том числе основного текста – 116 машинописных страниц, 47 рисунков и графиков, 8 таблиц.

1. АНАЛИЗ ТРЕБОВАНИЙ К СОВМЕСТНОМУ ИСПОЛЬЗОВАНИЮ СЦЕНАРИЕВ СЕТЕЙ 5G

Методы планирования ресурсов, используемые в сотовых сетях предыдущих поколений неэффективны для 5G из-за высокой плотности устройств и разнородности трафика. Например, процедура случайного доступа в 4G, рассчитанная на десятки устройств, становится узким местом в 5G, где одновременно активируются тысячи датчиков. Это приводит к коллизиям преамбул, увеличивающим задержку и снижающим надежность. Кроме того, передача разнородного трафика в одном частотном спектре требует новых подходов к мультиплексированию во избежание деградации качества обслуживания.

В данной главе проведен системный анализ архитектурных и функциональных особенностей сетей 5G с выявлением актуальных задач распределения ресурсов.

1.1. Концептуальное представление сетей 5G

Концептуальное представление сетей 5G приведем на уровне общих характеристик в сравнении сетей пятого поколения с предыдущими поколениями сотовых сетей и рассмотрения архитектуры сетей 5G в соответствии со стандартами консорциума 3GPP (The 3rd Generation Partnership Project) – партнерского проекта третьего поколения в области телекоммуникаций [35].

1.1.1. Общая характеристика сетей 5G

Сети сотовой связи предыдущих поколений имели следующий функционал:

- 1G: Услуги передачи речи по аналоговой сети;

- 2G: Услуги передачи речи по цифровой сети, низкоскоростные услуги передачи данных;
- 3G: Высокоскоростные услуги передачи данных, с возможностью передачи голоса по сети IP, мобильный доступ к интернету MBB (Mobile Broadband).
- 4G: Мобильный широкополосный доступ MBB на базе LTE, LTE-A, передача голоса (VoLTE).

Сети 5G существенно расширяют функционал сетей 4G благодаря появлению следующих особенностей [3]:

- Усовершенствованный мобильный широкополосный доступ eMBB;
- Сверхнадежные коммуникации с низкой задержкой ULLRC;
- Массивные межмашинные коммуникации Massive IoT/ПоТ, мМТС.

На основе этих трех сценариев строится все многообразие услуг и возможностей сетей 5G, наиболее характерными из которых являются:

- высокая скорость передачи данных через различные технологии радиодоступа, Гигабайты в секунду;
- решения «Умный дом» (Smart Home) и «Умное здание» (Smart Building): видеонаблюдение, управление и автоматизация бытовой техники, управление системами безопасности, хранилища контента, климатика и пр.;
- решение «Умный город», основными сервисами которого могут быть: безопасный город, электронное правительство, электронное здравоохранение, электронное образование, электронный банкинг, электронный сбор показаний ЖКХ Smart Meters, умные электросети Smart Grid, и пр.;
- новые видеоуслуги 4К/8К: объемное видео, экран сверхвысокой четкости, возможность эффекта присутствия;
- работа в облаке: хранение данных в облачном хранилище, работа прикладных программ непосредственно из облака на любом устройстве и из любого местоположения;
- дополненная и виртуальная реальность (AR/VR): развлечения, игры,

телеприсутствие;

- промышленная автоматизация, построенная на технологиях интернета вещей IoT и искусственного интеллекта;
- бизнес-критичные приложения: электронная медицина, связь при чрезвычайных ситуациях, тактильный интернет и другие.
- беспилотный транспорт: беспилотные автомобили, беспилотные тракторы, беспилотные поезда для метро и пригородных железных дорог, дроны и другие виды общественного и специального транспорта, системы помощи водителю.

Спектр услуг платформы 5G носят синергетический и масштабируемый характер. По сути, 5G становится платформой для разработки новых услуг и приложений – любая отдельная услуга подключения может быть реализована как подходящая комбинация eMBB, mMTC и URLLC. Например, в продвинутом сельскохозяйственном сценарии дистанционно управляемая машина должна поддерживать надежное выполнение команд в реальном времени (URLLC), время от времени отправляя видеопоток (eMBB), а также собирая данные с различных датчиков и устройств интернета вещей в сельскохозяйственной среде (mMTC). Любая услуга реализуется через широкополосный доступ к ресурсу.

Некоторые унаследованные инфраструктуры сетей 2G, 3G и 4G также вошли состав сетей 5G, поскольку продолжают эксплуатироваться. На рисунке 1.1 приведен общий вид возможностей сети 5G как платформы разработки.

Качество предоставляемых услуг 5G регламентируется нормативными документами. Международный союз электросвязи МСЭ-R определил основные показатели качества сетей 5G:

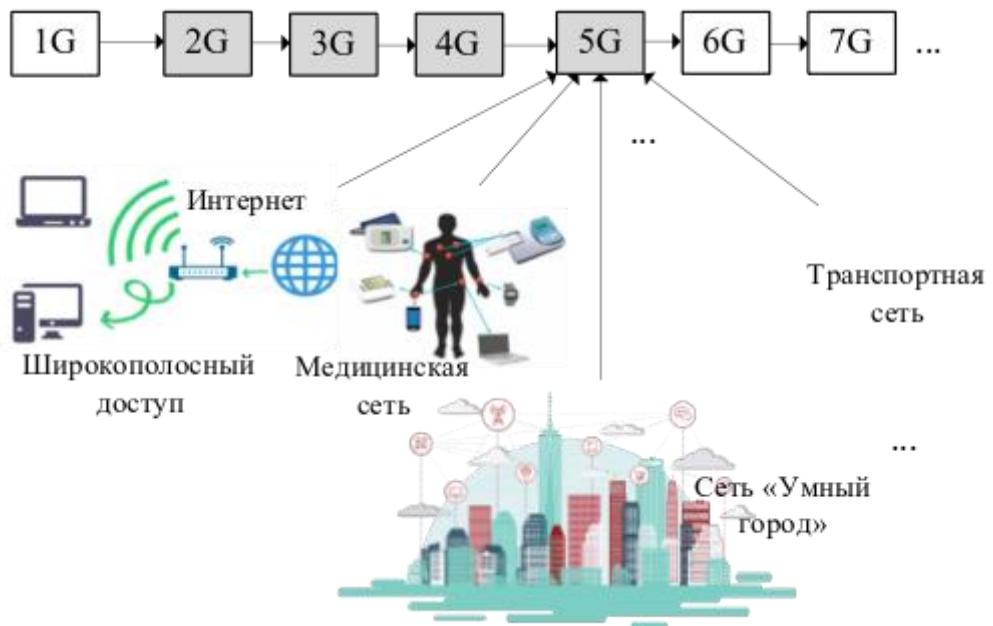


Рисунок 1.1 – Сети 5G, как платформа для новых услуг и приложений

1) Скорость передачи данных – количество бит, передаваемых за единицу времени, [бит/с]. Соответственно, различают два вида скорости: в восходящем канале и в нисходящем канале. Пиковая скорость передачи данных в сетях 5G может достигать 20 Гбит/с. Для сравнения в 4G скорость передачи данных до 300 Мбит/с. На рисунке 1.2 приведена диаграмма роста скорости передачи данных в поколениях сотовых сетей [1; 3].

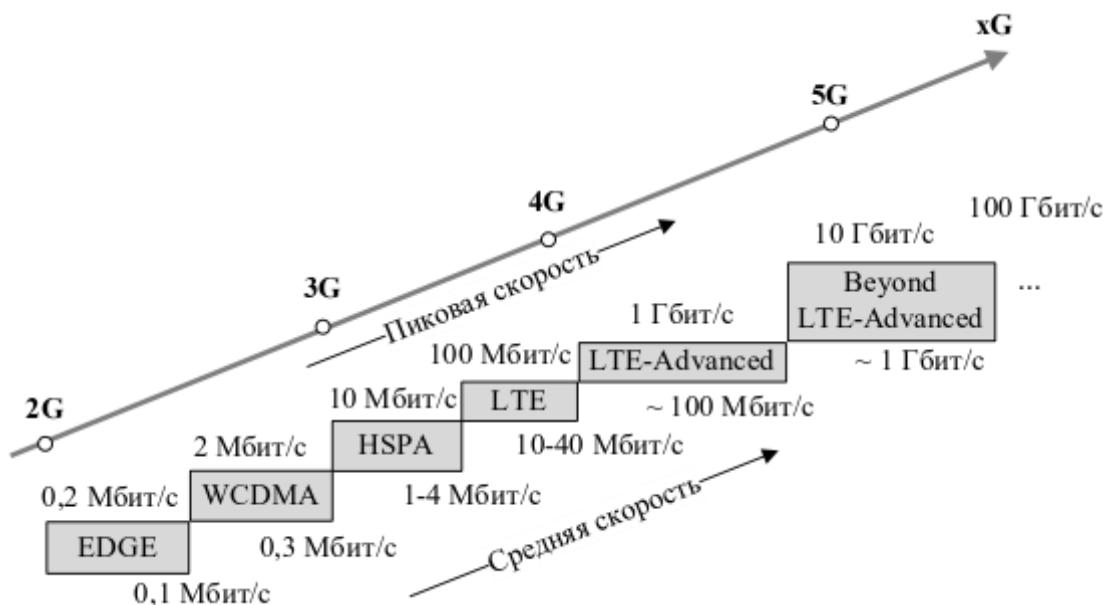


Рисунок 1.2 – Диаграмма роста скорости передачи данных в поколениях сотовых сетей

EDGE (Enhanced Data rates for GSM Evolution) – цифровая технология беспроводной передачи данных для мобильной связи, которая функционирует как надстройка над 2G.

WCDMA (Wideband Code Division Multiple Access – широкополосный множественный доступ с кодовым разделением) – технология радиоинтерфейса, использующая широкополосный множественный доступ с кодовым разделением каналов.

HSPA (High Speed Packet Access – высокоскоростная пакетная передача данных) – технология беспроводной широкополосной радиосвязи, использующая пакетную передачу данных и являющаяся надстройкой к мобильным сетям WCDMA.

LTE (Long-Term Evolution – долговременное развитие, часто обозначается как 4G LTE) – стандарт беспроводной высокоскоростной передачи данных для мобильных телефонов.

LTE-Advanced – стандарт мобильной связи, улучшающий стандарт LTE, в том числе по скорости передачи данных.

Beyond LTE-Advanced – стандарт мобильной связи, который наложен на стандарт LTE-Advanced.

2) Задержка – время, требуемое пакету данных для прохождения от источника к месту назначения, [с]. Минимальная задержка для сервисов URLLC [4; 6] – 0,5 мс, для сервисов eMBB – 4 мс. В среднем задержка при передаче данных в сетях 5G – не более 2 мс.

3) К показателям надежности передачи данных в сетях 5G относят вероятность потери данных из-за ошибок в процессе передачи. Коэффициент потери пакетов выражается в процентах и рассчитывается как количество потерянных пакетов по сравнению с общим количеством отправленных.

На рисунке 1.3 показаны основные параметры сети 5G в сравнении с показателями сети 4G: пиковая скорость 5G в 20 раз выше по сравнению с 4G, скорость на пользователя (средняя) может достигать 100 Мбит/с и более,

эффективность использования спектра – количество информации, которую можно передать на единицу частотного диапазона в сети 5G в 3 раза выше, чем в 4G, мобильность пользователя - скорость, с которой может перемещаться пользователь с терминалом 5G по площади покрытия сети без потери хендовера между базовыми станциями достигает 500 км/час, задержка в сети 5G снижается до 1 мс и менее, в то время как в сети 4G можно достичь минимум 10-миллисекундной задержки, плотность терминалов в сети 5G повышается на порядок, энергоэффективность сети 5G на порядок лучше, чем в сети предыдущего поколения, емкость трафика на единицу площади – скорость передачи данных на квадратный метр площади покрытия сети в 5G на два порядка выше, чем в сети 4G.

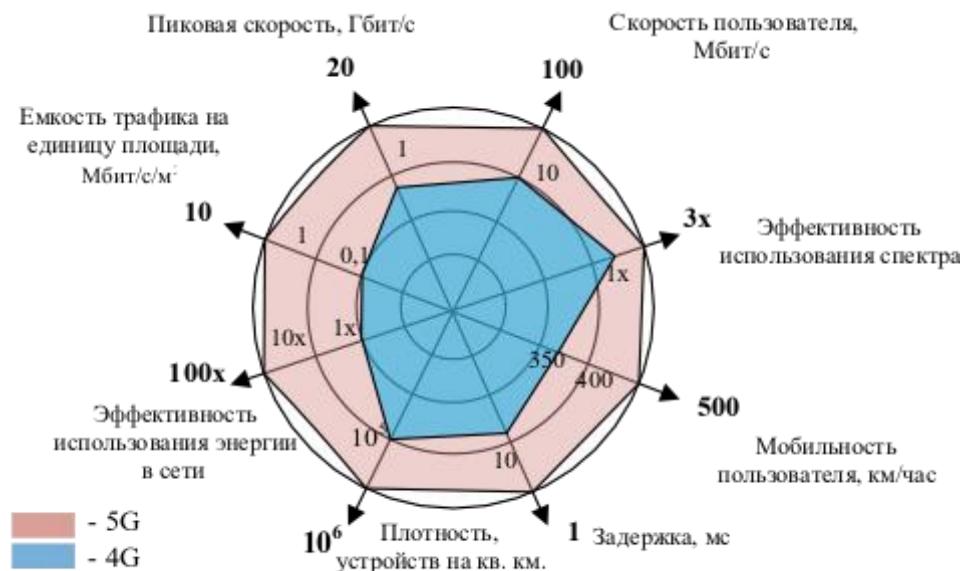


Рисунок 1.3 – Основные параметры сравнения 5G и 4G

На рисунке 1.4 показаны соотношения по степени важности для сценариев использования 5G.



Рисунок 1.4 – Соотношения по степени важности параметров сети 5G для сценариев использования

1.1.2. Основные компоненты сети 5G

Концептуальное представление сетей 5G на уровне основных компонентов и их взаимодействия приведено на рисунке 1.5.

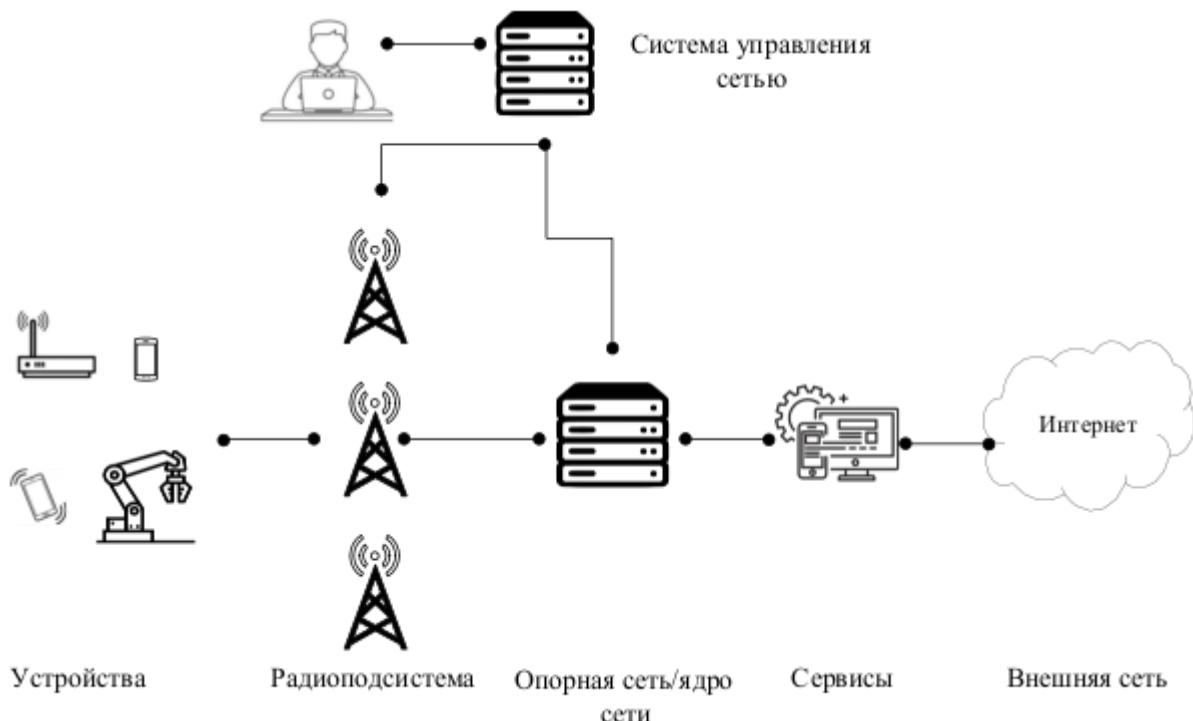


Рисунок 1.5 – Концептуальное представление компонентов сети 5G

Сотовая сеть состоит из двух подсистем: радиоподсистемы и опорной сети. Радиоподсистема включает базовые станции (generalized Node B, gNB) и контроллеры 2G базовых станций и контроллеры 3G – радиосети. Главной функцией радиоподсистемы является управление радиоресурсами на базовых станциях.

Опорная сеть состоит из нескольких систем, которые обеспечивают обработку трафика, тарификацию, авторизацию в сети и управление политиками обслуживания абонентов.

Все эти системы сегодня в основном реализованы в виде программного обеспечения, которое работает на серверах общего пользования.

Для мониторинга производительности и сбоев, настройки сети и обновления ее программного обеспечения используются системы управления сетью.

Пользовательское оборудование (User Equipment, UE), такое как смартфон, датчик или модем 5G, инициирует запрос на обслуживание.

Таким образом, сотовая сеть обеспечивает доступ от устройства пользователя до выхода во внешние сети (Интернет) для обмена пакетами данных PDU (Packet Data Unit). Для передачи трафика конкретной услуги или приложения устанавливается PDU-сессия – ассоциация между пользовательским устройством и шлюзом UPF. Она определяет требования QoS, использование слайсинга, информацию о дальнейшей маршрутизации пользовательских данных.

Сеть 5G включает в себя следующие основные сетевые функции [1;3;5]:

- функция управления доступом и мобильностью (AMF – Access and Mobility Management Function);
- функция управления сессиями (SMF – Session Management Function);
- функция плоскости пользователя (UPF – User Plane Function).

Взаимодействие между сетевыми функциями представлено двумя способами:

- сервис-ориентированное, когда одни сетевые функции (например, AMF) позволяют другим авторизованным сетевым функциям получать доступ

к их сервисам.

– интерфейсное, которое показывает какое взаимодействие существует между сервисами сетевых функций, описанных как взаимодействие точка-точка между любыми двумя сетевыми функциями (например, AMF и SMF).

На плоскости управления 5G сетевые функции должны использовать только сервис-ориентированные интерфейсы для их взаимодействия.

Через сеть 5G пользовательский трафик проходит следующий путь (рисунок 1.6):

- между UE и gNB – Data Radio Bearer (DRB) – логический канал связи, используемый в сети LTE для предоставления высокоскоростных услуг передачи данных устройствам;
- между gNB и UPF – GPRS Tunneling Protocol – User Plane tunnel (GTP-U tunnel) – протокол туннелирования, обеспечивающий передачу пользовательского трафика между элементами сотовой сети. Пользовательский трафик попадает в туннель на базовой станции, а выходит из туннеля на элементе UPF, например, маршрутизатор или шлюз;
- между UPF и внешней сетью – при выходе из туннеля, маршрутизатор снимает GTP-обертку и трафик передается дальше во внешнюю сеть.



Рисунок 1.6 – Путь прохождения пользовательского трафика в 5G

Использование туннеля между gNB и UPF для передачи пользовательского трафика внутри сотовой сети реализует, во-первых, бесшовный переход при перемещении пользовательского устройства между

базовыми станциями, и, во-вторых, с точки зрения опорной сети весь трафик пользователя – это один или несколько туннелей. Соответственно, внутри сотовой сети все политики управления качеством обслуживания QoS, приоритезации, маршрутизации и т. п. применяется к туннелю, а не к каждому соединению каждого приложения в отдельности.

Модель QoS сети 5G основана на потоках QoS. Каждая PDU-сессия может содержать несколько потоков QoS, каждый из которых идентифицируется идентификатором потока QoS (QFI) и связан с набором параметров QoS, определенных в индикаторе QoS 5G (5QI). Такая детализация важна для разнородного трафика, поскольку один пользовательский сеанс может одновременно передавать трафик с различными требованиями, например, поток виртуальной реальности (высокий приоритет, низкая задержка) наряду с фоновой синхронизацией файлов (более низкий приоритет).

1.1.3. Концепция сетевых сегментов в 5G

Логическое представление сетей 5G включает две функциональные плоскости:

1. *Плоскость пользователя* (User Plane, UP), выполняющей функции распределенной пересылки и обработки данных, обеспечения возможности совместной работы на нескольких площадках, многосоставных механизмов и технологий мультистандартного слияния для создания гибкой топологии сети доступа. Можно сказать, что UP отвечает за передачу данных и пакетов между устройствами и внешними сетями.

2. *Плоскость управления* (Control Plane, CP), выполняющей реконфигурируемую функцию централизованного управления сетью, обеспечения доступа по требованию, мобильность и управление сессиями, поддержания управления и контроля ресурсов. Можно сказать, что CP отвечает за установление и управление соединениями.

Разделение на функциональные плоскости позволяет оптимизировать работу различных сценариев использования сетей 5G.

Физическое представление сети 5G соответствует следующим архитектурным особенностям сетей 5G [2;5;6]:

- разделение сетевых узлов на элементы, обеспечивающие работу протоколов «плоскости пользователя» и элементы, обеспечивающие работу протоколов «плоскости управления»;
- разделение сетевых элементов на сетевые слои, основываясь на услугах, предоставляемых конкретным группам конечных пользователей;
- реализация сетевых элементов в виде виртуальных сетевых функций;
- поддержка одновременного доступа к централизованным и локальным службам для реализации облачных и граничных вычислений;
- определение конвергентной архитектуры, объединяющей различные типы сетей доступа: 3GPP (New Radio, NR) и не 3GPP (WiFi и пр.) с единой опорной сетью/ядром сети 5G.

Разделение сети 5G на элементы и слои – это сетевая архитектура, которая позволяет мультиплексировать виртуализированные и независимые логические сети в рамках одной физической сетевой инфраструктуры. Каждый сегмент сети представляет собой изолированную сквозную сеть, адаптированную под различные требования конкретного приложения с самыми разными требованиями к уровню обслуживания.

С точки зрения бизнес-модели каждый сегмент сети администрируется оператором мобильной виртуальной сети. Поставщик инфраструктуры – владелец телекоммуникационной инфраструктуры сдает свои физические ресурсы в аренду операторам мобильной виртуальной сети, которые используют базовую физическую сеть.

В зависимости от доступности выделенных ресурсов оператор мобильной виртуальной сети может самостоятельно развертывать несколько сегментов сети, адаптированных под различные приложения, предоставляемые его пользователям.

Общая архитектура, которая объединяет элементы каждого решения в единую структуру приведена на рисунке 1.7.

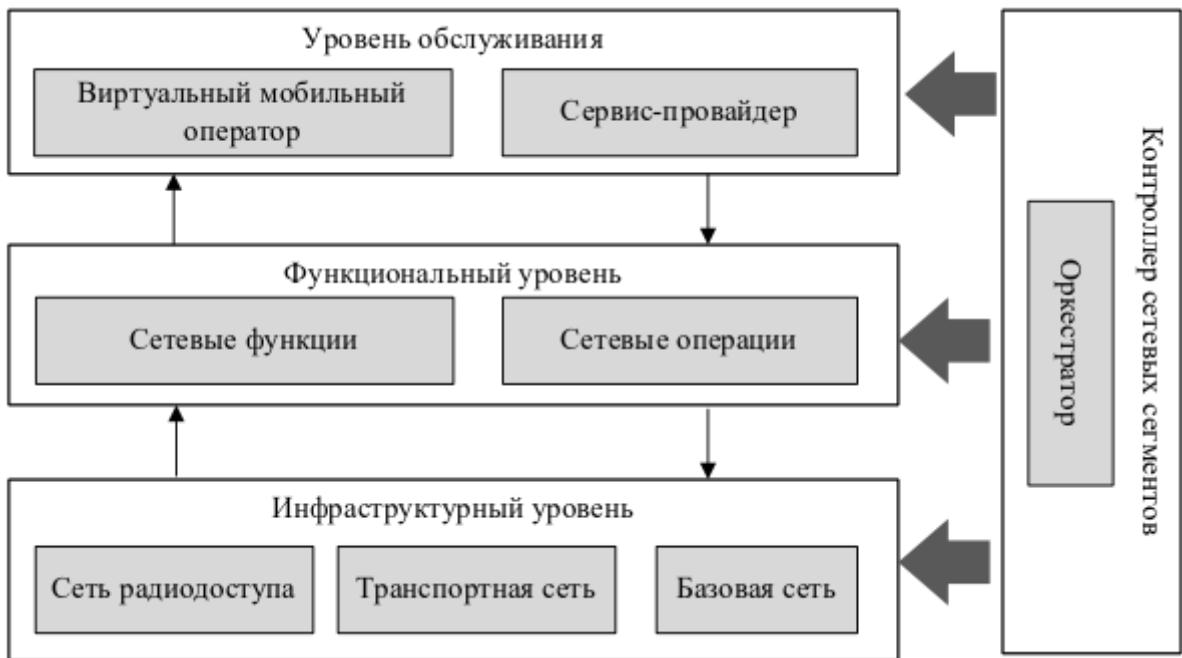


Рисунок 1.7 – Архитектура сетевых сегментов сети 5G

Архитектура сетевых сегментов состоит из двух основных блоков: один предназначен для фактической реализации сегмента, а другой – для управления и настройки сегмента.

Первый блок представляет собой трехуровневую архитектуру: уровень обслуживания, уровень сетевых функций, уровень инфраструктуры, каждый из которых, выполняя отдельные задачи, вносит свой вклад в определение и развертывание сегмента.

Второй блок представляет собой централизованный контроллер сетевых сегментов, который отслеживает и управляет функциями между тремя уровнями, чтобы эффективно координировать существование нескольких сегментов.

Уровень обслуживания напрямую взаимодействует с сетевыми бизнес-объектами, например, оператором мобильной виртуальной сети и сторонними поставщиками услуг, которые совместно используют базовую физическую

сеть, что обеспечивает единое видение требований к обслуживанию. Каждая служба формально представлена как экземпляр службы, который объединяет все сетевые характеристики в виде требований SLA (Service Level Agreement – соглашение об уровне обслуживания в виде договора об уровне предоставляемого сервиса между поставщиком услуг и заказчиком), которые, как ожидается, будут полностью удовлетворены при создании подходящего среза.

Функциональный уровень сети отвечает за создание каждого сетевого сегмента в соответствии с запросами экземпляров сервисов, поступающими с верхнего уровня. Он состоит из набора сетевых функций, которые реализуют четко определенные модели поведения и интерфейсы. Несколько сетевых функций размещаются поверх инфраструктуры виртуальной сети и объединяются в цепочку для создания сквозного сетевого сегмента, отражающего характеристики сети, запрошенные сервисом. Настройка сетевых функций выполняется с помощью набора сетевых операций, которые позволяют управлять их полным жизненным циклом – от размещения при создании сегмента до освобождения ресурсов, когда функция больше не нужна.

Чтобы повысить эффективность использования ресурсов, одна и та же сетевая функция может одновременно использоваться несколькими сегментами, что приводит к усложнению управления операциями. И наоборот, однозначное соответствие между каждой сетевой функцией и каждым сегментом упрощает процедуру настройки, но может привести к неэффективному использованию ресурсов.

Инфраструктурный уровень представляет собой фактическую физическую топологию сети: сеть радиодоступа, транспортная сеть и базовая сеть, на которой мультиплексируются все сегменты сети. Он предоставляет ресурсы физической сети для размещения нескольких сетевых функций, составляющих каждый сегмент.

Пул доступных ресурсов включает в себя разнородный набор компонентов инфраструктуры, таких как центры обработки данных – ресурсы

для хранения и вычислительных мощностей, устройства, обеспечивающие подключение к сети, например, маршрутизаторы – сетевые ресурсы и базовые станции – ресурсы для радиосвязи.

Контроллер сетевых сегментов определяется как сетевой оркестратор, который взаимодействует с различными функциями, выполняемыми на каждом уровне, для согласованного управления каждым запросом на создание сегмента. Оркестратор позволяет эффективно и гибко создавать сегменты, которые можно перенастраивать в течение их жизненного цикла. С технической точки зрения контроллер сетевых сегментов выполняет несколько задач, которые обеспечивают более эффективную координацию между вышеупомянутыми уровнями:

- Комплексное управление сервисами: сопоставление различных экземпляров сервисов с требованиями SLA и соответствующими сетевыми функциями, способными удовлетворить ограничения сервисов.
- Определение виртуальных ресурсов: виртуализация физических сетевых ресурсов для упрощения операций по управлению ресурсами, выполняемых для распределения сетевых функций.
- Управление жизненным циклом сегментов: мониторинг производительности сегментов на всех трех уровнях для динамической перенастройки каждого сегмента с учетом возможных изменений требований SLA.

Из-за сложности выполняемых задач, направленных на достижение различных целей, контроллер сетевых сегментов может состоять из нескольких оркестраторов, которые независимо друг от друга управляют подмножеством функций каждого уровня. Чтобы соответствовать требованиям к обслуживанию, различные оркестраторы должны координировать свои действия, обмениваясь информацией высокого уровня о состоянии операций, связанных с созданием и развертыванием сегментов.

Концепция сегментирования сети реализуется через требование изоляции сегментов – важного требования, позволяющее реализовать заключающуюся в

одновременном сосуществовании нескольких сегментов, использующих одну и ту же инфраструктуру. Это свойство достигается за счет того, что производительность каждого сегмента не должна влиять на производительность других сегментов.

1.2. Описание сценариев использования сетей 5G

Рассмотрим сценарии использования сети 5G с учетом ее основных компонентов и концепции сетевых сегментов.

1.2.1. Сценарий расширенной мобильной широкополосной связи

Сценарий использования расширенной мобильной широкополосной связи eMBB является основой для приложений с интенсивным использованием данных, которые используют передачу мультимедиа, совместную работу на рабочем месте и другие общественные услуги [3].

Технические требования к eMBB, изложены в концепции ITU-R IMT-2020 [1]:

- пиковые скорости передачи данных на уровне 20 Гбит/с в нисходящей линии связи и 10 Гбит/с в восходящей линии связи,
- удобную для пользователя скорость передачи данных на уровне 100 Мбит/с в нисходящей линии связи и 50 Мбит/с в восходящей линии связи,
- а также пропускную способность с плотностью трафика, равной скорость до 10 Мбит/с/м² в закрытых точках доступа.

Сценарий eMBB можно разделить на две широкие, часто пересекающиеся области: зоны широкого охвата и зоны с высокой плотностью устройств. Широкомасштабный охват обеспечивает непрерывную работу пользователей с высокой скоростью передачи данных, обычно достигающей 100 Мбит/с и более, даже в условиях высокой мобильности, таких как высокоскоростные поезда и автомобили. Для этого требуются усовершенствованные механизмы

передачи обслуживания и стабильная производительность на уровне сотовой связи. И наоборот, сценарий с высокой плотностью применяется в местах с высокой концентрацией пользователей, таких как стадионы, концертные площадки и городские центры, где основная проблема заключается в общей пропускной способности сети и плотности трафика в районе.

Сфера применения eMBB в сетях 5G охватывает следующее многообразие приложений: мобильная потоковая передача видео и аудио сверхвысокой четкости до 4K/8K и 360-градусного видео, что требует стабильных мультигигабитных соединений с минимальными перебоями в работе; иммерсивные приложения дополненной и виртуальной реальности, что требует высокой пропускной способности для рендеринга сложной графики и низкой задержки для предотвращения укачивания; беспроводные облачные офисные и вычислительные сервисы, включая совместную работу с большими массивами данных в режиме реального времени и облачные игры, что требует симметричных высокоскоростных каналов и надежного подключения.

Успешная поддержка этих приложений зависит от организации сквозной архитектуры 5G: пользовательское оборудование инициирует запрос на обслуживание, опорная сеть и радиоподсистема работают согласованно, чтобы соответствовать высоким показателям производительности eMBB, базовая станция использует передовые антенные системы, включая Massive Multiple-Input Multiple-Output (MIMO) с формированием луча, для направления пространственных потоков к пользователю, тем самым значительно повышая спектральную эффективность и пропускную способность сотовой связи.

Упрощенная структурная схема потока данных eMBB представлена на рисунке 1.8.

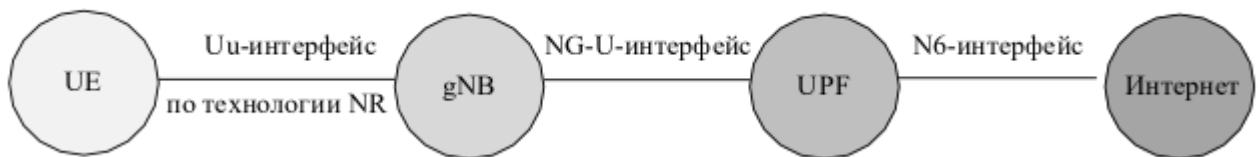


Рисунок 1.8 – Структурная схема потока данных eMBB

Пользовательский трафик передается по Uu-интерфейсу между пользовательским оборудованием и сетью радиодоступа по технологии «новое радио 5G» New Radio (NR), которая используется в сетях 5G. Uu-интерфейс обеспечивает передачу данных, управление и сигнализацию между сетью и устройствами.

Далее данные из gNB передаются через интерфейс пользовательской плоскости NG-U (Next-Generation User Plane) в функцию пользовательской плоскости UPF, которая действует как интеллектуальный узел маршрутизации и шлюз пакетов. UPF является ключевым узлом для обеспечения соблюдения политик QoS, формирования трафика и реализации функций, требующих высокой пропускной способности, таких как дублирование пакетов для обеспечения надежности.

Наконец, UPF по интерфейсу N6 в сторону внешних сетей подключается к сети передачи данных, в которой размещается фактический контент сервиса eMBB, такой как сервер потоковой передачи видео или облачная игровая платформа.

Стандартизованные значения 5QI соответствуют предварительно настроенному набору характеристик ресурсов. Для типичных служб eMBB первостепенное значение имеют следующие 5QI:

- 5QI 6/7/8/9 – классы QoS для стандартного доступа в Интернет. Например, для 5QI 8 уровень приоритета по умолчанию равен 80, время задержки пакетов (PDB) – 300 мс, а частота ошибок в пакетах (PER) - 10^{-6} , что подходит для видеоконференций и просмотра веб-страниц.

- 5QI 80/90: Для более требовательных интерактивных приложений, таких как облачные игры и виртуальная/дополненная реальность, используются 5QI 80 (приоритет 68, время ожидания 10 мс, (PER) - 10^{-6}) или 5QI 90 (приоритет 25, время ожидания 20 мс, (PER) - 10^{-4}) Малое значение PDB обеспечивает интерактивность, в то время как низкий PER гарантирует целостность визуального потока.

Ключевые показатели эффективности для eMBB напрямую зависят от этих параметров 5QI. Скорость передачи данных, с которой сталкивается пользователь, является основным показателем, измеряемым на прикладном уровне. Бюджет задержки пакетов (PDB) определяет верхнюю границу для 98-го процентиля задержки пакетов, что имеет решающее значение для интерактивных приложений. Коэффициент потери пакетов с ошибками (PELR) определяет допустимый уровень потери пакетов, не связанный с перегрузкой, что имеет решающее значение для поддержания качества несжатых видеопотоков. Кроме того, сеть должна обеспечивать высокую плотность соединений в точках доступа и надежную мобильность, чтобы поддерживать эти параметры QoS при перемещении пользователя на высоких скоростях.

1.2.2. Сценарий сверхнадежной связи с низкой задержкой

В то время как усовершенствованная мобильная широкополосная связь удовлетворяет потребность в высоких скоростях передачи данных, ориентированных на человека, сценарий сверхнадежной связи с низкой задержкой (URLLC) представляет собой сдвиг парадигмы в сторону критически важной машинной связи [2;6].

URLLC разработан не для обеспечения максимальной пропускной способности, а для гарантированной и детерминированной доставки пакетов данных в чрезвычайно жестких рамках задержки и надежности. Этот сценарий является основополагающим для приложений, где стоимость сбоя, будь то с точки зрения целостности или безопасности работы, непомерно высока. Целевыми показателями производительности, определенными 3GPP [3], являются задержка на уровне пользователя в 1 мс и надежность в 99,999% для 32-байтового пакета. Эти требования являются не просто постепенными улучшениями по сравнению с предыдущими поколениями; они требуют фундаментальной перестройки архитектуры радиоинтерфейса, стека протоколов и сетевого ядра для достижения

уровня детерминированной производительности, ранее достижимого только с помощью специализированных проводных систем.

Концептуальная основа URLLC заключается в совместной оптимизации задержки и надежности. Эти два показателя тесно взаимосвязаны; достижение сверхнизкой задержки невозможно без сверхвысокой надежности, и наоборот. Это создает уникальную инженерную задачу, поскольку традиционные методы повышения надежности, такие как повторные передачи, по своей сути увеличивают задержку. Таким образом, реализация сценария URLLC требует новых технологий, которые разрешают этот компромисс [56].

Варианты использования URLLC разделяются на две основные области: системы управления с замкнутым контуром и дискретные системы командования и контроля. Приложения с замкнутым контуром, такие как промышленная автоматизация, управление движением робототехники и автономное управление транспортными средствами предполагают непрерывные циклы обратной связи. В этих системах данные, полученные с помощью датчиков, передаются на контроллер, который определяет команду и отправляет ее на исполнительный механизм. Общее время цикла, включающего измерение, передачу данных, определение и приведение в действие команды, должно быть ограничено несколькими миллисекундами для поддержания стабильности и производительности системы. Любой сбой или недопустимая задержка в сегменте связи могут привести к катастрофическому отказу системы или снижению качества услуги.

С другой стороны, дискретные системы управления используют одноразовые или прерывистые сообщения. К этой категории относятся такие приложения, как удаленная хирургия, где тактильные ощущения хирурга должны мгновенно передаваться на роботизированный хирургический инструмент, а тактильная обратная связь должна передаваться без заметной задержки. Это также включает в себя интеллектуальную сетевую защиту, при которой сигнал обнаружения неисправности должен активировать автоматический выключатель в течение миллисекунд, чтобы предотвратить

каскадные сбои в работе сети. Общим требованием для всех вариантов использования URLLC является необходимость в детерминированной производительности, где статистическая разница в задержке (джиттер) является таким же важным показателем производительности, как и сама средняя задержка.

Архитектура 5G поддерживает URLLC благодаря усовершенствованиям в сети радиодоступа и базовой сети, а общий принцип проектирования сценария URLLC заключается в минимизации всех источников задержки и обеспечении избыточности на нескольких уровнях.

Предоставление услуг URLLC представлено следующей схемой (рисунок 1.9).

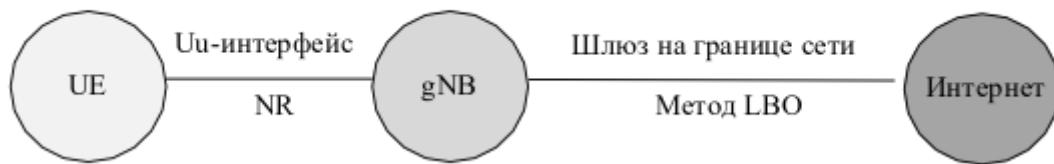


Рисунок 1.9 – Структурная схема потока данных URLLC

Основной архитектурной особенностью реализации сценария URLLC является технология Multi-Access Edge Computing (MEC) – концепция архитектуры сети, которая приближает вычислительные ресурсы к конечным пользователям (границные вычисления). В отличие от традиционных облачных вычислений, МЕС распределяет задачи обработки данных ближе к источнику данных или устройству пользователя, которая объединяет возможности облачных вычислений и экземпляры пользовательских функций UPF на границе сети, рядом с gNB.

Local Breakout (LBO) – метод маршрутизации трафика данных с UE непосредственно в интернет без прохождения через основную сеть.

Таким образом, PDU-сессия – это установление сеанса между UE с локальным шлюзом UPF. Наличие границных вычислений гарантирует, что трафик не будет проходить через всю сеть в централизованный центр обработки

данных, тем самым сокращая задержку распространения и транспортировки. Этот процесс управляет функцией SMF на основе политик функции сетевого взаимодействия.

На уровне радиоинтерфейса стек протоколов для реализации сценария URLLC претерпел значительные изменения, чтобы соответствовать гарантированной задержке в 1 мс. Главным новшеством является бесплатный (сконфигурированный) доступ по восходящей линии связи. В отличие от традиционного планирования, когда UE должен запрашивать ресурсы и ожидать предоставления – многоступенчатый процесс, который приводит к задержке, доступ без предоставления предварительно настраивает UE на периодическую передачу ресурсов. Это позволяет устройству URLLC осуществлять передачу сразу после генерации данных, что значительно сокращает задержку восходящего канала связи.

В нисходящем канале механизмы упреждающей передачи позволяют трафику URLLC прерывать текущие передачи eMBB, гарантируя, что критический пакет не будет поставлен в очередь после загрузки большого объема некритических данных. Кроме того, физический уровень использует надежные схемы модуляции и кодирования (Modulation and Coding Scheme, MCS), часто с модуляциями более низкого порядка, такими как метод цифровой модуляции (Quadrature Phase Shift Keying, QPSK), и использует частотное и пространственное разнесение благодаря многосвязности для борьбы с замираниями и помехами.

Система QoS 5G, основанная на значениях 5QI позволяет задавать приоритеты трафика URLLC. Стандартизированная таблица 5QI содержит специальные записи для служб с низкой задержкой и высокой надежностью, которые запускают специальные механизмы обработки ресурсов в gNB и UPF.

Наиболее релевантными стандартизованными значениями 5QI для URLLC являются:

- 5QI 85 – класс QoS URLLC для критически важных систем автоматизации и контроля. Он определен с уровнем приоритета 21, допустимой

задержкой пакетов в 5 мс и вероятностью ошибок пакетов, равной 10^{-5} . Относительно высокий приоритет обеспечивает преимущество перед другими типами трафика.

- 5QI 82/83: разработан для дискретной автоматизации. Определен с уровнем приоритета 19/22 соответственно, допустимая задержка пакетов, равная 10 мс и вероятность ошибок пакетов, равная 10^{-4} .
- 5QI 84: разработан для интеллектуальных транспортных систем, согласно он имеет значение вероятности ошибок пакетов, равное 10^{-5} , допустимая задержка пакетов равное 30 мс.
- 5QI 86: разработан для интеллектуальных транспортных систем, он определяет значение вероятности ошибок пакетов, равной 10^{-4} , но уменьшенное значение допустимой задержки пакетов, равной 5 мс, что, например, отражает потребности приложений V2X (Vehicle-to-Everything) – «автомобиль, подключенный ко всему» – технология беспроводной связи, которая позволяет автомобилям обмениваться данными с другими элементами транспортной системы.

Отличительной особенностью QoS для URLLC является усиленная жесткость к показателям качества: допустимая задержка пакетов, составляющая 30 мс или менее, является крайне ограниченным параметром, который включает задержку на радиоинтерфейсе, обработку узлов и обратную передачу; для достижения целевого показателя вероятности ошибок пакетов, в системе используется резервирование на нескольких уровнях, что включает в себя дублирование пакетов, когда один и тот же пакет отправляется одновременно по двум независимым каналам радиосвязи, и мультисоединение, когда UE поддерживает соединения с несколькими сотами, обеспечивая бесперебойную и мгновенную передачу в случае сбоя одного канала.

Планировщик gNB использует целевое значение вероятности ошибок пакетов для выбора подходящей схемы модуляции и кодирования и мощности передачи, чтобы гарантировать успешное декодирование с первой попытки.

1.2.3. Сценарий массовых коммуникаций машинного типа

Сценарий массовых коммуникаций машинного типа mMTC использует принципиально иную парадигму, чем требования eMBB к высокой пропускной способности и детерминированные требования URLLC.

Основная цель сценария mMTC – обеспечить повсеместное подключение к большому количеству устройств с низким энергопотреблением и низкой вычислительной мощностью устройств, которые, как правило, передают небольшие, нечастые пакеты данных [16]. Задача заключается не в максимизации скорости передачи данных или минимизации задержек, а в достижении масштабируемости, широкого охвата и энергоэффективности. Сценарий mMTC является основополагающим для Интернета вещей – интеграции датчиков, исполнительных механизмов и счетчиков в цифровую структуру умных городов, промышленности и сельского хозяйства.

Основные показатели эффективности использования сценария mMTC, определенные 3GPP [4], нацелены на достижение плотности подключения до 1 000 000 устройств на кв. км, увеличение срока службы батареи устройства до 10 и более лет и расширение зоны покрытия для доступа к устройствам в труднодоступных местах, таких как подвалы или отдаленные сельские районы.

Основная проблема mMTC заключается не в объеме данных с одного устройства, а в совокупной передаче сигналов и загрузке данных, генерируемых миллионами устройств, пытающихся получить доступ к общему каналу связи. Это создает проблему масштабируемости, с которой традиционные архитектуры сотовой связи, рассчитанные на относительно небольшое число постоянно активных пользователей, плохо справляются [7; 16; 28; 70]. Таким образом, реализация сценария mMTC требует новых процедур доступа к сети, управления состоянием устройств и сигнализации в основной сети для эффективной поддержки большой по объему совокупности устройств, которая к тому же является прерывистой и часто бездействующей [8;75].

Варианты использования mMTC определяются приложениями, которые

используют периодические или запрошенные отчеты о данных из конечных точек. Эти приложения можно разделить на системы мониторинга/телеметрии и отслеживания активов [20;60].

Крупномасштабный мониторинг и телеметрия составляют основу интеллектуальных инженерных сетей, например, интеллектуальных счетчиков газа, воды и электричества, мониторинга окружающей среды, например, качества воздуха, влажности почвы и инфраструктуры "умный город", например, управление отходами, интеллектуальное освещение. В этих приложениях устройства передают небольшие пакеты данных с регулярными, часто длительными интервалами, при этом основным требованием является высокая энергоэффективность, обеспечивающая долгие годы автономной работы без ручного вмешательства.

Отслеживание активов и логистика также представляют собой важную область, где устройства, подключенные к посылкам, транспортным средствам или домашнему скоту, сообщают об их статусе и местоположении. Приложения, работающие по сценарию mMTC, терпимо относятся к задержкам, отдавая приоритет надежной и энергоэффективной доставке обновлений о местоположении, а не мгновенности.

Общей чертой всех вариантов использования mMTC является асимметричный и случайный характер трафика: в отличие от смартфонов, которые остаются в подключенном состоянии, устройства mMTC проводят большую часть своего времени в режиме энергосбережения. Преобладает передача данных по восходящей линии, в то время как связь по нисходящей линии обычно используется для нечастых изменений конфигурации, обновлений программного обеспечения или запросов на опрос [34].

Архитектура системы 5G поддерживает mMTC благодаря сочетанию усовершенствованных технологий 4G IoT [50], а именно узкополосного интернета вещей – стандарта сотовой связи для устройств телеметрии с низкими объемами обмена данными (Narrow Band Internet of Things, NB-IoT) и LTE для компьютеров – стандарта сотовой связи, разработанный специально для устройств интернета

вещей (Long-Term Evolution for Machines, LTE-M), которые интегрированы в платформу 5G New Radio (NR), и встроенных функций 5G NR-light (5G Reduced Capability, RedCap – версии технологии 5G, предназначеннной для устройств среднего уровня производительности), обеспечивающей баланс между высокопроизводительными возможностями стандартных устройств 5G NR и сверхнизкой сложностью устройств LTE-M и NB-IoT.

Установление сеанса передачи данных для устройства mMTC представлено на рисунке 1.10.

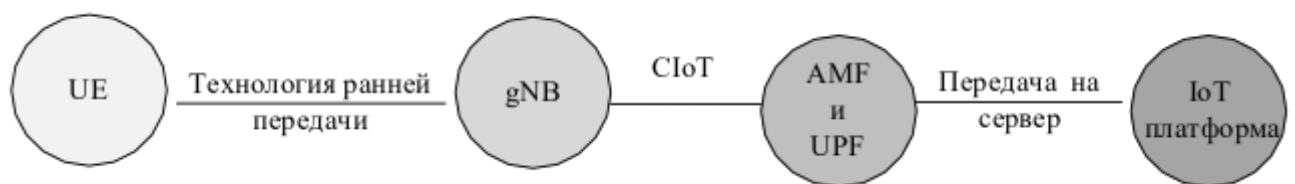


Рисунок 1.10 – Структурная схема потока данных mMTC

Пользовательское оборудование UE, например, датчик или трекер, в основном работает в режиме глубокого сна (режим энергосбережения) или использует расширенный прерывистый прием. Подобные механизмы позволяют устройству отключать большую часть своих схем на длительное время, лишь на короткое время включаясь для запланированной передачи данных или прослушивания канала.

В сценарии mMTC используется ранняя передача данных (Early Data Transmission, EDT), что позволяет UE передавать свои данные на начальных этапах процедуры произвольного доступа, тем самым завершая транзакцию одним сообщением и быстро переходить в спящий режим. Эта технология является основной причиной энергоэффективности, поскольку позволяет сети запланировать передачу устройства на время, оптимальное для загрузки сети, или позволить устройству дождаться следующего запланированного периода пробуждения, тем самым избегая частых и энергозатратных включений радиосвязи.

Предварительно настроенные схемы для передачи данных аналогичны для сценария URLLC, но используются с другой целью:

- на уровне радиосвязи технологии mMTC, такие как NB-IoT, работают в очень узкой полосе пропускания (180 кГц), что обеспечивает значительное увеличение бюджета канала связи за счет усиления обработки, обеспечивая глубокое проникновение в помещения;
- основная сеть, в частности функция управления доступом и мобильностью AMF, предназначена для управления одновременных периодических обновлений регистрации без перегрузки;
- архитектура 5G поддерживает сотовый Интернет вещей (Cellular IoT, CIoT) – технологию, которая позволяет устройствам подключаться к интернету с помощью сотовых сетей. Оптимизация CIoT на уровне управления позволяет небольшим пакетам данных интегрироваться в сигнальные сообщения, что снижает задержку передачи сигналов и энергопотребление устройства.

Модель QoS 5G, основанная на той же платформе 5QI, что и eMBB и URLLC, применяется и в mMTC, но с акцентом на эффективность использования ресурсов и приоритизацию трафика, а не на низкую задержку. Сеть должна отличать аварийный сигнал, поступающий от детектора дыма, от обычных ежечасных показаний температуры, поступающих от погодного датчика.

Для большинства приложений mMTC достаточно стандартного потока QoS с негарантированной скоростью передачи данных. Обычно используется значение 5QI 75/79, которое является стандартным классом QoS для "некритичных, устойчивых к задержкам" данных V2X. Эти классы имеют отличаются только уровнем приоритета по умолчанию, равным 25/65 соответственно, и имеют допустимую задержку передачи пакетов, равную 50 мс, и вероятность потери пакетов из-за ошибок, равный 10^{-2} .

Основным отличием QoS для mMTC является не требование обеспечения гарантированной задержки, а общее поведение сети. Сеть должна быть способной обрабатывать огромное количество одновременных попыток

доступа во время событий, которые приводят в действие множество устройств одновременно, например, отключение электроэнергии, в результате чего все интеллектуальные счетчики отправляют окончательные показания.

1.3. Структура фрейма сети 5G

5G обеспечивает не только более высокую скорость передачи данных и меньшую задержку, но и новую структуру фреймов, которая лучше подходит для работы с разнообразными беспроводными устройствами и приложениями, которые используются сегодня [64].

Система 5G адаптируется к потребностям каждого отдельного приложения, повышая эффективность сети в целом. Одним из ключевых аспектов гибкости структуры 5G является метод распределения частотно-временных ресурсов. Частотно-временные ресурсы разделены на блоки ресурсов, которые затем подразделяются на более мелкие элементы сети. Это обеспечивает оптимальное распределение сетевых ресурсов и более быстрое и бесперебойное подключение пользователей так, что сеть может определять и удовлетворять различные потребности в подключении, обеспечивая, например, высокоскоростное подключение для потоковой передачи видео в формате 4K и подключение с низкой задержкой для автономных транспортных средств.

Передача данных осуществляется порциями – фреймами. Структура фрейма в сетях 5G основана на использовании слотов и символов, что позволяет сети динамически регулировать продолжительность временных интервалов в зависимости от потребностей в обслуживании. Каждый фрейм в свою очередь делится еще на 10 кусочков – субфреймов, субфреймы состоят из слотов, каждый слот – из нескольких символов OFDM (Orthogonal Frequency Division Multiplexing), которые являются наименьшими единицами передачи, как показано на рисунке 1.11.

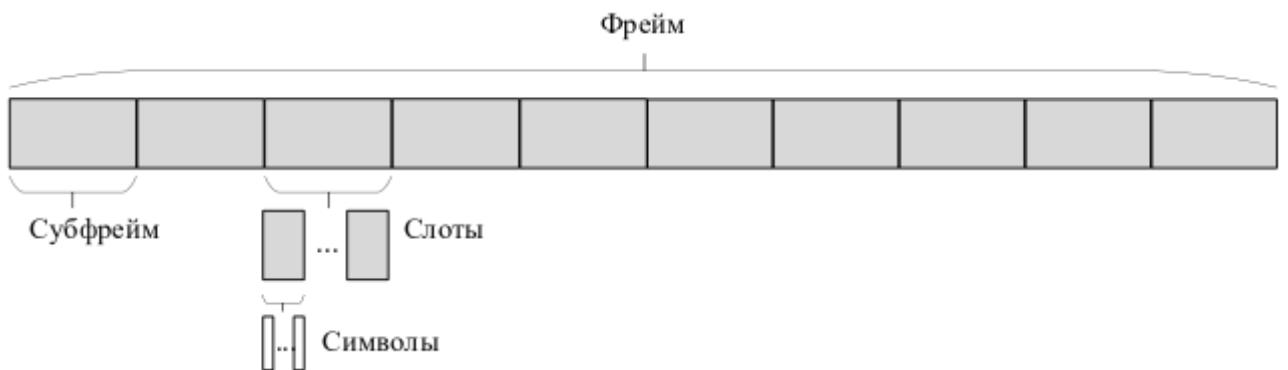


Рисунок 1.11 – Структура фрейма

Количество слотов, из которых состоят субфреймы, может быть разным и зависит оно от ширины спектра поднесущей (нумерологии).

В сети 5G преимущественно применяется технология цифровой модуляции с использованием большого количества близко расположенных ортогональных поднесущих – мультиплексирование с ортогональным частотным разделением каналов OFDM (Orthogonal frequency-division multiplexing).

Нумерология в системах OFDM относится к набору параметров, которые определяют структуру физического уровня. Эти параметры включают расстояние между поднесущими, длительность символа и длину циклического префикса. Независимо от нумерологии, длина одного фрейма и длина одного субфрейма одинакова. Длина фрейма всегда равна 10 мс, а длина субфрейма всегда равна 1 мс. Разница заключается в том, чтобы разместить разное количество слотов в одном субфрейме. При этом количество символов в слоте не меняется в зависимости от нумерации, оно меняется только в зависимости от типа конфигурации слота. Учет физических свойств различных нумерологий достигается путем изменения количества интервалов в пределах одного субфрейма. Для слота с обычным циклическим префиксом количество символов в слоте всегда равно 14, а для слота с расширенным циклическим префиксом количество символов в слоте всегда равно 12.

Наличие нескольких нумерологий обеспечивает дополнительную

гибкость при планировании различных служб. Для приложений, которые допускают больший разброс задержек, предпочтительны более узкие расстояния между поднесущими. В таблице 1.1 представлен обзор нумерологий 5G-NR и соответствующего количества слотов на фрейм и субфрейм.

Таблица 1.1 – Нумерологии 5G и соответствующие параметры

Индекс нумерологии	Количество символов в слоте	Количество слотов в фрейме	Количество слотов в субфрейме	Длительность слота, мс
0	14	10	1	1
1	14	20	2	0.5
2	14	40	4	0.25
3	14	80	8	0.125
4	14	160	16	0.0625
5	14	320	32	0,03125
6	14	640	64	0,015625

При этом планировщик обычно работает на уровне слотов. В первом выпуске стандарта 5G New Radio (NR Release 15) сформулированы спецификации физического уровня, где есть функция, которая позволяет передавать только часть временного интервала, используя минимум два символа OFDM в нисходящей и восходящей линиях связи соответственно. Это называется мини-интервалом. Такие короткие передачи используются в сценариях, требующих минимальной задержки, таких как сервисы URLLC.

Другим важным аспектом NR является концепция частичной пропускной способности – Bandwidth Part (BWP). BWP обеспечивает большую гибкость в распределении ресурсов в рамках данной несущей. С помощью BWP несущую можно разделить на более мелкие части и использовать для различных целей. Каждый BWP имеет свои собственные параметры, включая полосу

пропускания и нумерацию. Полоса пропускания настраивается для каждого UE в зависимости от его способности поддерживать максимальную поддерживаемую полосу пропускания. Таким образом, несколько UE с различными возможностями могут обслуживаться на одном широкополосном NR-носителе. Сервисы с большим объемом данных получают более длительные интервалы, в то время как сервисы быстрого реагирования, получают более короткие интервалы. Такая гибкость повышает эффективность и оперативность реагирования сети 5G.

Еще одним аспектом NR является концепция мини-слотов 5G. Мини-слоты позволяют фреймовой структуре 5G обеспечивать более быстрое реагирование на определенные приложения. В чрезвычайных ситуациях, когда важна каждая доля секунды, использование мини-слота может обеспечить быструю доставку важной информации.

Кроме того, фрейм 5G включает в себя как автономные, так и несамостоятельные субфреймы, обеспечивая еще один уровень гибкости. Это позволяет использовать различные методы передачи данных в зависимости от конкретных потребностей приложения, сохраняя качество обслуживания и еще больше сокращая время ожидания.

1.4. Постановка задачи исследования

Сети пятого поколения представляют собой сложную инфраструктуру, предназначенную для обслуживания разнородного трафика с различными требованиями к качеству обслуживания. Высокая плотность подключенных устройств, включая как пользовательские терминалы, так и устройства интернета вещей, создают значительную нагрузку на доступные ресурсы сети, такие как частотный спектр, энергия и вычислительные мощности. В таких условиях возникает проблема эффективного распределения пула ресурсов, обеспечивающего выполнение показателей QoS для различных классов трафика.

Анализ архитектуры сети 5G и требования к качеству обслуживания при использовании разных сценариев показывает, что ключевыми задачами проектирования современных сотовых сетей являются:

- обеспечение сосуществования трафика URLLC и mMTC в восходящем канале связи. URLLC предъявляет жесткие требования к задержкам (менее 1 мс) и надёжности передачи данных (до 99,999%), что требует немедленного выделения ресурсов. В то же время mMTC характеризуется массовостью подключений, но при этом допускает более высокие задержки. Проблема заключается в том, что mMTC-устройства часто используют механизмы случайного доступа, такие как Grant-Free передача, что приводит к коллизиям и увеличению интерференции. Это создает дефицит ресурсов для URLLC, особенно в условиях высокой плотности устройств. Для решения этой задачи необходимы методы динамического резервирования ресурсов, включая выделение мини-слотов для URLLC и приоритезацию трафика с помощью динамического изменения числа доступных преамбул в зависимости от загруженности сети.

- обеспечение сосуществования трафика eMBB и URLLC в нисходящем канале связи. eMBB ориентирован на предоставление услуг с высокой пропускной способностью, но при этом менее чувствителен к задержкам по сравнению с URLLC. Однако приоритетное выделение ресурсов для URLLC может негативно сказаться на производительности eMBB, особенно в условиях ограниченной полосы пропускания. Для минимизации такого влияния необходимы адаптивные алгоритмы планирования ресурсов, основанные на методах мультиплексирования потоков, которые способны динамически перераспределять ресурсы между eMBB и URLLC с учетом текущей нагрузки и требований QoS. Дополнительно может применяться гибкое разделение полосы частот, при котором часть ресурсов, выделенных для eMBB, временно заимствуется для обслуживания URLLC-трафика.

- организация динамического перераспределения нагрузки между

базовыми станциями. Неравномерное распределение пользователей и устройств может приводить к перегрузке одних базовых станций при недогрузке других, что приводит к ухудшению QoS. Для решения этой задачи необходимы механизмы координации между базовыми станциями, где управление ресурсами осуществляется централизованно через облачную платформу. Это позволяет оптимизировать распределение нагрузки благодаря динамическому переключению пользователей между станциями (Handover) на основе прогнозируемой и текущей нагрузки.

Таким образом, обеспечение требуемого уровня QoS в сетях 5G с высокой плотностью устройств требует комплексного подхода, включающего как приоритизацию критически важного трафика, так и оптимизацию доступа для массовых устройств интернета вещей.

Сформулируем задачу диссертационного исследования, как предложение эффективных методов распределения ресурсов сети для обеспечения требуемого уровня QoS сети 5G:

$$\begin{cases} \bar{T}_{BK} \leq \bar{T}_{BK}^{dop}, \text{при URLLC} \cup \text{mMTC} \\ \bar{T}_{HK} \leq \bar{T}_{HK}^{dop}, \text{при eMBB} \cup \text{URLLC} \end{cases} \quad (1.1)$$

где \bar{T}_{BK} – среднее время задержки в восходящем канале;

\bar{T}_{BK}^{dop} – допустимое среднее время задержки в восходящем канале, определяемое QoS.

$$\begin{cases} P_{BK}^{Err} \leq P_{BK}^{Err,dop}, \text{при URLLC} \cup \text{mMTC} \\ P_{HK}^{Err} \leq P_{HK}^{Err,dop}, \text{при eMBB} \cup \text{URLLC} \end{cases} \quad (1.2)$$

где P_{BK}^{Err} – вероятность ошибки восходящего канала;

$P_{HK}^{Err,dop}$ – допустимая вероятность ошибки восходящего канала, определяемое QoS.

$$a(t) = \left[a_i(t), i = \overline{1, m} \right], \quad (1.3)$$

$$\forall t: \sum_t a_i(t) \leq 1,$$

где $a_i(t)$ – доля ресурсов канала, выделенная i -му устройству в момент времени t с учетом того, что доступная полоса частот для передачи сообщений является ограниченной;
 m – количество UE.

Выводы по первой главе

Рассмотрены три основных типа трафика в сетях 5G: расширенная мобильная широкополосная связь, сверхнадёжная связь с низкой задержкой и массовые коммуникации машинного типа. Показано, что каждый из этих сценариев предъявляет уникальные требования к пропускной способности, задержке и надёжности, что создаёт сложные задачи для проектирования и оптимизации сетей.

Особое внимание уделено передаче M2M-трафика в сотовых сетях, где выявлены существенные различия между коммуникациями типа "человек-человек" (H2H) и "машина-машина" (M2M). Показано, что традиционные методы распределения ресурсов, разработанные для H2H, недостаточно эффективны для M2M-устройств, которые характеризуются высокой плотностью подключений, низким энергопотреблением и разнородными требованиями к качеству обслуживания. Это обуславливает необходимость разработки новых моделей и методов, адаптированных к особенностям M2M-трафика.

Концептуальная модель сотовых сетей 5G, представленная в главе, подчёркивает важность гибкости и адаптивности при распределении ресурсов. Введённое понятие нумерологии 5G позволяет динамически настраивать параметры сети, такие как интервалы между поднесущими и длительность

символов, что обеспечивает поддержку разнородных сценариев использования. Это особенно актуально для mMTC и URLLC, где критически важны низкие задержки и высокая надёжность.

Постановка задачи исследования отражает необходимость комплексного подхода к управлению ресурсами в условиях высокой плотности устройств и гетерогенного трафика. Основные проблемы включают обеспечение совместного обслуживания mMTC и URLLC в восходящем канале, мультиплексирование eMBB и URLLC в нисходящем канале, а также динамическую ассоциацию пользователей с базовыми станциями. Решение этих задач требует разработки инновационных методов, сочетающих математическое моделирование, алгоритмическую оптимизацию и современные технологии машинного обучения.

Проведенный анализ подтверждает актуальность темы диссертации и формирует теоретическую основу для последующих глав, в которых предлагаются конкретные методы и алгоритмы для повышения эффективности распределения ресурсов в сотовых сетях 5G. Результаты, изложенные в первой главе, служат фундаментом для дальнейших исследований, направленных на решение выявленных проблем и совершенствование работы сетей 5G в условиях высокой нагрузки и разнородного трафика.

2. РАСПРЕДЕЛЕНИЕ РЕСУРСОВ В ВОСХОДЯЩЕМ КАНАЛЕ СЕТИ 5G

Проблема случайного доступа рассматривается как основное узкое место в модели восходящего канала, поскольку восходящий канал связи в сетях 5G характеризуется массовыми подключениями устройств, генерирующих короткие пакеты данных с низкой периодичностью [17; 32; 33; 74].

Вторая глава посвящена решению задачи сосуществование трафика URLLC и mMTC в восходящем канале связи.

Для строгой формулировки решаемой задачи введем следующие обозначения и параметры системы:

- D_m^i – средняя скорость передачи данных для i -го пользователя mMTC;
- D_u^i – средняя задержка передачи пакетов для i -го пользователя URLLC;
- P_{loss}^i – вероятность потери пакетов для i -го пользователя URLLC;
- L – максимальное число доступных преамбул;
- m_0 – число преамбул для URLLC устройств;
- m_1 – число преамбул для mMTC устройств.

Задача оптимизации параметров мультиплексирования формулируется следующим образом:

$$\sum_i D_u^i + \gamma D_m^i \rightarrow \min, \quad (2.1)$$

при условии:

$$m_0 + m_1 = L,$$

$$\forall i: P_{\text{loss}}^i < P_{\text{max}}.$$

2.1. Методы случайного доступа к каналу

Случайный доступ – это процесс, который позволяет пользовательскому оборудованию эффективно и динамично осуществлять синхронизацию по

восходящей линии связи и получать «идентификационный номер» для связи по общему радиоканалу. Этот процесс называется процедурой Random Access Channel (RACH) – канал произвольного доступа, которая также известна как "Нисходящая синхронизация + RACH".

Существует два типа процедур RACH с точки зрения допустимых физических ресурсов: времени и частотного местоположения PRACH, и индекса последовательности PRACH [8; 61]:

- Произвольный доступ на основе конкуренции (RACH, основанной на конфликтах);
- Произвольный доступ без конкуренции (RACH, не основанной на конфликтах).

При произвольном доступе, основанном на конкуренции, UE случайным образом выбирает преамбулу из набора преамбул, которые доступны другим UE.

Во время этой процедуры множество пользователей могут выбирать одну и ту же преамбулу с произвольным доступом и передавать ее на базовую станцию. Учитывая, что несколько пользователей могут выбрать одну и ту же преамбулу, существует вероятность возникновения помех или коллизии. После отправки преамбулы пользователи ожидают ответа от базовой станции. Если они получают положительный ответ, то отправляют дополнительную информацию для установления соединения. Впоследствии сеть подтверждает успешный доступ одного из активных устройств с помощью уникального идентификатора в пределах соты C-RNTI (Cell Radio Network Temporary Identifiers). Пользователи, которым не удается получить подтверждение, узнают о наличии помех и инициируют процесс отмены, прежде чем повторить попытку.

Подход на базе произвольного доступа на основе конкуренции обычно используется для первоначального доступа, когда UE не синхронизировано с сетью или когда сети не хватает предварительных знаний о пользователе. Но в некоторых случаях такого рода конфликты неприемлемы по какой-либо причине (например, из-за ограничения по времени) [29; 30; 31].

При произвольном доступе без конкуренции сеть сообщает каждому UE о том, когда именно и какую подпись в преамбуле он должен использовать. Сеть сама распределяет подписи в преамбуле таким образом, чтобы они не конфликтовали. Чтобы запустить процесс произвольного доступа без конкуренции, UE должно находиться в режиме подключения до начала процесса RACH [59].

Преамбула состоит из трех частей, представленных на рисунке 2.1.

Циклический префикс	Тело преамбулы	Заданный интервал
---------------------	----------------	-------------------

Рисунок 2.1 – Структура преамбулы

Различное значение циклического префикса и защитного интервала определяет формат преамбулы. Каждый из форматов предназначался для использования в специфических условиях. Разные форматы в теории позволяют организовывать соты радиусом от 14 до 108 км. Определено два значения длины преамбулы: 839 отсчетов – длинная преамбула и 139 отсчетов – короткая преамбула.

В 5G/NR существует больше форматов преамбулы по сравнению с LTE [5]. В LTE существует всего 4 различных типа преамбул, которые называются типом 0,1,2,3. В формате 5G/NR определено суммарно 13 форматов преамбул: 0,1,2,3, A1, A2, A3, B1, B2, B3, B4, C0, C2.

Еще одним отличием 5G от LTE является поддержка двухэтапного RACH для начального доступа. В LTE RACH для первоначального доступа всегда состоит из 4 этапов, тогда как в 5G поддерживается двухэтапный RACH даже для первоначального доступа.

Схема четырехэтапной процедуры случайного доступа представлена на рисунке 2.2,а. Устройство, которое намеревается сначала передать данные, случайным образом выбирает преамбулу из набора преамбул, доступных в данном

слоте. Преамбулы в сетях 4G и 5G генерируются с использованием последовательностей Задоффа-Чу. Эти преамбулы представляют собой последовательности с постоянной амплитудой и коэффициентом корреляции последовательности с её любой циклически сдвинутой копией (за исключением сдвига на 0 или определённое количество элементов) равным 0, а при сдвиге на 0 или определённое количество элементов нормированная функция автокорреляции равна 1, что облегчает точное определение преамбулы и времени ее появления.

Однако последовательности Задоффа-Чу трудно генерировать в режиме реального времени, кроме того, требуется большой объем памяти для хранения кода из-за требования ортогональности [7]. Таким образом, количество доступных ортогональных преамбул ограничено, и базовая станция сообщает об этом через общее сообщение конфигурации физического канала произвольного доступа. Как правило, на каждый слот случайного доступа в данной ячейке приходится 64 преамбулы, и 10 из них зарезервированы для бесконфликтного доступа, в то время как остальные 54 преамбулы распределены между конкурирующими устройствами.

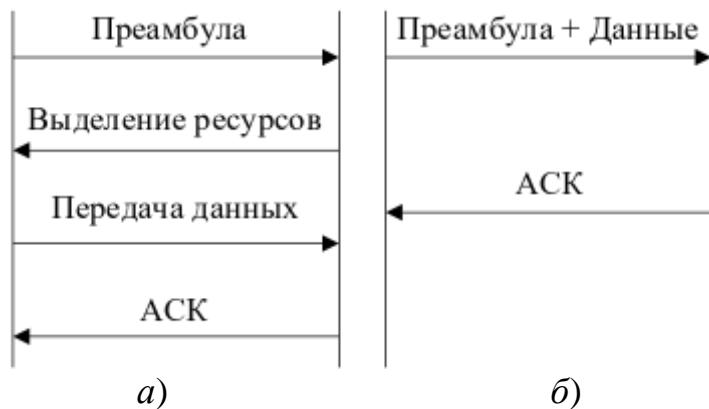


Рисунок 2.2 – Процедуры доступа к разделяемому ресурсу:

a – четырехэтапная; *б* – двухэтапная

Перед передачей данных каждое устройство случайным образом выбирает преамбулу из числа доступных вариантов и передает ее в следующий свободный интервал произвольного доступа, указанный в конфигурации физического канала произвольного доступа. Базовая станция периодически

распространяет информацию о доступных преамбулах, их форматах и настройках физического канала RACH на все устройства в пределах своей ячейки посредством сообщений системного информационного блока 2 (SIB2).

В случае, если два или более устройств выбирают одну и ту же преамбулу и передают ее одновременно в один и тот же интервал произвольного доступа, возникает конфликт. В таком случае всем участвующим устройствам часто требуется повторно отправить преамбулу после случайного периода задержки, выбранного из заранее определенного интервала ожидания. Если конфликтов обнаружено не будет, то базовая станция отправит сообщение с ответом на случайный доступ (RAR), за которым последует серия сообщений, связанных с разрешением конфликтов. Как только все эти шаги будут успешно выполнены, процесс первоначального доступа будет считаться успешным.

Базовая станция периодически передает сообщения, содержащие важные сведения о процедуре произвольного доступа ко всем устройствам в ячейке: информацию об индексе конфигурации физического канала произвольного доступа, который определяет частоту интервала произвольного доступа во временной области и тип преамбулы.

Двухэтапная процедура доступа (рисунок 2.2, б) может быть применена в случае, если UE находится в активном режиме. Принцип, лежащий в основе двухэтапной процедуры, заключается в объединении преамбулы и информации о данных в одно сообщение. Затем базовая станция ответит на успешный прием. Из-за уменьшения количества передаваемых сообщений двухэтапная процедура может снизить задержку и нагрузку на сеть по мере увеличения количества точек доступа. Устройство UE считает двухэтапную процедуру успешно завершенной после получения положительной квитанции ACK, после успешного декодирования данных. В противном случае, в случае неудачного обнаружения преамбулы или сбоя в декодировании данных, UE повторно передаст «Преамбула +Данные» или вернется к четырехэтапной процедуре доступа. В сообщении обратной связи от базовой станции к UE может

содержатся указание на отключение для следующей передачи RACH или резервный код, позволяющий вернуться к четырехэтапной процедуре.

Проблема перегрузки канала произвольного доступа является серьезной проблемой, связанной с процедурами произвольного доступа, в первую очередь из-за ограниченного числа ортогональных преамбул, доступных в каждом временном интервале для устройств в сети. В сценариях массовой машинной связи mMTC, где используется большое количество устройств, частые коллизии при передачах между этими устройствами во время интервала RACH приводят к большому числу повторных передач и увеличению задержки.

Одним из подходов к управлению доступом в сетях LTE-A и 5G является использование методов контроля доступа.

2.2.Модель трафика восходящего канала

Трафик устройств можно разделить на два вида – периодичный, при котором устройства отправляют данные с определенной периодичностью и апериодичный – данные передаются ситуационно.

В работе рассмотрена модель burst-трафика, представленная на рисунке 2.3. Такой режим работы обусловлен самой природой данных, которые генерируют устройства. В отличие от периодичного трафика, burst-трафик характерен для событийно-ориентированных систем, каким и являются сети интернета вещей. Ключевыми параметрами, описывающими каждый всплеск трафика, являются его длина, интенсивность и объем. Длина всплеска – это промежуток времени, в течение которого идет непрерывная передача данных на высокой скорости. Интенсивность, или скорость передачи внутри этого окна, обычно близка к максимальным техническим возможностям устройства. Объем данных – это произведение длины всплеска на его интенсивность, то есть общее количество мегабайт, переданных за один активированный сеанс связи. Также важен промежуток между всплесками – случайная или псевдослучайная величина, которая определяет период неактивности между последовательными

передачами. Подобная модель создает специфическую нагрузку на сеть: в отличие от плавного и предсказуемого периодического трафика, burst-трафик порождает резкие, почти мгновенные пики нагрузки. Сетям, рассчитанным на среднюю, а не пиковую пропускную способность, становится сложно обрабатывать такие концентрированные выбросы данных от множества устройств одновременно. Это может приводить к задержкам, потерям пакетов и необходимости применения специальных механизмов управления очередями и приоритизации трафика.

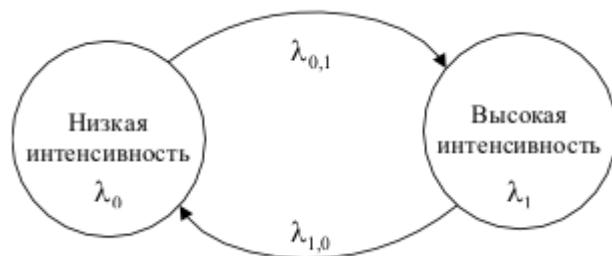


Рисунок 2.3 – Модель burst-трафика

Рассматриваемая модель burst-трафика включает в себя два режима работы сети – обычный режим, при котором каждое устройство передает трафик с низкой интенсивностью λ_0 и burst-режим с высокой интенсивностью трафика λ_1 . При этом время между появлением сообщений заданы произвольным распределением с математическим ожиданием $1/\lambda_0$, $1/\lambda_1$ соответственно. Переходы между этими состояниями заданы интенсивностью перехода λ_{01} – интенсивность перехода из состояния 0 в состояние 1, λ_{10} – интенсивность перехода из состояния 1 в состояние 0. Все параметры такой модели могут быть заданы произвольными распределениями с конечными математическим ожиданием и дисперсией.

2.3. Описание процесса передачи данных в восходящем канале

Обозначим время, затрачиваемое на передачу (рисунок 2.4):

- преамбулы T_{pr} ;

- сообщения T_{tx} ;
- ответа или обратной связи T_{dp} .

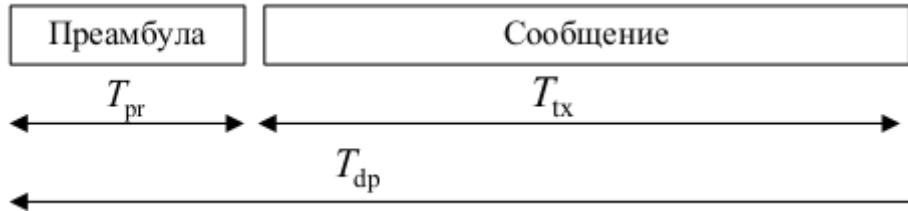


Рисунок 2.4 – Передача преамбулы и сообщения по восходящему каналу связи

Если пакет успешно декодирован, базовая станция отправляет положительную квитанцию ACK по обратной связи, в противном случае она отправляет отрицательную квитанцию NACK. После получения и обработки обратной связи UE может решить, следует ли выполнять повторную передачу.

Отметим, что T_{pr} и T_{dp} являются фиксированными величинами, а время передачи T_{tx} может варьироваться от одного до нескольких TTI (Transmission Time Interval – это интервал времени, в течение которого данные передаются в радиоинтерфейсе беспроводной сети) в зависимости от размера сообщения, качества канала и стратегии планирования [40;46].

Рассмотрим беспроводной канал "точка-точка" и предположим, что каждое сообщение представляет собой кодовое слово. Достигимая скорость передачи кода длиной n определяется [23; 69] как

$$R(p, n, \varepsilon) = \log(1 + p) - \sqrt{\frac{V(p)}{n}} Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \quad (2.2)$$

где p – отношение сигнал/шум (SNR),

ε – вероятность ошибки (в кодовом слове или пакете);

$V(p)$ – дисперсия канала, которая определяется как

$$V(p) = \frac{2 + p}{(1 + p)^2} (\log \varepsilon)^2. \quad (2.3)$$

В качестве альтернативы вероятность ошибки может быть рассчитана как

$$\varepsilon = Q \sqrt{\frac{n}{V(p)}} \log(1+p) - \frac{N_{\text{bit}}}{n}, \quad (2.4)$$

где N_{bit} представляет собой количество битов данных в пакете, а n может быть интерпретировано как количество битов для передачи по каналу.

Здесь скорость передачи кода, N_{bit}/n , должна быть ниже пропускной способности, $\log_2(1+p)$, для обеспечения низкой вероятности ошибки. В результате канал можно рассматривать как канал со стиранием с вероятностью стирания для каждой пакетной передачи ε .

В этой модели используются следующие допущения:

- Состояние канала постоянно для всех устройств и задается вероятностью приема ошибочного пакета ε .
- Предполагается, что в исходящем канале связи ошибки отсутствуют, таким образом устройство всегда узнает, было ли корректно доставлено сообщение или нет.

Рассмотрим процедуру нахождения средней задержки для сценария mMTC. Пусть P_e – вероятность ошибки при передаче фрейма с данными (конфликт преамбул или ошибки, произошедшие при передаче по беспроводному каналу), тогда $Q = (1 - P_e)$ – вероятность успешной передачи сообщения.

Пусть K – число повторных передач, потребовавшееся для успешной отправки сообщения, тогда $\Pr\{K = k\} = QP_e^{k-1}$. Тогда математическое ожидание числа передач можно рассчитать, как

$$E[K] = QP_e^0 + 2QP_e^1 + \dots + kQP_e^{k-1} = \sum_{k=1}^{\infty} kQP_e^{k-1} = \frac{Q}{(P_e - 1)^2} = \frac{1}{1 - P_e} = \frac{1}{Q}. \quad (2.5)$$

Общее время обслуживания T можно рассчитать, как произведение числа повторных передач на время каждой передачи $E[S] = E[K]T$.

Без ограничения общности будем считать, что $T = 1$, тогда

$$E[S] = E[K] = \left(\Pr\{\text{Нет конфликта}\} \Pr\{\text{Нет ошибок}\} \right)^{-1}. \quad (2.6)$$

При этом вероятность конфликта при передачи устройством сообщения\

$$\Pr\{\text{Нет конфликта}|M = m\} = \left(1 - \frac{1}{L}\right)^m \quad (2.7)$$

где M – число активных устройств в системе, не включая рассматриваемое; L – число использующихся преамбул в системе.

Исходя из формулы 2.5 следует, что для уменьшения задержки передачи сообщения необходимо рассматривать две задачи – уменьшение вероятности конфликтов при выборе преамбулы и уменьшения вероятности ошибки при передаче сообщения.

2.4. Обеспечение надежности передачи данных в межмашинном взаимодействии

Для достижения высокой надежности используются протоколы с гибридной обратной связью (Hybrid Automatic Repeat and Request, HARQ): кодированные пакеты передаются в канал, и, если приемник не может их декодировать из-за замирания канала, помех или любых других причин то выполняется повторная передача.

В целом, протоколы HARQ позволяют достичь высокой надежности [45;67], но частные повторные передачи увеличивают задержку декодирования пакетов [47;65].

2.4.1. Метод K -повторений

Эффективным средством снижения задержки в HARQ является использование разнесения передачи, например, один и тот же пакет может

передаваться несколько раз, что называется K -повторением [6; 57]. Поскольку вероятность успешной передачи увеличивается с увеличением числа повторений K , то уменьшение числа повторений приводит к короткой задержке декодирования за счет спектральной эффективности. Это часто приемлемо для соблюдения строгих ограничений по задержке, когда пропускная способность достаточно большая [49].

Предположим, что каждый закодированный пакет передается K раз в соответствии с K -повторением. Это приводит к повышению надежности за счет снижения спектральной эффективности в K раз. Величина $1/K$ упоминается как коэффициент спектральной эффективности или кодовая скорость.

Схема передачи сообщений при использовании K -повторений показана на рисунке 2.5, где устройство сконфигурировано для автономной передачи одного и того же пакета $K = 3$ раз в последовательных ТТИ. После получения K копий сообщения базовой станции необходимо разделить принятую последовательность, обработать принятый пакет и отправить обратную связь на устройство.

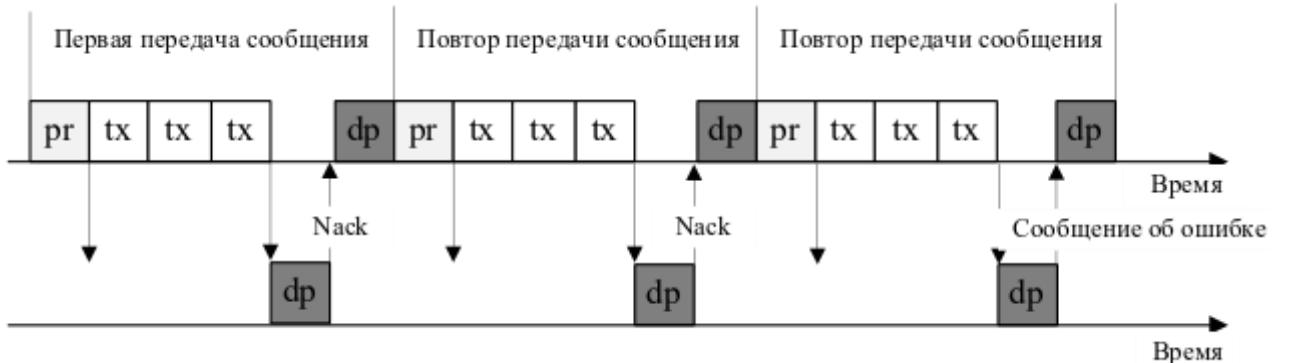


Рисунок 2.5 – Схема передачи с использованием K -повторений

Приемник способен декодировать пакет, если хотя бы одна из K копий может быть правильно декодирована. Таким образом, вероятность ошибки декодирования составит $\varepsilon_K = \varepsilon^K$.

2.4.2. Алгоритм скользящего сетевого кодирования

Использование линейного сетевого кодирования может обеспечить значительное повышение надежности по сравнению с K -повторением при той же спектральной эффективности [36]. Однако такая система требует доставки всех M пакетов данных, то есть отдельные приоритеты для первых пакетов не определяются. Таким образом, успешного декодирования начального пакета, приемнику необходимо получить как минимум M пакетов. Это создает фундаментальное ограничение – минимальная задержка для первого пакета составляет M единиц времени. Когда значение M превышает K , становится очевидным, что задержка декодирования при использовании сетевого кодирования оказывается выше, чем в схеме с K -повторениями.

Отметим, что M представляет собой минимально возможную задержку в сетевом кодировании, тогда как K является максимальным показателем задержки при K -повторении. Даже при равенстве $M = K$, практическая задержка передачи пакетов в сетевом кодировании обычно превышает аналогичный показатель для K -повторения. Таким образом, можно сделать вывод, что технология сетевого кодирования обеспечивает повышенную надежность по сравнению с K -повторением, однако достигается это за счет увеличения времени передачи, что делает сетевое кодирование не совсем подходящим для сценария URLLC.

При постоянной скорости генерации пакетов передающему устройству приходится ожидать накопления достаточного количества исходных пакетов для выполнения операций сетевого кодирования, что создает дополнительную задержку на передающей стороне. Для сокращения времени декодирования в системах сетевого кодирования можно уменьшать размер групп пакетов. Например, при необходимости передачи 20 пакетов их можно разделить на 4 группы по 5 пакетов, что снижает M с 20 до 5. Однако, при низкой вероятности ошибок в канале порядка 10^{-3} вероятность ошибочного декодирования может превышать аналогичный показатель для K -повторения, например, где $K=3$. Более того, задержка декодирования, например, при $M=5$ остается выше, чем

при K -повторении. Следовательно, уменьшение размера групп пакетов в сетевом кодировании приводит к снижению общей эффективности системы.

Исходя из вышеописанных преимуществ и недостатков использования сетевого кодирования, предлагается алгоритм, основанный на свойствах сетевого кодирования, позволяющий уменьшить задержку декодирования и декодировать пакеты «на лету». При таком подходе пакеты сетевого кодирования, которые представляют собой линейные комбинации исходных пакетов данных, передаются вместе с исходными пакетами данных. В частности, пакеты сетевого кодирования генерируются с использованием скользящего окна исходных пакетов данных, и по этой причине предлагаемый подход называется скользящим сетевым кодированием (Sliding Network Coding, SNC) [79; 86; 88].

Чтобы проиллюстрировать идею скользящего сетевого кодирования, рассмотрим пример с числом пакетов в одной передаче $K=2$.

Предположим, что передаются 2 блока, которые состоят из четырех последовательных пакетов

$$(V_{1,m}, V_{2,m}) = (X_m, X_{m-1} \oplus X_m),$$

где X – пакет, который передается в k -м блоке $V_{k,m}$.

В таблице 2.1 приведены пример таких пакетов.

Таблица 2.1 – Пример скользящего кодирования

m	1	2	3	4	...
$V_{1,m}$	X_1	X_2	X_3	X_4	...
$V_{2,m}$	X_1	$X_1 + X_2$	$X_2 + X_3$	$X_3 + X_4$...

Предположим, что получатель должен декодировать пакет X_{m-1} . Например, если $m = 4$, и приемник должен декодировать X_3 . Тогда $V_{1,4} = X_4$ и $V_{2,4} = X_3 + X_4$. Если все предыдущие пакеты были успешно декодированы, тогда вероятность ошибки декодирования для X_3 может быть рассчитана по следующей формуле:

$$P_3 = \varepsilon^2 \left(1 - (1 - \varepsilon)^2\right) = 2\varepsilon^3 - \varepsilon^4 (1 - \varepsilon)^2.$$

Первый множитель получается из вероятности того, что не будут корректно декодированы блоки $V_{1,3}$ и $V_{2,3}$ (так как считаем, что X_2 было успешно декодировано, то можем получить $X_3 = V_{2,3} + V_{1,2}$). Вторая часть выражения показывает вероятность того, что будут ошибочно декодированы блоки $V_{2,4}$ и $V_{1,4}$. Для успешного декодирования X_3 должны быть успешно декодированы оба блока.

Также можно показать, что X_1 декодируется при $m=2$ со следующей вероятностью ошибки декодирования:

$$P_{31} = \varepsilon^2 \left(1 - (1 - \varepsilon)^2\right) = 2\varepsilon^3 - \varepsilon^4.$$

Следовательно, может потребоваться уменьшить вероятность ошибки декодирования для первого пакета данных, X_1 .

Чтобы проиллюстрировать это, рассмотрим пример в таблице 2.2. Можем видеть, что X_1 может быть декодирован при трех принятых блоках с вероятностью ошибки декодирования

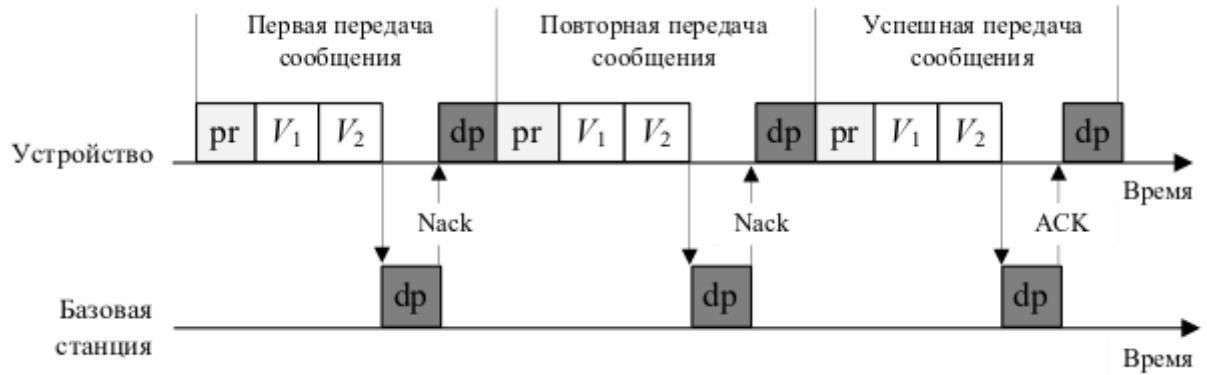
$$P_1 = \varepsilon^2 \left(1 - (1 - \varepsilon)^2\right)^2 = 4\varepsilon^4 + O(\varepsilon^5).$$

Очевидно снижение вероятности ошибки декодирования X_1 за счет задержки. То есть X_{m-2} может быть декодирован в блоке m .

Таблица 2.2 – Пример скользящего кодирования с $K=2$ и задержкой в 2 блока

m	1	2	3	4	...
$V_{1,m}$	X_1	X_2	X_3	X_4	...
$V_{2,m}$	X_1	$X_1 + X_2$	$X_1 + X_3$	$X_2 + X_4$...

Временная диаграмма передачи сообщений с использованием скользящего сетевого кодирования с $K=2$ будет выглядеть, как на рисунке 2.6.

Рисунок 2.6 – Временная диаграмма передачи пакетов SNC при $K=2$

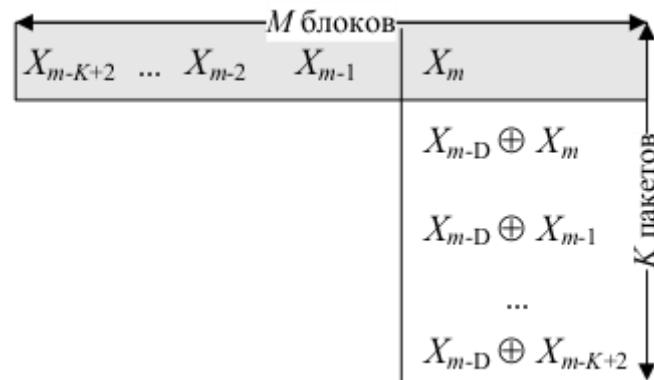
Рассмотрим обобщение схемы скользящего сетевого кодирования, описанной ранее. В общем случае, любую схему скользящего сетевого кодирования можно описать тремя параметрами $SNC(K, D, M)$:

K – число пакетов в одной передаче, определяет эффективность использования ресурсов канала;

D – задержка кодирования / декодирования, определяет задержку декодирования i -го блока сообщения;

M – число блоков сообщения, определяет число блоков на которое разбивается исходное сообщение.

На рисунке 2.7 изображена обобщенная схема $SNC(K, D, M)$.

Рисунок 2.7 – Обобщенная схема $SNC(K, D, M)$

Несмотря на простоту использования схем кодирования с величиной задержки равной $D=K-1$, данная схема сильно увеличивает максимальную задержку декодирования пакета вплоть до квадратичной зависимости от числа пакетов. Но за счет большей задержки декодирования такая схема гарантирует максимальную надежность при передаче данных. При этом легко показать, что параметр задержки не должен быть меньше двоичного логарифма числа пакетов.

2.4.3. Результаты сравнения методов скользящего сетевого кодирования и K -повторений

Для оценки эффективности метода скользящего сетевого кодирования выполнено сравнение предложенной схемы со схемой K -повторений с помощью имитационного моделирования. Сравнение происходило по следующий двум важным критериям:

1. Средняя задержка передачи сообщения;
2. Вероятность потери сообщения из-за серии неудачных передач.

Параметры моделирования приведены в таблице 2.3.

Таблица 2.3 – Параметры моделирования

Интенсивности $\lambda_0, \lambda_1, \lambda_{01}, \lambda_{10}$ (1/TTI)	0.001, 0.25, 0.1, 0.8
Законы распределения для $1/\lambda_0, 1/\lambda_1, 1/\lambda_{01}, 1/\lambda_{10}$	W, W, Exp, Exp
Количество копий при передаче, k	2
Максимальное количество передач одного сообщения, K	3
Количество блоков для сетевого кодирования, M	4
Вероятность ошибки в канале, P_{err}	0.1 ... 0.01
Количество преамбул, L	64

На рисунке 2.8 изображен график зависимости средней задержки от общего числа устройств в системе. Из этого графика видно, что метод сетевого кодирования обеспечивает меньшую задержку передачи сообщения, чем метод K -повторений. При этом из рисунка 2.9 следует, что вероятность неудачной передачи сообщений вследствие превышения лимита повторных отправок для метода, основанном на сетевом кодировании не превышает эту вероятность для метода K -повторений.

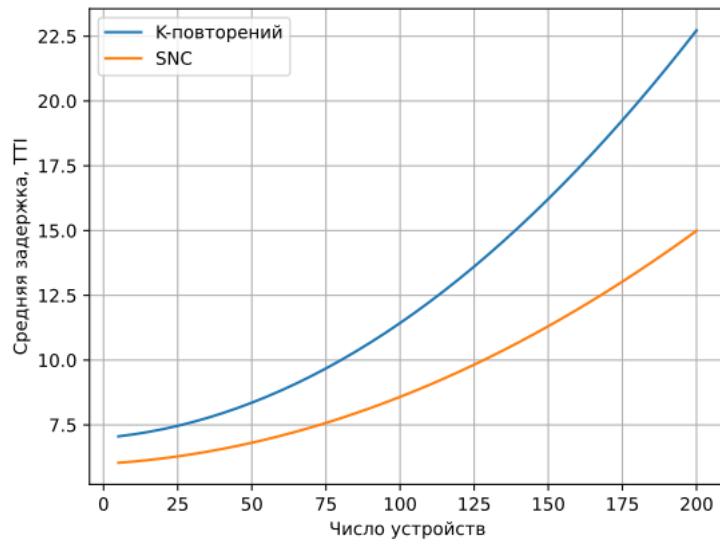


Рисунок 2.8 – График зависимости средней задержки от общего числа устройств в системе

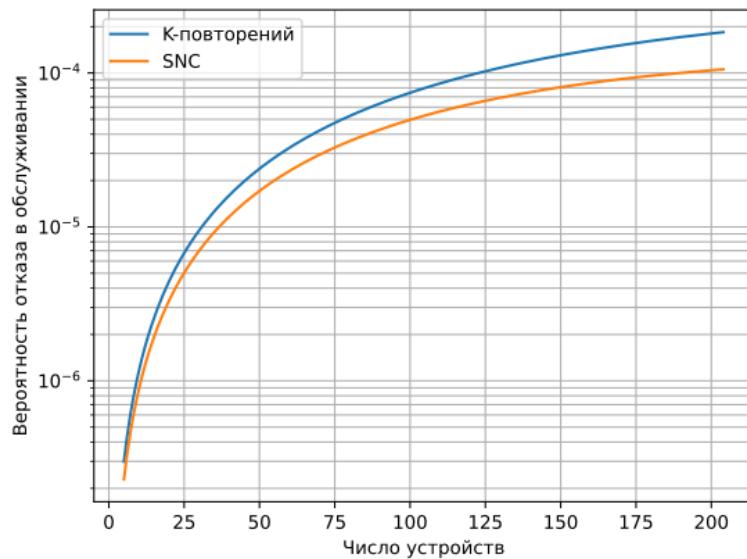


Рисунок 2.9 – График зависимости вероятности неудачной передачи сообщений после k повторных отправок от общего числа устройств в системе

По результатам моделирования можно сделать вывод, что предложенная схема сетевого кодирования может использоваться для достижения двух важных целей – снижение средней задержки передачи сообщения по восходящему каналу связи при сохранении исходной вероятности неудачной передачи или же для увеличения надежности передачи при сохранении средней задержки.

2.5. Метод динамического изменения числа преамбул

Актуальной задачей в высоконагруженных системах с множественным случайным доступом является большое количество коллизий, возникающих при одновременной передаче по каналу связи [37; 81; 89].

Предлагается новый подход, который использует концепцию нумерологии 5G NR для распределения ресурсов. Основная цель – динамически распределить эти слоты между высокоприоритетными URLLC и низкоприоритетными mMTC пользователями на основе заранее определенных критериев. Приоритеты основаны на конкретных потребностях в надежности и задержке каждого типа трафика.

Предполагается, что пакеты, передаваемые устройствами, относительно малы, что позволяет успешно передавать их в течение одного интервала времени в выбранной нумерации. Это предположение справедливо для передачи небольших пакетов в сценарии mMTC. Отметим, что в общем случае подход не ограничен определенным размером пакета.

Распределение интервалов для высокоприоритетного и низкоприоритетного трафика регулируется заранее установленными максимальными и минимальными ограничениями для каждого типа трафика. Эти параметры тщательно откалиброваны на основе всестороннего анализа структуры трафика, как в [9] в течение длительного периода времени, значительно превышающего продолжительность подкадра или кадра в целом.

Настроив максимальное ограничение на количество слотов для

высокоприоритетного трафика таким образом, чтобы оно было меньше общего количества доступных слотов в подкадре, можно гарантировать, что по крайней мере один слот остается доступным для трафика с низким приоритетом.

Процедура динамического изменения числа преамбул для определенного вида трафика состоит из двух основных этапов:

1. Оценка числа активных устройств, участвовавших в передаче во время предыдущего фрейма;

2. Оценка числа новых заявок и активных устройств, появившихся в системе.

Оценка количества тегов может рассматриваться как задача многоклассовой классификации или как задача нелинейной регрессии. Поскольку ошибка предсказания числа устройств будет оцениваться через среднеквадратичное отклонение, а не через точность, методы машинного обучения, основанные на регрессии, показывают лучшие результаты по сравнению с классификаторами [22; 82]. В связи с этим выбран метод регрессии опорных векторов (Support Vector Regression, SVR), которая является разновидностью метода опорных векторов, применяемого для задач регрессии.

Наборы данных для обучения модели являются синтетическими, то есть получены с применением моделирования методом Монте-Карло [58; 60; 76] для получения достаточного количества возможных сценариев, которые могут возникнуть во время процедуры опроса.

Распределение устройств в ячейках также задавалось методом Монте-Карло.

Моделирование было выполнено для фреймов размером $L > 16$. Для каждого размера фрейма и количества тегов были реализованы случайные 100 000 распределений пустых слотов E , успешных слотов S и слотов с конфликтом C .

При малых размерах фрейма точность предсказания снижается из-за большого числа конфликтов при увеличении числа активных устройств. Для решения этой проблемы и дообучения модели была создана дополнительная выборка путем моделирования системы с использованием алгоритма слотовой интервальной ALOHA.

Далее обученная модель использовалась для предсказания числа

активных устройств после конца каждого фрейма. Применив этот этап для обоих типов трафика, получаем оценку числа активных устройств как mMTC так и URLLC (рисунок 2.10).

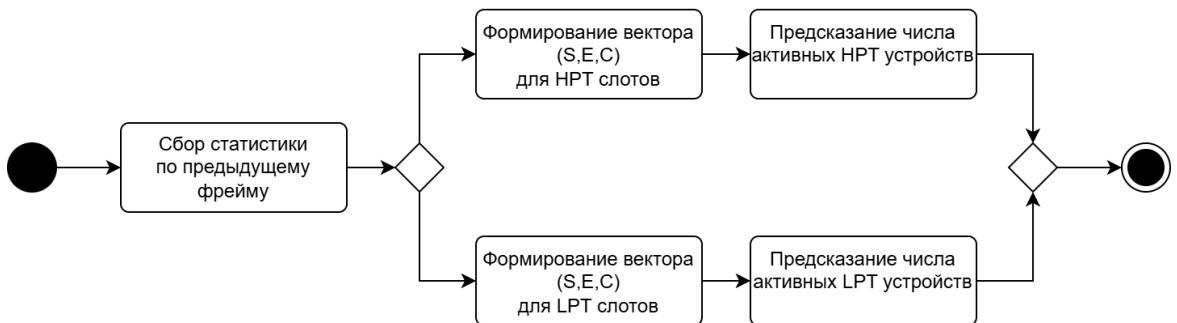


Рисунок 2.10 – UML-диаграмма метода предсказания числа активных устройств в сети

В результате проверки обученной модели на тестовой выборке был рассчитан коэффициент детерминации равный 0.998, что показывает высокую точность предсказаний и обеспечивает эффективность использования данного метода для предсказания числа активных устройств.

Для предсказания числа пришедших в систему устройств предлагается использовать модель на основе рекуррентной нейронной сети. Метод предсказания будет состоять из следующих этапов:

1. Сеть обучается на данных, включающих как равномерный трафик (периодические данные), так и всплески трафика (событийные данные).
2. Для обучения используется алгоритм оптимизации Adam и функция потерь MSE (среднеквадратичная ошибка).
3. В отличие от традиционного метода "rolling", который требует буферизации данных и их обновления при поступлении новых данных, предлагаемый алгоритм использует только свежие данные для обновления состояний сети.
4. Алгоритм сохраняет состояния нейронной сети перед началом предсказаний и обновляет их при поступлении новых данных, что позволяет избежать вычислительных издержек, связанных с обработкой устаревших данных.

На рисунке 2.11 приведена временная диаграмма смоделированного

трафика, а также предсказанного обученной моделью. Для сравнения были выбраны две модели нейронной сети – сеть с долговременной краткосрочной памятью (Long Short-Term Memory, LSTM) сеть и обычная рекуррентная нейронная сеть (Recurrent Neural Networks, RNN).

В таблице 2.4 приведены численные показатели эффективности предсказания с помощью обученных моделей – среднеквадратичное отклонение (Mean Squared Error, MSE) и средневзвешенная процентная ошибка предсказания (Weighted Average Percentage Error, WAPE).

Таблица 2.4 – Численные результаты предсказания на тестовой выборке

	LSTM	RNN
MSE	81.72	90.34
Среднее значение	25	25
WAPE	0.177	0.19

Таким образом исходя из близких результатов на тестовой выборке для дальнейшего исследования была выбрана модель LSTM.

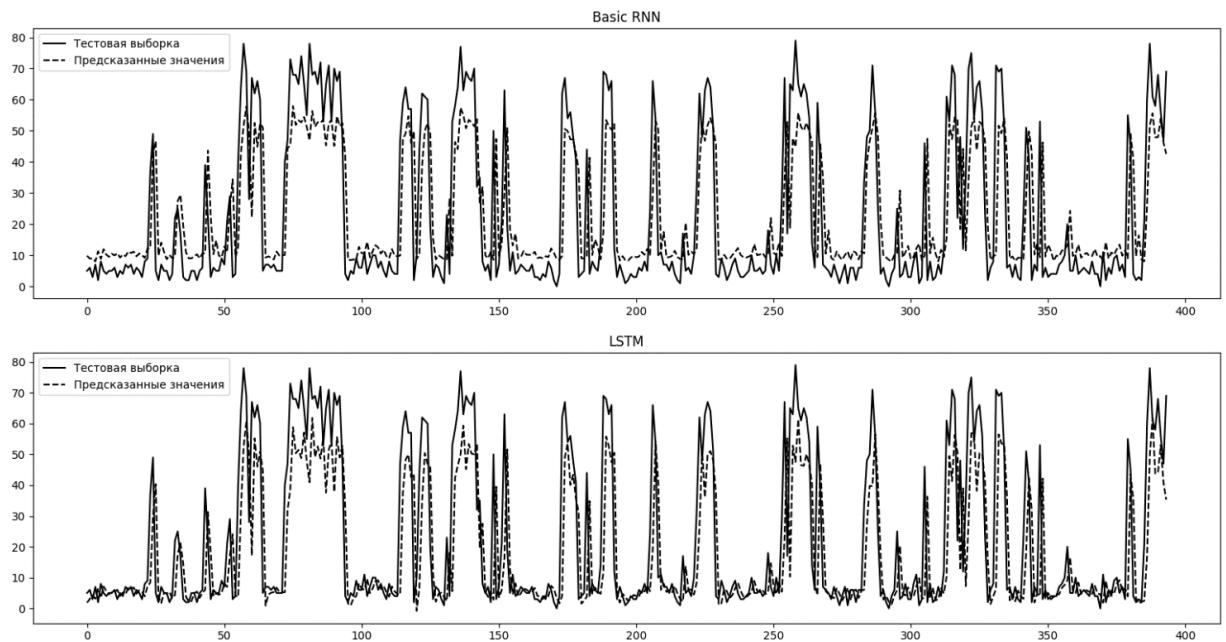


Рисунок 2.11 – Временная диаграмма реального и предсказанного трафика

На основании оценки текущего числа активных устройств и интенсивности входящего трафика базовая станция осуществляет

динамическое перераспределение преамбул между высокоприоритетными и низкоприоритетными устройствами. Ключевая задача алгоритма – оптимально распределить доступные слоты для высокоприоритетных передач, обеспечивая высокую пропускную способность, минимальную задержку доступа и снижение вероятности потерь пакетов. При этом система должна поддерживать базовый уровень обслуживания mMTC трафика, предотвращая его полное вытеснение даже в условиях высокой нагрузки URLLC трафиком.

Для этого вводится общее число преамбул в подкадре – L , определяемое структурой сети. Если m_0 и m_1 – количество слотов, выделенных для URLLC и mMTC соответственно, то выполняется условие $m_0+m_1=L$. Алгоритм ограничивает выделение ресурсов для URLLC, задавая максимальное ($i_{0,\max} < L$) значение числа выделенных преамбул. Верхняя граница гарантирует, что как минимум один слот в каждом подкадре останется для mMTC, предотвращая его полную блокировку.

В процессе работы, базовая станция gNB непрерывно анализирует состояние канала. Для каждого типа трафика x , где $x=0$ соответствует URLLC, а $x=1$ соответствует mMTC в каждом субфрейме t фиксируются следующие показатели:

- число пустых слотов $e_{x,t}$,
- количество успешных передач $s_{x,t}$,
- случаи коллизий $c_{x,t}$.

Очевидно, что сумма $e_{x,t} + s_{x,t} + c_{x,t} = m_{x,t}$, так как каждый слот может находиться только в одном из этих состояний.

Дополнительно gNB оценивает число новых запросов $\lambda_{x,t}$, отражающее количество устройств, сгенерировавших пакеты в текущем подкадре. На основе этих данных система обновляет распределение слотов $m_{x,t}$ для следующего интервала ($t+1$), обеспечивая адаптацию к изменяющимся условиям сети.

Этапы метода динамического изменения числа преамбул:

1. Для значений x , равных 1 и 2, gNB, используя данные, полученные из наблюдений $(e, s, c)_{x,t}$, наряду с расчетным количеством активных устройств в начале субфрейма t , переходит к оценке количества устройств, оставшихся в

системе по завершению этого конкретного субфрейма (обозначается $w_{x,t}$).

2. Для $x=0,1$ gNB, используя модель машинного обучения оценивает количество вновь прибывших сообщений в течение субфрейма t , $\lambda_{x,t}$.

3. Общее количество активных устройств в конце субфрейма t , готовых к передаче в подкадре $(t+1)$, равно сумме $w_{x,t}$ и $\lambda_{x,t}$.

4. Расчет общего необходимого числа преамбул L ;

5. Для обеспечения приоритетности трафика URLLC сначала генерируется $m_{0,t+1} = \min(u_{0,\max}, \max(w_{0,t} + u_{0,\min}))$ преамбул – слотов для передачи, а затем конфигурируются оставшиеся $m_{1,t+1} = L - m_{0,t+1}$ преамбул для mMTC устройств.

На рисунке 2.12 представлена полная схема метода динамического изменения числа преамбул.

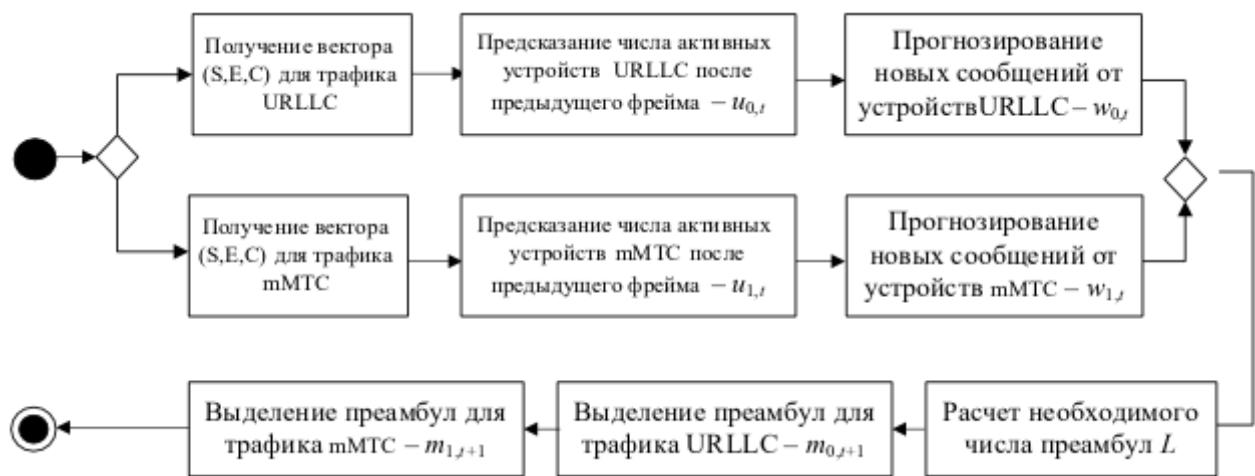


Рисунок 2.12 – UML диаграмма метода динамического изменения числа преамбул

2.6. Модель восходящего канала связи сети 5G

Математическое моделирование восходящего канала сети 5G выполнено имитационным методом. Приведем описание объектов и процессов модели восходящего канала и полученные результаты моделирования.

2.6.1. Описание объектов и процессов модели восходящего канала

Модель восходящего канала, рассматриваемая в работе, представляется системой массового обслуживания (СМО) типа $G|G|1$ с приоритетным обслуживанием (рисунок 2.13).

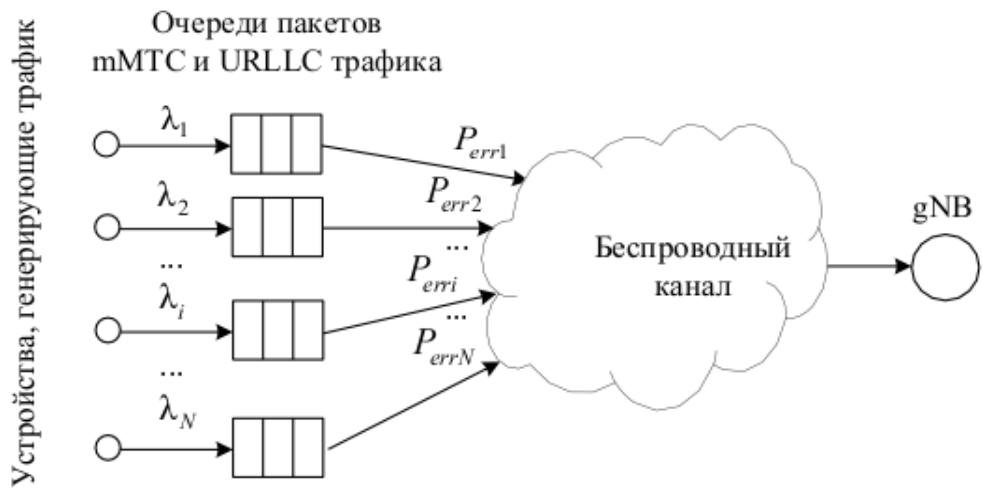


Рисунок 2.13 – Модель восходящего канала связи

Модель представляет собой замкнутую систему, которая адекватно отражает основные аспекты работы восходящего канала сети 5G: конкуренцию за ресурсы, гетерогенность трафика, неидеальность канала и интеллектуальное управление.

Модель – дискретно-событийная, построена по схеме процессов. Все время взаимодействия N устройств с базовой станцией (generalized Node B, gNB) сети 5G разбито на фреймы одинаковой длины, поэтому системное время в модели задано последовательностью фреймов.

В модели одновременно развиваются несколько процессов:

Пользовательское устройство соты, которое задается интенсивностью λ_i , $i = \overline{1, N}$ высокоприоритетного URLLC трафика, либо низкоприоритетного mMTC трафика. Распределение интервалов времени между поступлениями пакетов задано распределением Вейбулла, изменение значения λ_i выполняется согласно модели burst-трафика.

В каждом новом цикле имитации работы восходящего канала выполняются следующие процедуры:

- устройство активно, если у него на момент начала фрейма есть сообщение;
- активным устройством случайным образом, с равномерным распределением, выбирается одна из доступных преамбул, отправленных gNB;
- сообщение кодируется выбранным методом помехоустойчивого кода;
- проверяется условие «Преамбула выбрана только одним устройством?». Если «Да», то сообщение передается на gNB по беспроводному каналу, иначе – пополняется статистика потерянных сообщений;
- проверяется условие «Сообщение искажено ошибками канала связи?». Если «Да», то сообщение пополняет статистику потерянных сообщений, иначе число успешных передач увеличивается на единицу, сообщение удаляется из очереди.

Очередь, в которой находятся необслуженные пакеты каждого пользовательского устройства. Каждое из устройств имеет свою неограниченную очередь сообщений.

Базовая станция, которая управляет ресурсами, динамически перераспределяя преамбулы на основе наблюдаемой статистики. В каждом новом цикле имитации работы восходящего канала выполняются следующие процедуры:

- оценивается текущее число активных устройств каждого типа, используя модель машинного обучения SVR;
- прогнозируется интенсивность поступления новых сообщений на следующий фрейм с помощью рекуррентной нейронной сети LSTM;
- на основе прогнозных оценок вычисляется новое оптимальное распределение преамбул m_0 – для URLLC трафика и m_1 – для mMTC трафика на следующий фрейм с обеспечением минимальных гарантий для каждого класса трафика;
- преамбулы m_0 и m_1 отправляются широковещательной рассылкой.

Беспроводной канал, который задается вероятностью ошибки P_{err} для каждого устройства. Значение P_{err} определяется моделированием полной группы несовместных случайных событий на интервале $[0.1, 0.01]$.

На рисунке 2.14 и 2.15 представлена UML диаграммы процесса моделирования базового сценария и сценария с использованием SNC и метода динамического изменения числа преамбул.

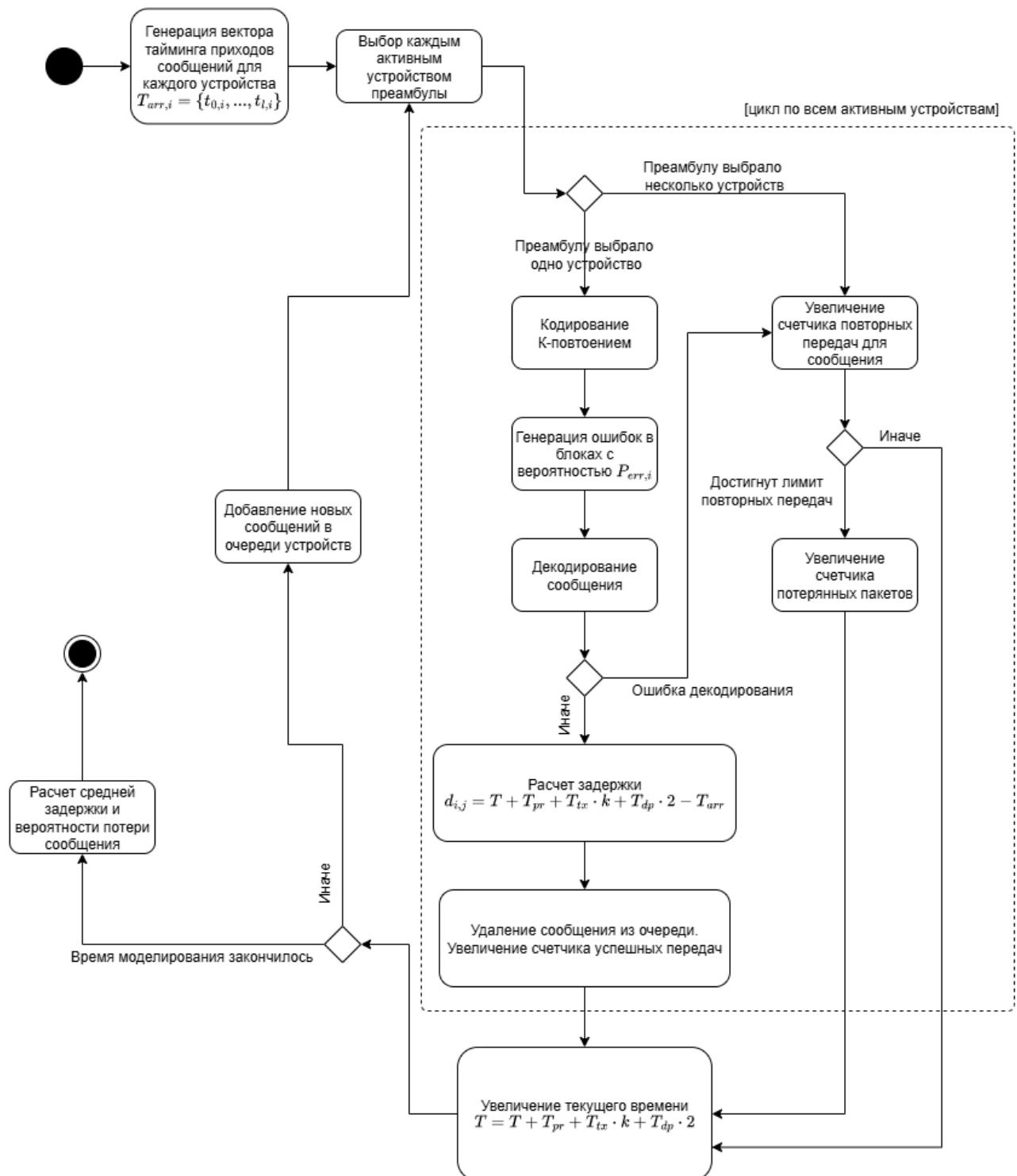


Рисунок 2.14 – UML-диаграмма моделирования базового сценария

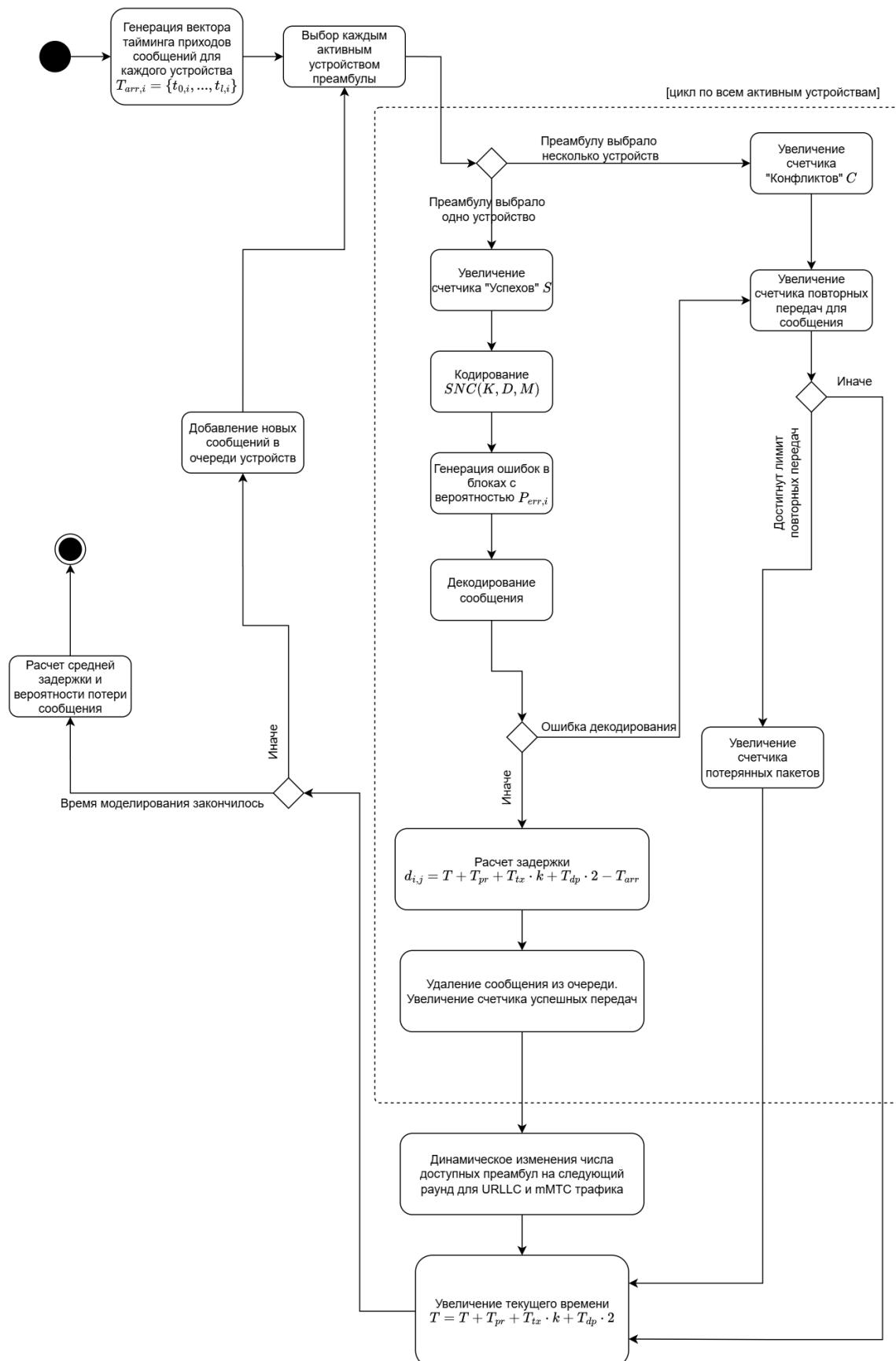


Рисунок 2.15 – UML-диаграмма моделирования предлагаемого метода

2.6.2. Результаты моделирования

Для анализа эффективности предлагаемых методов снижения задержки при передаче трафика по восходящему каналу выполнено имитационное моделирование рассматриваемой системы. Также получены результаты сравнения предложенного метода оценки числа активных устройств с методом из работы [80] для случая 2, при котором число активных устройств в фрейме неизвестно. Обозначим алгоритм, предлагаемый в работе [80] как M^* , а метод предложенный в настоящей диссертации как A^* . Параметры моделирования представлены в таблице 2.5.

Таблица 2.5 – Параметры моделирования

Число устройств, N	5 ... 200
Интенсивность, (1/ TTI)	$\lambda_0=0.001, \lambda_1=0.25, \lambda_{01}=0.1, \lambda_{10}=0.8$
Закон распределения для burst-трафика	W(\bullet), Exp(\bullet)
Доля устройств с высоким приоритетом	40%
Количество пакетов при передаче, k	2
Максимальное количество передач одного сообщения, K	3
Количество блоков для сетевого кодирования, m	4
Вероятность ошибки в канале, P_{err}	0.1 ... 0.01
Время передачи данных, ТТИ	$T_{pr} = 1, T_{tx} = 2, T_{dp} = 1$
Количество преамбул, L	64

Совместное применение скользящего сетевого кодирования и динамического управления позволяет добиться снижения средней задержки передачи сообщения как для URLLC трафика, так и для mMTC трафика. На

на рисунке 2.16 приведены результаты имитационного моделирования передачи трафика по восходящему каналу, из которого видно, что при 100-200 устройствах в сете задержка высокоприоритетного трафика снижается на 15-30%, а низкоприоритетного трафика на 10-20%. Очевидно, что и энергоэффективность улучшается за счет сокращения повторных передач.

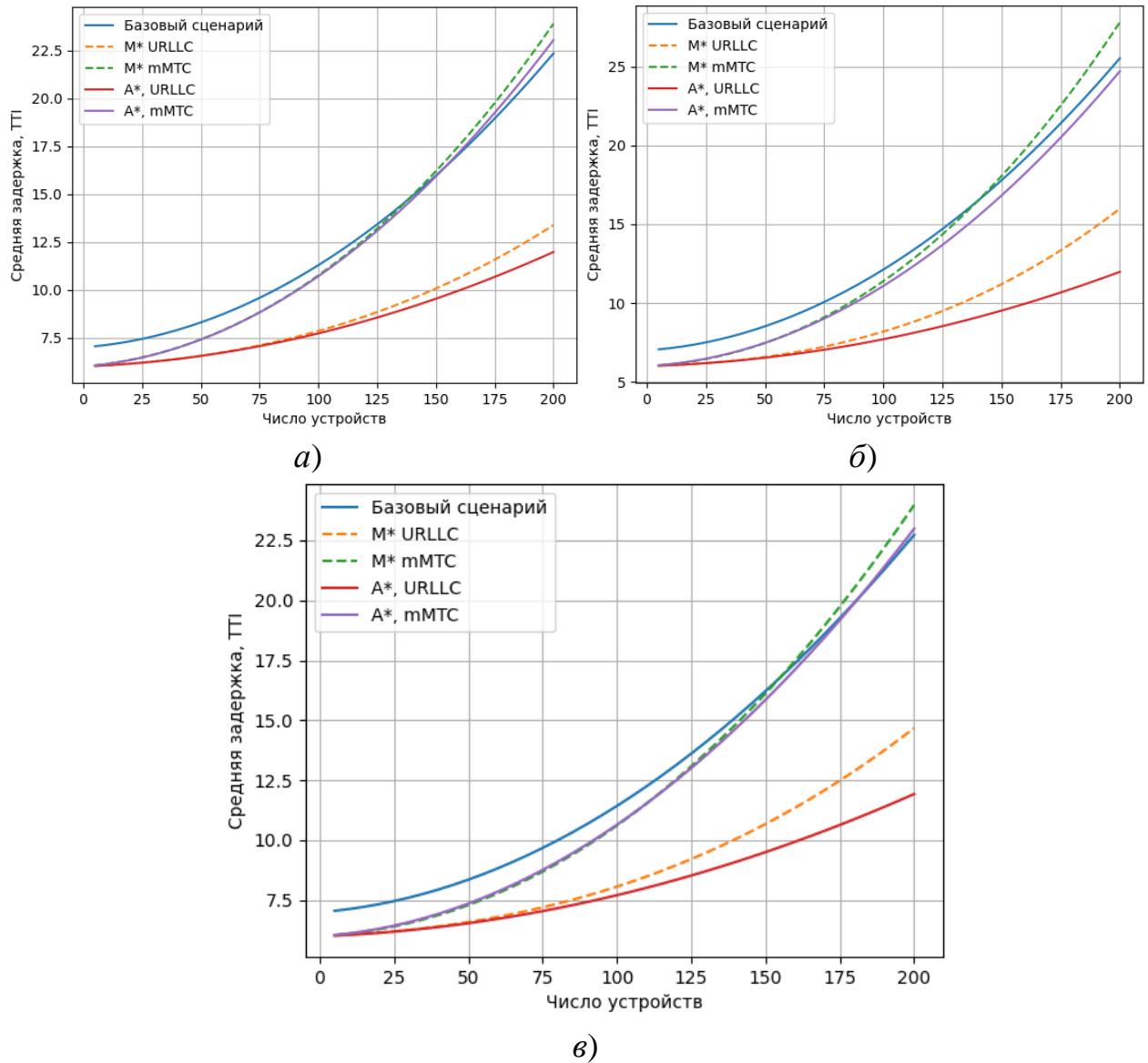


Рисунок 2.16 – График зависимости средней задержки для каждого класса трафика в восходящем канале от числа устройств в системе: *a* – $M|G|1$; *б* – $LN|G|1$; *в* – $W|G|1$

Проигрыш в средней задержке для низкоприоритетного трафика при большом числе устройств обуславливается высоким процентом числа

высокоприоритетных устройств, что влияет на недостаточность ресурсов для передачи низкоприоритетного трафика. Однако в реальных системах процент высокоприоритетного трафика обычно ниже, чем использовался при моделировании.

Выводы по второй главе

Проведен анализ существующих схем передачи данных по восходящему каналу связи в сотовых сетях.

Предложены методы распределения ресурсов в восходящем канале связи, позволяющие динамически менять параметры канала в зависимости от текущей и предсказанной ситуации в нем:

- Метод динамического распределения ресурсов сети между различными видами трафика с применением ML-прогнозирования: прогнозирование числа активных устройств методом SVR-регрессии и предсказания количества новых запросов с применением RNN-сети, обрабатывающей временные ряды трафика.
- Алгоритм скользящего сетевого кодирования, позволяющий декодировать пакеты «на лету» и значительно снизить вероятность ошибок и среднюю задержку передачи сообщений за счет совмещения преимуществ сетевого кодирования и механизма повторных передач.

Результаты моделирования системы показали, что использование предлагаемого метода динамического изменения числа преамбул совместно с алгоритмом скользящего сетевого кодирования позволяют добиться значительного уменьшения средней задержки передачи сообщения.

3. МЕТОД АДАПТИВНОГО МУЛЬТИПЛЕКСИРОВАНИЯ БИТОВЫХ ПОТОКОВ В НИСХОДЯЩЕМ КАНАЛЕ СВЯЗИ

Совместная передача eMBB и URLLC в нисходящем канале требует новых методов мультиплексирования, чтобы избежать деградации качества обслуживания. Традиционные подходы, приводят к потере пропускной способности eMBB. Предлагаемый в этой главе метод основан на решётчатом кодировании и методе «выкалываний» позиций передачи eMBB трафика, что позволяет интегрировать разнородный трафик в один ресурсный блок без потерь.

3.1. Анализ существующих методов мультиплексирования

Одной из актуальных задач при проектировании сетей 5G и последующих поколений является эффективное совмещение в общем радиочастотном ресурсе разнородных типов трафика (мультиплексирование), прежде всего, расширенной мобильной широкополосной связи eMBB и сверхнадежной связи с малой задержкой URLLC [15; 78]. Трафик eMBB предъявляет требования к высокой пиковой скорости передачи данных, в то время как URLLC характеризуется строгими ограничениями по задержке и надежности.

Фундаментальные исследования в данной области сосредоточены на разработке математических моделей мультиплексирования, описывающих влияние трафика URLLC на производительность eMBB, тем самым раскрывая сложность задачи – поиска методов существования трафика eMBB и трафика URLLC в нисходящем канале с выполнением требований QoS. Так, в работах [10; 11; 12; 13; 14] предложены и проанализированы три ключевые модели снижения скорости eMBB при использовании стратегии прокалывания (puncturing) или наложения (overlapping) ресурсов URLLC:

- Линейная модель, в которой потеря пропускной способности eMBB прямо пропорциональна доле занятого мини-слота.

- Выпуклая модель, где потери описываются выпуклой функцией.
- Пороговая модель, предполагающая незначительное влияние URLLC до достижения определенного порога, за которым следует существенная деградация eMBB.

При этом основные методы мультиплексирования можно разделить на два основных типа – ортогональные и неортогональные.

Ортогональные методы предполагают разделение ресурсов в частотной или временной области с выделением отдельных полос или временных интервалов для различных типов трафика. Исторически этот подход получил широкое распространение благодаря своей простоте и надежности. В его основе лежит принцип изоляции сервисов, что исключает взаимное влияние разнородного трафика.

Основным преимуществом ортогональных методов является гарантированное качество обслуживания (QoS) для каждого типа трафика. Однако фундаментальным недостатком такого подхода является статическое распределение ресурсов, не адаптируемое к динамическим изменениям трафика и состояния канала. Как показывают исследования [10; 40; 42; 47; 63], эффективность использования полосы частот в системах с ортогональным разделением редко превышает 60-70%, что делает их малопригодными для сетей 5G/6G с их требованиями к высокой спектральной эффективности.

Более того, жесткое закрепление ресурсов за определенными типами трафика приводит к значительным потерям пропускной способности в условиях неравномерной нагрузки. Например, при временном разделении ресурсов неиспользуемые URLLC-слоты не могут быть перераспределены для передачи eMBB-трафика, даже при наличии очереди на обслуживание [48].

В этой связи, перспективным решением является неортогональное мультиплексирование потоков данных в нисходящем канале связи [21; 41; 43; 44; 72], в рамках которого ресурсы, выделенные для eMBB, могут динамически использоваться для URLLC-трафика.

Неортогональные схемы (Non-Orthogonal Multiple Access, NOMA) теоретически обеспечивают более высокую спектральную эффективность за счет одновременной передачи нескольких потоков в одном частотно-временном ресурсе [53;63]. Различные пользователи разделяются посредством различных уровней мощности, кодовых последовательностей или других признаков.

В контексте мультиплексирования eMBB и URLLC наибольший интерес представляют методы, основанные на различном распределении мощности. URLLC-пользователям, как правило, выделяется большая мощность передачи для обеспечения требуемой надежности, в то время как eMBB-пользователи работают с меньшей мощностью, но используют более высокоуровневые схемы модуляции.

Однако практическая реализация NOMA сталкивается с серьезными проблемами. Во-первых, методы NOMA зависят от точности оценки состояния канала. Даже незначительные ошибки в оценке состояния канала могут привести к значительному ухудшению производительности системы. Во-вторых, процедуры подавления интерференции на приемной стороне обладают высокой вычислительной сложностью, что ограничивает их применение в устройствах с ограниченными ресурсами. В-третьих, координация мощности между разнородными пользователями требует сложных алгоритмов оптимизации, плохо масштабируемых в больших сетях.

Проведенный анализ позволяет сделать вывод о том, что ни один из рассмотренных подходов в отдельности не может эффективно решить задачу совместной передачи eMBB и URLLC трафика в условиях современных требований к производительности беспроводных сетей. Ортогональные методы страдают от низкой спектральной эффективности, методы прокола приводят к значительным потерям пропускной способности eMBB, а неортогональные схемы обладают неприемлемо высокой сложностью реализации.

Таким образом, существует потребность в гибридном подходе, сочетающем преимущества перечисленных методов при минимизации их недостатков. Такой подход должен обеспечивать адаптивное управление ресурсами на основе текущего состояния канала и характеристик трафика,

управление избыточностью для различных типов сервисов и эффективные механизмы совместной обработки разнородных потоков.

3.2. Формулировка задачи мультиплексирования гетерогенного трафика в нисходящем канале сети 5G

Рассмотрим ситуацию одновременной передачи двух классов трафика – eMBB и URLLC для разных получателей от базовой станции (рисунок 3.1). Так как трафик mMTC передается преимущественно по восходящему каналу связи, то не будем его учитывать ввиду малого влияния на загруженность нисходящего канала и качества обслуживания пользователей.

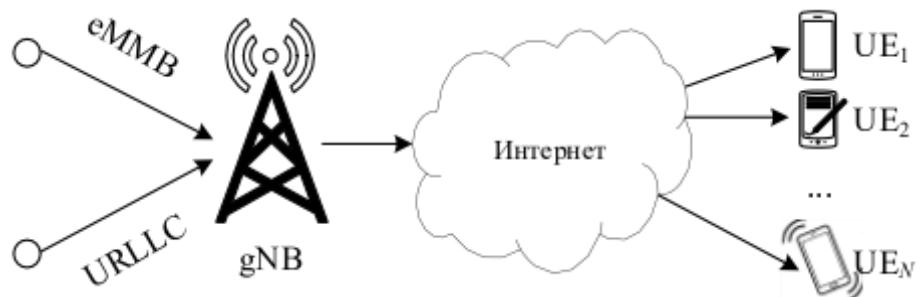


Рисунок 3.1 – Пояснения к постановке задачи мультиплексирования трафика в нисходящем канале сети 5G

Для каждого i -го устройства сотовой сети, на базовой станцией сотовой сети 5G генерируется трафик eMBB с интенсивностью $\lambda_{i,e}$ и трафик URLLC с интенсивностью $\lambda_{i,u}$.

Для строгой формулировки решаемой задачи введем следующие обозначения и параметры системы.

Обозначим:

- R_e^i – средняя скорость передачи данных для i -го пользователя eMBB;
- D_u^i – средняя задержка передачи пакетов для i -го пользователя URLLC;
- P_{loss}^i – вероятность потери пакетов для i -го пользователя URLLC;

- B – доступная полоса частот;
- P_{total} – максимальная доступная мощность передачи;
- γ_i – SNR для i -го пользователя.

Задача оптимизации параметров мультиплексирования формулируется следующим образом:

$$\sum_{i=1}^N R_e^i \rightarrow \max,$$

при условии:

$$\begin{aligned} D_u^i(\theta) &\leq D_{\max}^u, \quad \forall i = \overline{1, N}, \\ P_{\text{loss}}^i(\theta) &\leq P_{\max}^{\text{loss}}, \quad \forall i = \overline{1, N}, \\ \sum_{i=1}^N P_e^i + \sum_{i=1}^N P_u^i &\leq P_{\text{total}}, \\ \theta &\in \Theta, \end{aligned}$$

где θ – вектор параметров алгоритма мультиплексирования;

Θ – допустимые множества параметров;

D_{\max}^u – максимально допустимая задержка для URLLC- пользователей;

P_{\max}^{loss} – максимально допустима вероятность потери пакета URLLC.

Данная постановка задачи отражает компромиссный характер проблемы – максимизация суммарной пропускной способности eMBB-пользователей при соблюдении строгих ограничений по задержке и надежности для URLLC-пользователей.

Все время работы распределяется на интервалы равной продолжительности (слоты), в пределах интервала ограниченная частотная зона может быть распределена планировщиком базовой станции между несколькими устройствами.

При распределении ресурсов канала планировщиками используется выбранная схема мультиплексирования, после чего информационные биты

разных потоков располагаются в объединенном ресурсном блоке для дальнейшей передачи по каналу (рисунок 3.2).

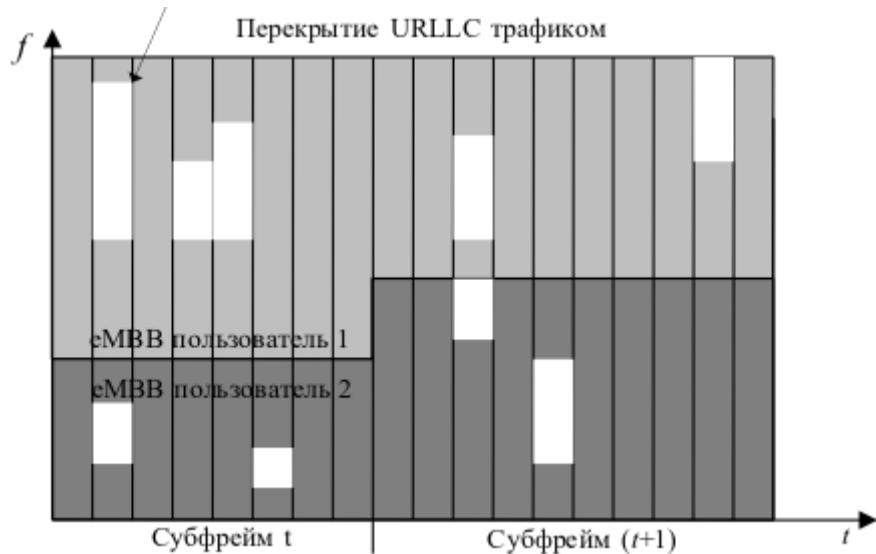


Рисунок 3.2 – Результат работы алгоритма мультиплексирования

3.3. Описание схемы адаптивного мультиплексирования потоков в исходящем канале

Известно, что схема Унгербоека [25; 26; 66] сочетает кодирование и модуляцию за счет увеличения евклидова расстояния между кодовыми словами и поглощает биты четности без расширения полосы пропускания за счет удвоения количества точек в созвездии за счет увеличения количества битов/символов на единицу. Такая конструкция в совокупности оптимизирует как кодирование канала, так и модуляцию, что, опять же, приводит к значительному улучшению качества кодирования без какого-либо расширения полосы пропускания. На основе этих идей был разработан метод адаптивного мультиплексирования потоков [52;54;85;90]. Архитектура предложенной системы адаптивного мультиплексирования состоит из следующих основных компонентов:

- Модуль анализа состояния канала – осуществляет оценку SNR, уровня интерференции и других параметров канала для каждого пользователя;

- Классификатор условий передачи – относит текущие условия к одной из трех категорий на основе заданных пороговых значений;
- Адаптивный мультиплексор – формирует совместный поток данных на основе выбранной схемы мультиплексирования;
- Совместный кодер LDPC-TCM – реализует предложенную схему кодирования и модуляции.

При этом обратная связь обеспечивает адаптацию параметров системы в реальном времени.

Рассмотрим пример работы предлагаемого метода мультиплексирвоания. В определенный момент времени на этой станции появляется сообщение eMBB и несколько сообщений URLLC. Алгоритм объединения и передачи сообщений следующий (рисунок 3.3):

1. Ко всем сообщениям добавляется контрольная сумма
2. Сообщения передаются в модуль кодирования LDPC. Скорость кодирования для трафика eMBB – 8/9, а для трафика URLLC – 1/2. Это обусловлено тем, что сообщения URLLC короче и более требовательны к надежности передачи.
3. Затем базовая станция проводит анализ состояния канала для каждого пользователя, основываясь на предыдущих передачах.
4. Исходя из состояния канала для каждого пользователя выбирается собственная схема мультиплексирования, в которой вартируются как количество бит eMBB потока, так и URLLC потока.
5. После сообщения объединяются в одно с помощью перемежителя (рисунок 3.4).
6. Общее сообщение подается на вход модулятора с решетчатым кодом (TCM), который выбирает размерность модуляции, необходимую для передачи символов $(l + r)$, где l – длина сообщения, r – избыточность применяемого сверточного кода.
7. Сообщение передается по каналу связи и поступает на вход декодера, который представляет собой алгоритм Витерби.

8. Сообщение квантуется в соответствии со следующим правилом: сохраняются наиболее значимые биты $\log_2(Q)$ LLR, где Q – количество уровней квантования.

9. Кодовые слова каждого потока передаются деинтерлейверу для разъединения бит сообщений.

10. Разъединенные сообщения проходят этап LDPC-декодера.

11. Проверяется контрольная сумма и принимается решение о корректности полученного сообщения.

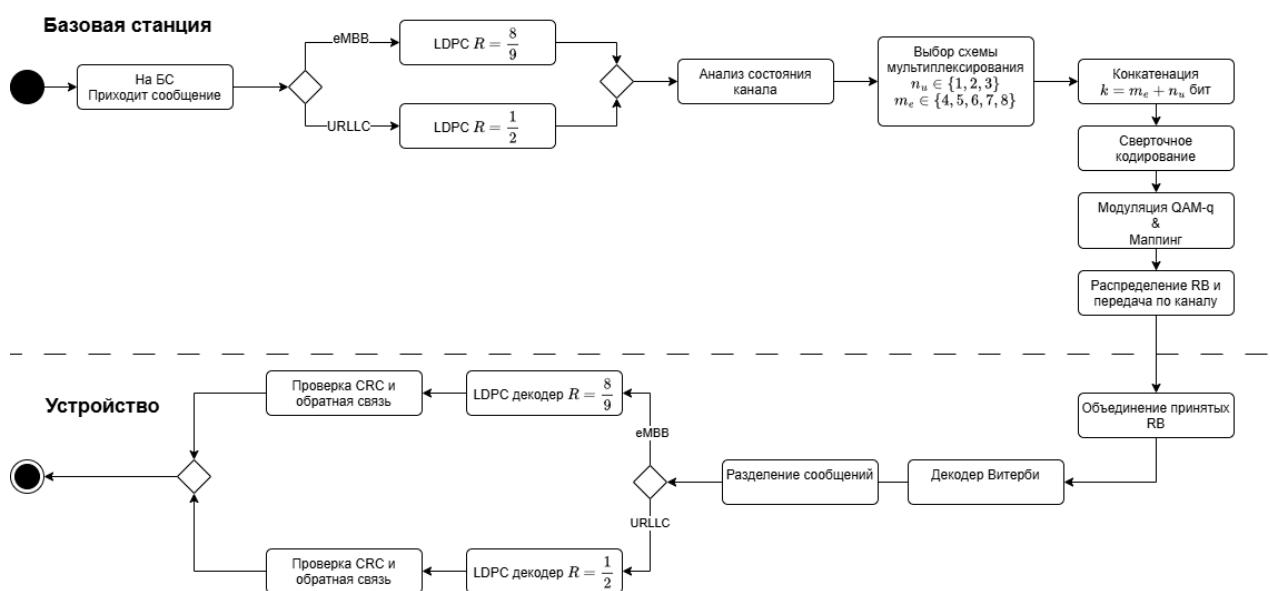


Рисунок 3.3 – UML диаграмма метода адаптивного мультиплексирования
ПОТОКОВ

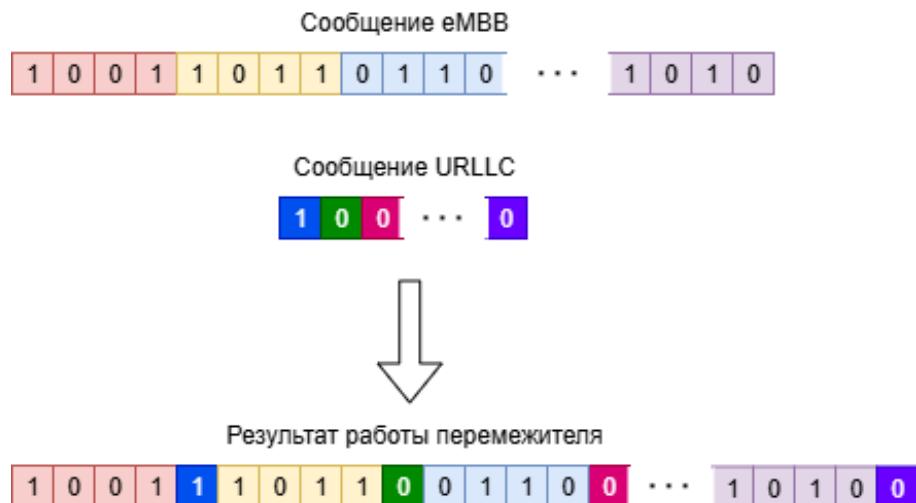


Рисунок 3.4 – Пример работы перемежителя сообщений

На рисунке 3.5 подробно рассмотрен процесс объединения битов сообщений с использование сверточного кодера. Он состоит из следующих этапов:

1. Перемежение битов сообщений eMBB и URLLC;
2. Свёрточное кодирование со скоростью кода $R = 1/2$;
3. Прореживание до скорости $(k-1)/k$;
4. Полученное кодовое слово модулируется и распределяется планировщиком как OFDM символ.

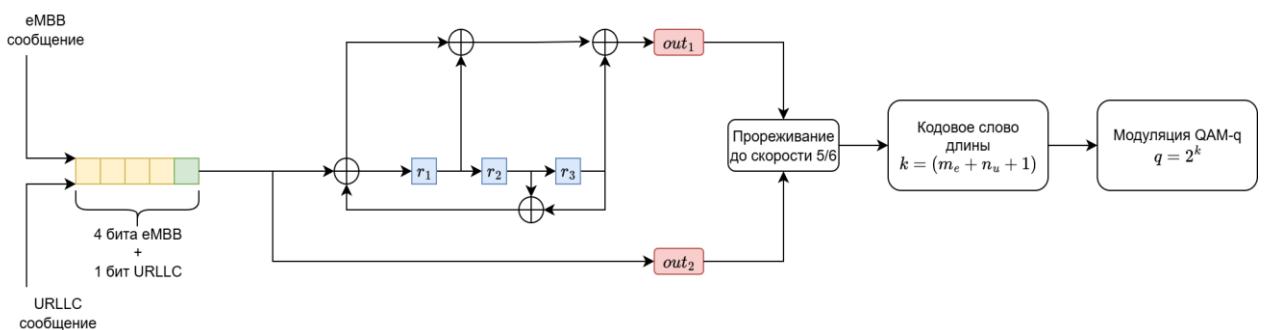


Рисунок 3.5 – Пример работы сверточного кодера

3.4. Модель нисходящего канала

Математическое моделирование нисходящего канала сети 5G выполнено имитационным методом. Приведем описание объектов и процессов модели нисходящего канала и полученные результаты моделирования.

3.4.1. Описание объектов и процессов модели нисходящего канала

Модель нисходящего канала, рассматриваемая в работе, представляется системой массового обслуживания типа $G|G|1$ с приоритетным обслуживанием (рисунок 3.6).

Математическое моделирование нисходящего канала сети 5G выполнено имитационным методом. Модель – дискретно-событийная, построена по схеме событий, где каждое новое событие начиналось и заканчивалось субфреймом длительностью 1 мс, что соответствует нумерологии 3 в 5G NR (New

Radio) глобального стандарта для эфирного интерфейса сетей 5G. Центральное место в модели отведено реализации алгоритмов мультиплексирования, основанного на комбинации LDPC-кодирования, решетчатого кодирования и адаптивной модуляции.

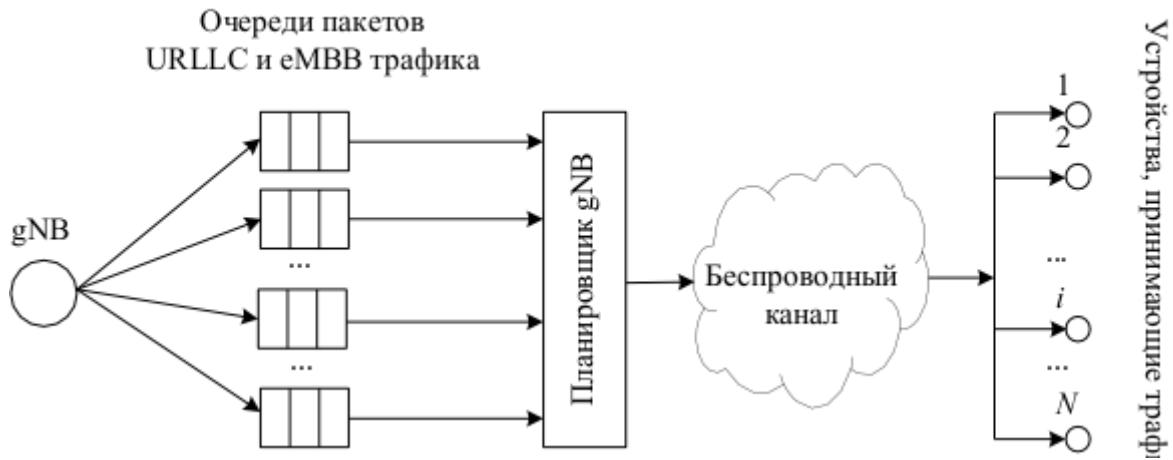


Рисунок 3.6 – Модель нисходящего канала связи

Имитационная модель нисходящего канала состоит из следующих объектов:

Базовая станция gNB, которая обрабатывает два типа сообщений от множества пользовательских устройств. В каждом новом цикле имитации работы исходящего канала выполняются следующие процедуры:

- для каждого i -го устройства генерируются два независимых простейших потока заявок: eMBB трафик с интенсивностью $\lambda_{i,e}$ и URLLC трафик с интенсивностью $\lambda_{i,u}$. Длина j -го eMBB сообщения \bar{V}_{ij} задается Бета-распределением вероятностей с математическим ожиданием 1024 байта из интервалом [512, 1536], что соответствует реальной природе объемного трафика, такого как видео или файлы. Размеры всех URLLC сообщений предполагались фиксированными и равными 32 байтами;
- каждое eMBB сообщение и URLLC сообщение снабжается контрольной суммой (CRC-8);

- каждое eMBB сообщение и URLLC сообщение независимо кодируется LDPC-кодами с разными скоростями: высокая скорость (8/9) для eMBB для максимизации пропускной способности и низкая (1/2) для URLLC для обеспечения сверхнадежности;
- в начале каждого слота gNB, основываясь на текущем состоянии системы: заполненности очередей, приоритетах и состоянии канала, принимает решение о выборе схеме мультиплексирования;
- закодированные битовые последовательности двух типов трафика поступают в блок адаптивного мультиплексирования. На основе выбранной схемы (1, 2 или 3 бита URLLC на OFDM-символ) производится их объединение с помощью перемежения;
- результирующий составной битовый поток подвергается решетчатому кодированию (trellis coded modulation, TCM – метод совместного кодирования и модуляции) с выбором размерности модуляции, необходимой для передачи символов $(l + r)$, где l – длина сообщения, r – избыточность применяемого сверточного кода;
- сформированные OFDM-символы размещаются в ресурсных блоках согласно решению планировщика и отправляются на пользовательские устройства.

Очередь, которая выделена на gNB для каждого пользовательского устройства. Очередь имеет неограниченной объем для каждого класса трафика. Это допущение моделирует сценарий с высокой нагрузкой, где потери происходят не из-за переполнения буфера, а из-за нарушения задержек.

Пользовательское устройство соты, которое выполняет функцию приемника.

В каждом новом цикле имитации работы восходящего канала выполняются следующие процедуры:

- на вход декодера, который представляет собой мягкий выход алгоритма Витерби поступает объединенный битовый поток данных;

- сообщение квантуется в соответствии со следующим правилом: сохраняются наиболее значимые биты $\log_2(Q)$ LLR, где Q – количество уровней квантования, LLR – Log Likelihood Ratio – статистическая мера, основанная на функции логарифмического правдоподобия.
- кодовые слова каждого класса трафика передаются деинтерлейверу для разъединения бит сообщений.
- разъединенные сообщения проходят этап LDPC-декодера;
- проверяется контрольная сумма и принимается решение о корректности полученного сообщения.

Важным допущением имитационной модели, упрощающим анализ, но сохраняющим общность, было предположение о постоянном состоянии канала для всех устройств, характеризуемом фиксированным отношением сигнал-шум (SNR). При моделировании использовалось допущение, что все устройства равномерно распределены по зоне действия базовой станции, а их SNR обратно пропорционален расстоянию от базовой станции.

В дополнение к основному допущению в модели исходящего канала связи приняты следующие допущения:

1. Информация о состоянии канала всегда доступна отправителю;
2. Рассматривается канал с аддитивным белым гауссовым шумом и случайной фазой;
3. На принимающей стороне известно местоположение передачи URLLC-трафика для пользователя;
4. Пользователи eMBB не знают о существовании URLLC-трафика и не декодируют его, то есть приемная сторона моделировалась с учетом асимметрии декодирования: предполагалось, что UE, получающее eMBB-трафик, не детектирует и не декодирует вкрапления URLLC, предназначенные другому абоненту, и наоборот.

На рисунках 3.7 и 3.8 представлены UML-диаграммы процесса моделирования базового сценария и сценария с мультиплексированием потоков.

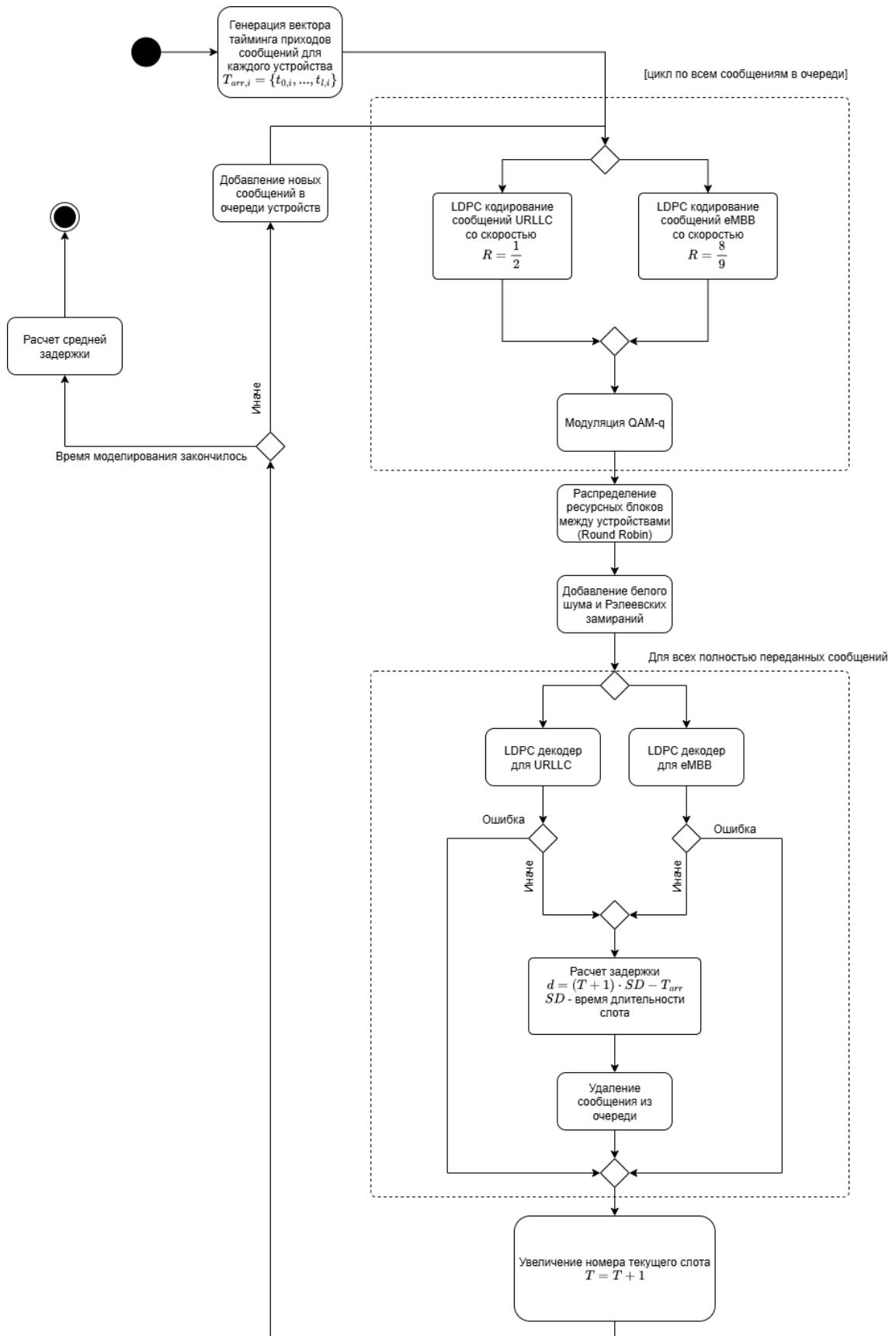


Рисунок 3.7 – UML-диаграмма моделирования базового сценария

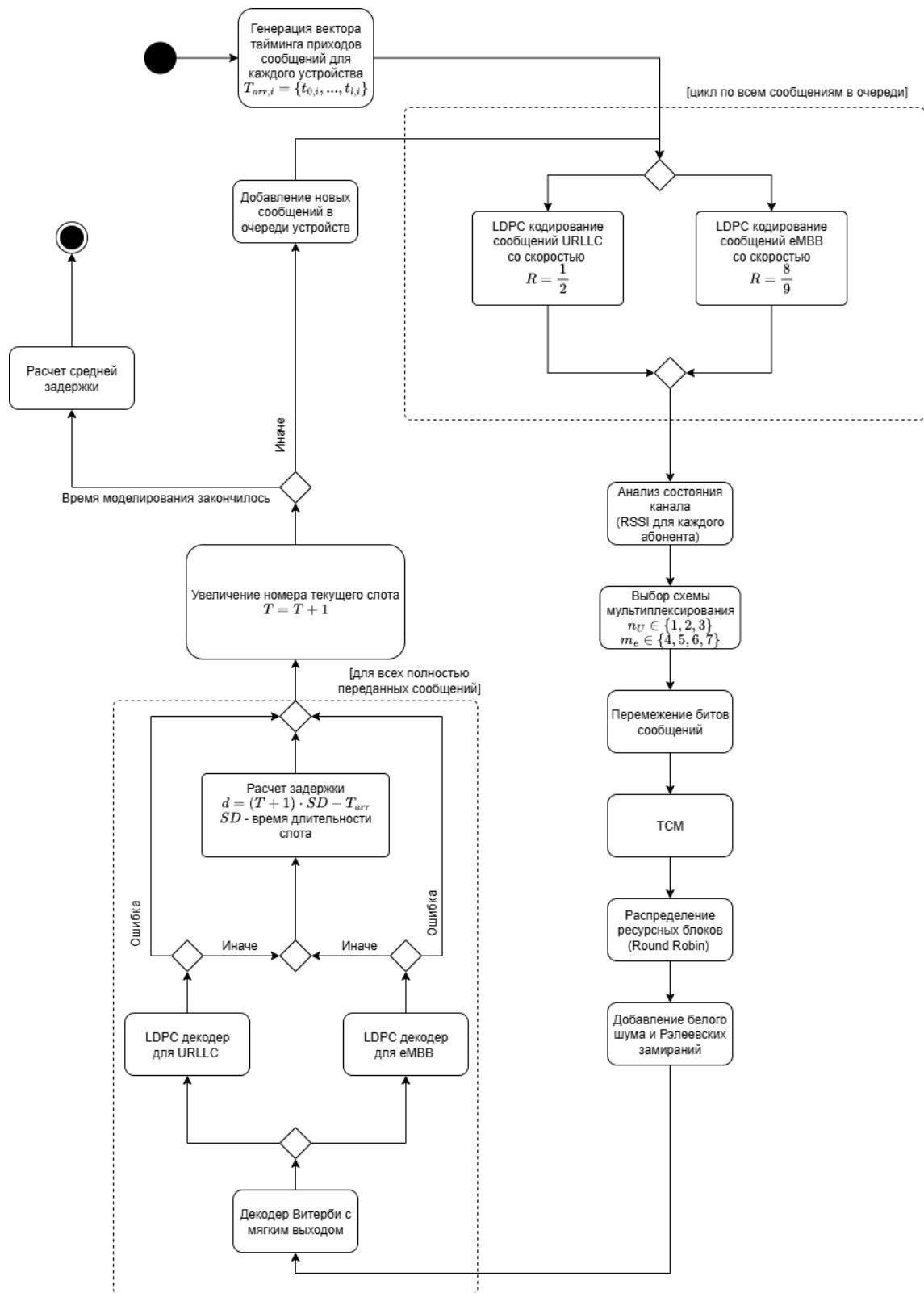


Рисунок 3.8 – UML-диаграмма моделирования сценария с мультиплексированием

3.4.2. Результаты моделирования

Так как подразумевается использования сценария передачи на миллиметровых волнах, то при моделировании рассматривалась сотовая сеть с полосой пропускания 50 МГц и 4 нумерологией стандарта 5G.

Основные параметры моделирования описаны в таблице 3.1

Таблица 3.1 – Параметры моделирования

Параметры моделирования	eMBB	URLLC
Длина сообщений	1024 байт	32 байт
Интенсивность сообщений для каждого UE (1/сек)	1	20
LDPC, скорость кода	8/9	1/2
TCM, скорость кода	5/6, 6/7, 7/8	
TCM, память кодера	3	
Модуляция	QAM-16 (64,128,256)	
Число RB в слоте	65	
Количество слотов в субфрейме	8	

Используя параметры, описанные выше, были получены графики зависимости средней задержки передачи сообщений eMBB и URLLC для различных схем мультиплексирования и различных законов распределения входного потока заявок (рисунки 3.9-3.11).

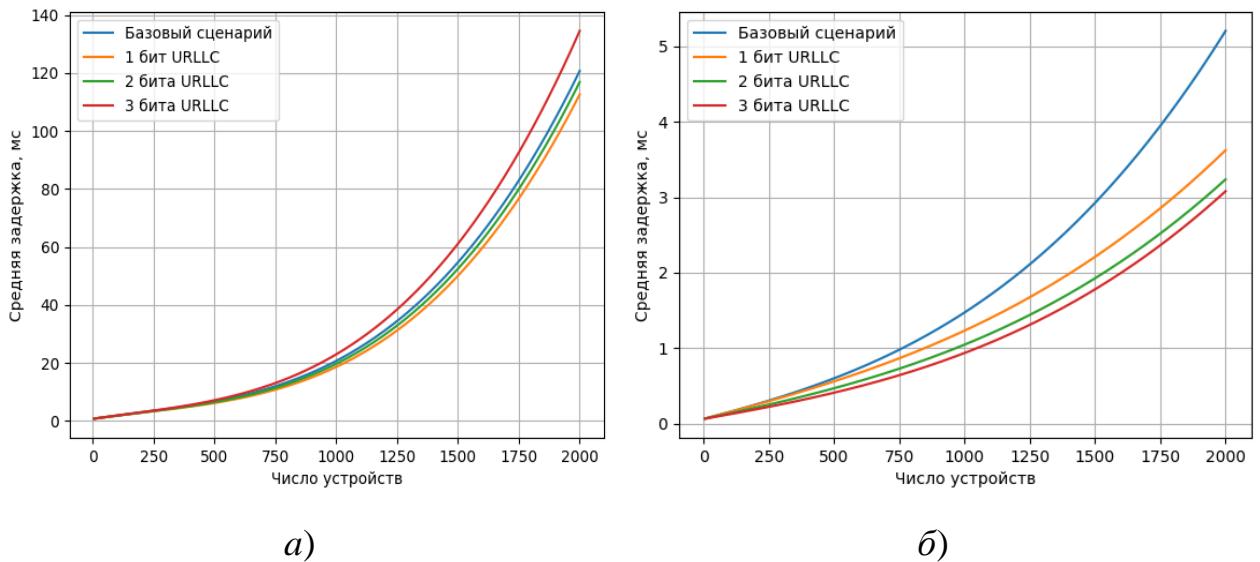


Рисунок 3.9 – График средней задержки передачи сообщения для СМО E|G|1:
 а – для eMBB потока; б – для URLLC потока

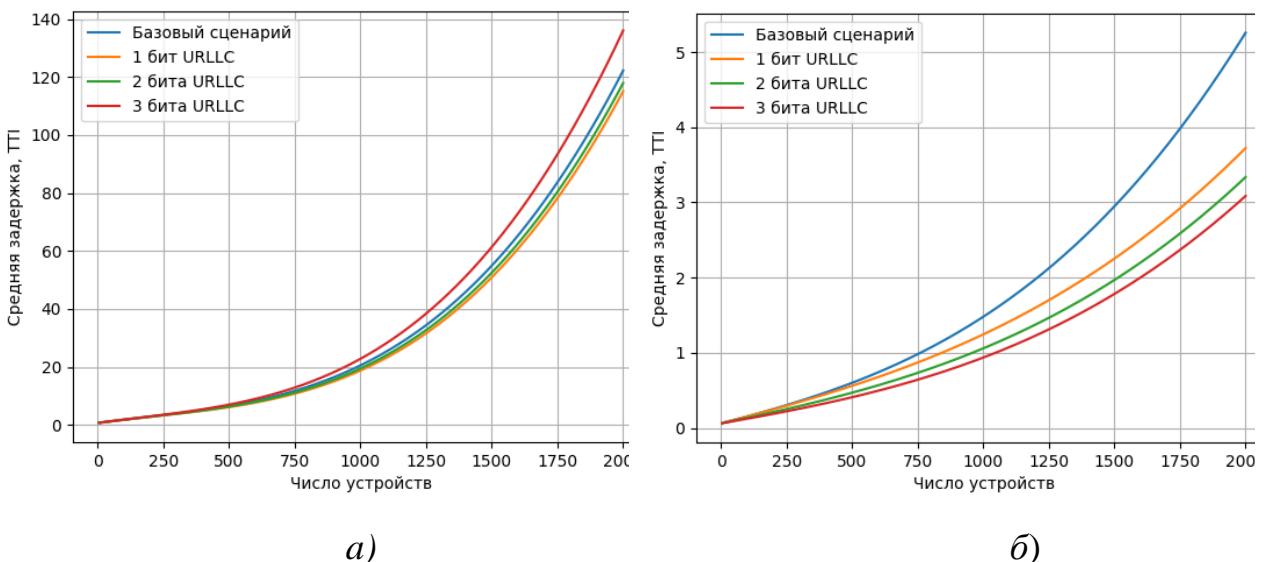


Рисунок 3.10 – График средней задержки передачи сообщения для СМО V|G|1:
 а – для eMBB потока; б – для URLLC потока

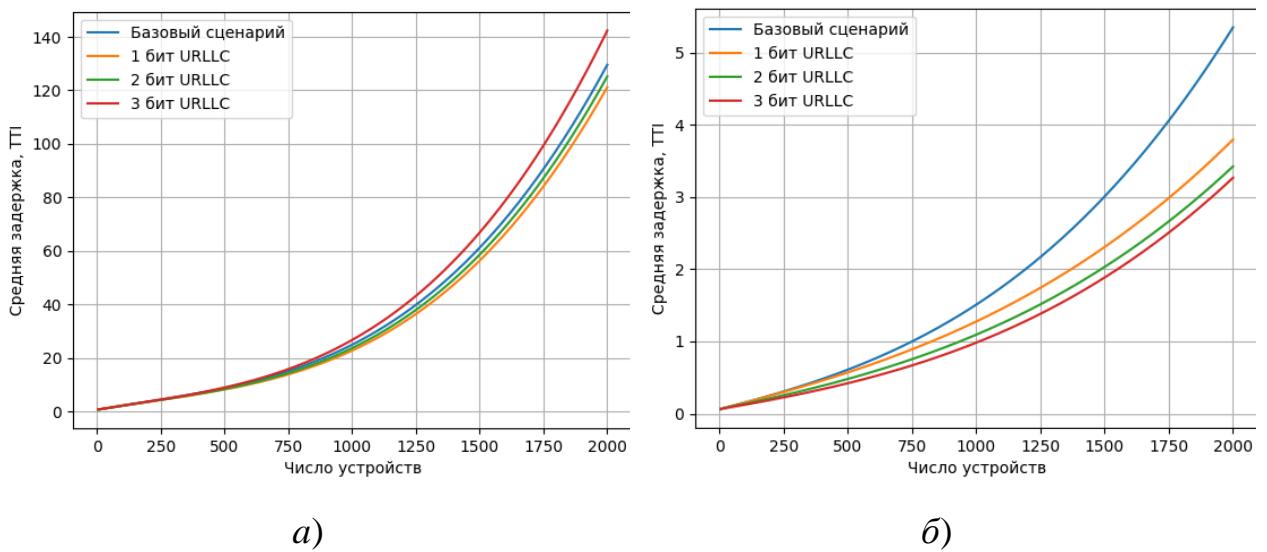


Рисунок 3.11 – График средней задержки передачи сообщения для СМО LN|G|1:
 а – для eMBB потока; б – для URLLC потока

Анализ зависимостей на рисунках 3.9-3.11 показывает, что благодаря использованию предлагаемой схемы мультиплексирования, средняя задержка передачи URLLC сообщения снижается для всех рассматриваемых схем мультиплексирования. При этом использование схемы мультиплексирования с 3-мя битами URLLC сообщения, средняя задержка передачи eMBB сообщения несколько выше, чем без использования мультиплексирования. Это обусловлено увеличение вероятности ошибки в канале из-за сильного увеличения размерности сигнального множества после объединения сообщений. Таким образом предлагаемый метод мультиплексирования с точки зрения потерь для eMBB трафика можно отнести к классу «пороговых».

Поскольку поток URLLC должен соответствовать требованиям немедленной передачи, будем предполагать, что если сообщение не было передано за субфрейм длиной 1 мс, то оно теряет свою актуальность для конечного пользователя. Для проверки соответствия этому требованию проведем моделирование по следующему алгоритму:

1. Набор сообщений генерируется в соответствии с распределением Пуассона

2. Сгенерированные сообщения занимают интервалы в кадре в соответствии с моделируемым сценарием (1,2 или 3 бита на символ OFDM)

3. Сообщения, которые не получили свободного интервала или были размещены не полностью, отбрасываются и пополняют буфер сообщений, передача которых была запрещена.

4. На основе количества отброшенных сообщений и общего количества вычисляется вероятность того, что сообщение не будет передано в течение 1 подкадра.

На рисунке 3.12 приведен график вероятности непередачи URLLC-сообщений за 1 субфрейм. График по оси абсцисс показывает количество URLLC-сообщений, поступающих за 1 мс. По оси ординат показана вероятность того, что полученное сообщение не будет передано по каналу в течение 1 мс из-за скопившейся очереди на базовой станции.

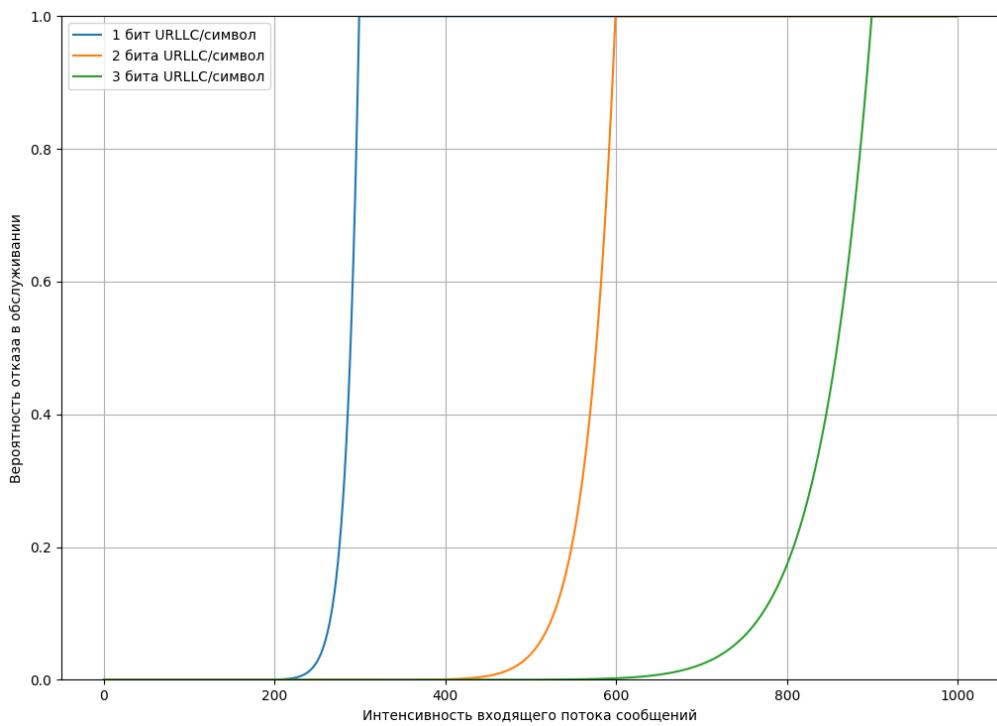


Рисунок 3.12— График вероятности отказа в обслуживании для URLLC сообщения

Основываясь на рисунке 3.12, можем заключить, что для одного из вариантов мультиплексирования, рассмотренного ранее (1 бит URLLC на символ), вероятность того, что по крайней мере одно сообщение не будет

передано за время, меньшее 1мс при интенсивности входного потока сообщений 200 сообщений за 1 субфрейм, составляет менее 10^{-5} . Для сценариев с 2-мя и 3-мя битами URLLC на символ мультиплексирование позволяет передавать до 400 и 500 сообщений URLLC в одном субфрейме с минимальной вероятностью отказа. Таким образом, эти параметры гарантируют срочность передачи, достаточную для использования в системах интернет-связи в промышленности.

Выводы по третьей главе

Предложен метод для адаптивного мультиплексирования потоков данных в нисходящей линии связи, соответствующий пороговой модели, предложенной Вецианой.

Предложен алгоритм кодирования мультиплексированного потока, основанный на методе совместного кодирования и модуляции. Результаты имитационного моделирования показывают, что предложенный алгоритм обеспечивает меньшую среднюю задержку передачи сообщения по нисходящему каналу связи как для eMBB так и для URLLC трафика, по сравнению с базовым сценарием без мультиплексирования.

Получены зависимости пропускной способности канала для потока URLLC, из которых видно, что данная схема позволяет обеспечить вероятность отправки URLLC-сообщения в одном мини-слоте более чем на 99.999% при скорости поступления до 200 сообщений на субфрейм при использовании схемы мультиплексирования с 1 битом URLLC.

4. МЕТОД ДИНАМИЧЕСКОЙ АССОЦИАЦИИ ПОЛЬЗОВАТЕЛЕЙ К БАЗОВЫМ СТАНЦИЯМ

Традиционная ассоциация устройств на основе показателя мощности принимаемого сигнала (Received Signal Strength Indicator, RSSI) приводит к неравномерной нагрузке на базовые станции. В данной главе приведено описание предлагаемого метода динамической ассоциации пользователей к базовым станциям. Метод использует искусственный нейронные сети рекуррентного типа с долговременной памятью (Long Short-Term Memory, LSTM) для прогнозирования трафика и оптимизации ассоциации на основе асимптотической нижней границы задержки.

4.1. Анализ существующих решений ассоциации пользователей к базовым станциям

Современные гетерогенные сети беспроводной связи 5G/6G характеризуются исключительной сложностью архитектуры, объединяющей разнородные элементы: макросоты, микросоты, пикосоты и фемтосоты, работающие в различных частотных диапазонах и обладающие разными характеристиками покрытия и пропускной способности. В таких условиях традиционные методы ассоциации пользователей с базовыми станциями, основанные исключительно на критерии максимального уровня принимаемого сигнала RSSI, демонстрируют свою несостоительность с точки зрения QoS. Причина объясняется тем, что мощные макросоты привлекают чрезмерное количество пользователей просто в силу своей большей мощности передачи, в то время как малые соты с меньшей мощностью, но часто с большей доступной пропускной способностью, остаются недогруженными [62;83;84].

С учетом гетерогенности сотовой сети важнейшим условием оптимизации работы сети 5G является задача балансировки нагрузки, которая предполагает оптимальное распределение пользователей между базовыми

станциями. Передача обслуживания является важной функцией 5G. Когда мобильное устройство перемещается с одной ячейки на другую во время подключения, передача обслуживания позволяет устройству оставаться на связи. Сеть обеспечивает передачу соответствующей информации с исходной ячейки на новую.

В режиме подключения, пользовательское оборудование UE регулярно измеряет качество канала с соседними ячейками и передает эту информацию в сеть. Затем сеть решает, следует ли перенести UE в новую ячейку – выполнить процесс хэндовера. Существуют различные типы хэндоверов в зависимости от типа исходной и целевой ячеек, а также от поддержки конкретных сетевых интерфейсов. Хорошая стратегия хэндовера эффективно использует радио- и сетевые ресурсы, сводя к минимуму перебои в обслуживании в процессе хэндовера.

Вообще, ассоциация пользователей с базовыми станциями может быть проведена по нескольким фундаментальным критериям:

- используемой метрике
 - RSSI, дБ – мощность сигнала, поступающего на антенные устройства;
 - SINR, дБ – Signal to Interference + Noise Ratio – отношение полезного сигнала к помехам от посторонних источников;
 - пропускная способность, бит/с – максимальная скорость передачи данных по заданному каналу связи.
- степени централизации принятия решений
 - централизованные;
 - распределенные;
 - гибридные;
- учету временной динамики
 - статические;
 - адаптивные;

- предиктивные.

Наиболее распространенным историческим подходом является ассоциация на основе максимального RSSI, которая формально может быть описана следующим образом:

$$gNB^* = \arg \max_{j \in M} P_j h_j^2, \quad (4.1)$$

где P_j – мощность j -й базовой станции;

h_j – коэффициент затухания в канале.

Такой подход, основанный на минимизации расстояния между пользователями и базовыми станциями, приводящий к разделению зоны покрытия по методу Вороного (рисунок 4.1), может оказаться крайне неэффективным при неравномерном размещении различных базовых станций, характеризующихся различными уровнями мощности передачи и конфигурациями антенн [77; 87].

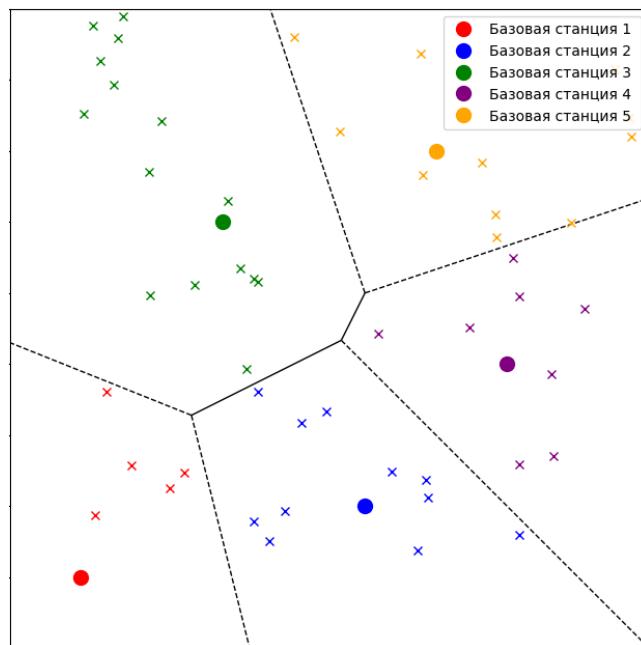


Рисунок 4.1 – Распределение пользователей по базовым станциям по методу Вороного

Основным недостатком данного подхода является полное игнорирование фактора загрузки базовых станций. Пользователи преимущественно ассоциируются с наиболее мощными станциями, создавая перегрузку даже в случаях, когда доступны менее загруженные альтернативы с лучшими условиями передачи.

Более совершенным методом является ассоциация на основе отношения сигнал-помеха плюс шум SINR. Этот подход учитывает интерференционную обстановку в сети и может быть формализован как:

$$gNB^* = \arg \max_{j \in M} \text{SINR}_{i,j} = \arg \max_{j \in M} \frac{P_j h_j^2}{\sum_{k \neq j} P_k h_k^2 + N_0}, \quad (4.2)$$

где $\text{SINR}_{i,j}$ – отношение сигнал-помеха плюс шум для пользователя i от базовой станции j ;

N_0 – спектральная плотность шума.

Однако и этот подход имеет существенные ограничения. Он не учитывает временную динамику трафика и может приводить к субоптимальным решениям в условиях быстро меняющейся нагрузки. Кроме того, вычисление SINR требует знания информации о интерференции от всех соседних станций, что создает значительную сигнальную нагрузку в сети.

Концепция «смещения» является широко распространенным эвристическим подходом к гетерогенным сетям. Это включает в себя искусственное масштабирование индикатора уровня принимаемого сигнала RSSI путем введения термина «смещение» - Cell Range Extension, который зависит от уровня базовой станции. Эта манипуляция предназначена для того, чтобы направить пользователя к базовым станциям более низкого уровня, эффективно «разгружая» нагрузку с макроячейки.

Таким образом, методы с использованием смещения представляют собой промежуточное решение, где к мощности малых сот искусственно добавляется положительное смещение:

$$\text{gNB}^* = \arg \max_{j \in M} (P_j + B_j) h_j^2, \quad (4.3)$$

где B_j – смещение для j -й базовой станции.

Хотя этот подход позволяет перераспределить часть нагрузки с макросот на малые соты, его фундаментальным недостатком является статичность параметра B_j , который не адаптируется к изменяющимся условиям сети и характеристикам трафика.

Современные исследования предлагают использование продвинутых алгоритмов машинного обучения для решения задачи ассоциации. Одной из наиболее важных задач в области сетевого планирования и управления является эффективное устранение неопределенности в отношении трафика [39]. Точное прогнозирование трафика позволяет свести к минимуму чрезмерное выделение ресурсов в процессе планирования и позволяет упреждающе распределять ресурсы, тем самым предотвращая снижение производительности сети.

В последние годы подходы, основанные на глубоком обучении, стали популярными альтернативами для прогнозирования структуры трафика в сетях мобильной и беспроводной связи. Эти методы предлагают многообещающее решение для оптимизации работы сети и обеспечения эффективного использования ресурсов [27;31;38;71]. В частности, в исследовании [31] рассматривалось использование искусственных нейронных сетей для прогнозирования трафика в сетях 4G. Эти прогнозы были использованы для оптимизации распределения ресурсов обратной связи в сети.

В [68] был предложен новый механизм прогнозирования трафика, основанный на архитектуре нейронной сети с долговременной памятью (LSTM). Эта модель была разработана для использования пространственных и

временных зависимостей в данных о трафике с целью разделения сети на сегменты в автомобильных коммуникациях.

Кроме того, в [18] были представлены модели машинного обучения, которые исследовали пространственно-временные корреляции. С помощью сетей LSTM были определены долгосрочные зависимости, имеющие решающее значение для точного прогнозирования трафика. Учитывая динамический характер перемещения пользователей внутри сотовой сети, в данном исследовании были использованы корреляции между потоками трафика между соседними базовыми станциями для повышения эффективности обучения и прогнозирования.

В другом исследовании, опубликованном в [19], представлена новая архитектура глубокой нейронной сети, известная как пространственно-временная междоменная сеть. Цель данной архитектуры – выявление сложных закономерностей в данных сотовой связи с использованием сети долговременной памяти (LSTM). Данный подход демонстрирует эффективность при моделировании пространственно-временных корреляций в данных сотовой связи.

Кроме того, в [68] представлен механизм отключения базовой станции, основанный на прогнозировании трафика. В нем использовалась модифицированная вейвлет-нейронная сеть для прогнозирования будущих схем трафика базовых станций. На основе полученных прогнозов предложен механизм энергосбережения для базовых станций.

При этом для ускорения процедуры передачи пользователя между станциями в стандарте 5G был разработан протокол DAPS (Dual Active Protocol Stack), при котором исходная и целевая ячейки поддерживают активные стеки протоколов одновременно в процессе передачи данных. Это позволяет пользовательскому оборудованию UE взаимодействовать одновременно с обеими ячейками, обеспечивая более плавный переход. В отличие от устаревших процессов передачи обслуживания LTE, при которых UE

прерывает соединение с исходной ячейкой перед установлением новой, DAPS позволяет выполнять параллельную обработку. Это сводит к минимуму риск сбоя радиолинии и позволяет снизить задержку при смене базовой станции к минимуму.

В свою очередь в настоящей работе рассматривается возможность применения рекуррентных нейронных сетей LSTM для динамической ассоциации пользовательских устройств с базовыми станциями на основе анализа текущей загруженности сети и прогнозирования нового трафика.

4.2.Модель ассоциации пользователей к базовым станциям

Рассмотрим развитие модели, предложенной в третьей главе, применительно к передаче URLLC трафика. Сеть содержит удаленный сервер, который координирует работу конечного числа N устройств, подключенных к одной из M базовых станций (рисунок 4.2).

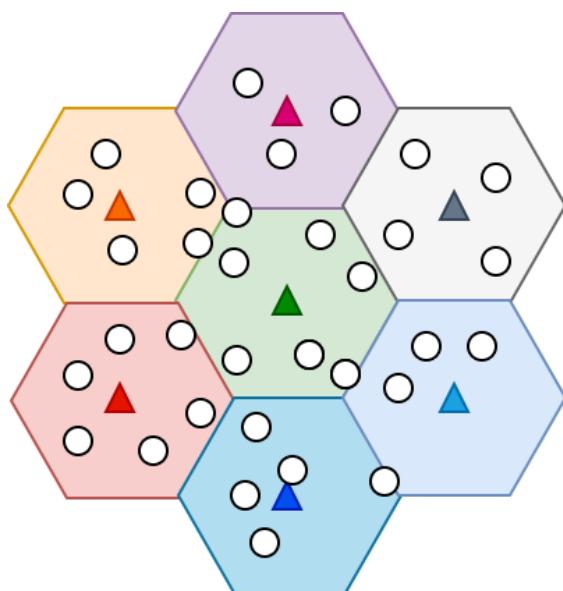


Рисунок 4.2 – Структура сотовой сети из 7 базовых станций

Классическое поведение пользователя исполнительного устройства в сценарии URLLC, например промышленного интернета вещей (Industrial Internet of Thing, IIoT) заключается в следующем (рисунок 4.3) [55; 73]:

- каждое i -е устройство получает команды через сообщения с уникальными номерами j и размерами $V_{i,j}$.
- временные интервалы между сообщениями распределяются по экспоненциальному закону, и их интенсивность зависит от интенсивности входящего потока;
- сформированные сообщения передаются по надёжному каналу на базовую станцию, где они сохраняются в отдельных очередях для каждого устройства перед передачей по беспроводному каналу связи;
- устройство использует управляющие сообщения из нисходящего канала связи, не подтверждая полученное сообщение.

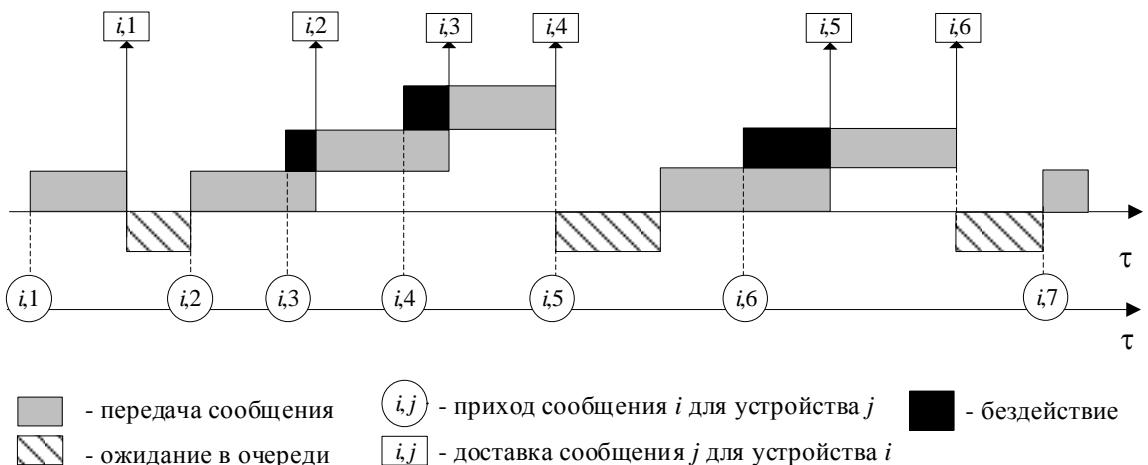


Рисунок 4.3 – Поведение пользовательского устройства

Так как рассматривается работа нисходящего канала связи в OFDM системе, то все время работы разбито на временные кадры одинаковой продолжительности. В одном временном кадре ограниченная полоса частот может быть распределена между несколькими устройствами.

Из-за наличия замираний в беспроводном канале связи, его состояние меняется со временем для каждого устройства: в каждый момент времени t , состояние канала может быть описано вектором

$C(t) = [C_{ij}(t), t = \overline{1, N}, j = \overline{1, M}]$, где $C_i(t)$ – максимально достижимая скорость для i -го устройства, подключенного к j -ой базовой станции в момент времени t .

Рассмотрим работу одной соты (рисунок 4.4). В ней находится N пользователей. Решающий вклад в эффективность данной системы вносит планировщик базовой станции, чьей задачей является распределение ресурса канала между устройствами. Результат работы планировщика можно представить в виде вектора $a(t) = [a_i(t), t = \overline{1, U_m}]$, где $a_i(t)$ – доля ресурсов канала, выделенная i -му устройству в момент времени t . Таким образом, обобщенным результатом работы планировщика для каждого устройства i является пара значений \bar{x}_i – среднее время передачи по беспроводному каналу и \bar{x}_i^2 – коэффициент вариации среднего времени передачи.

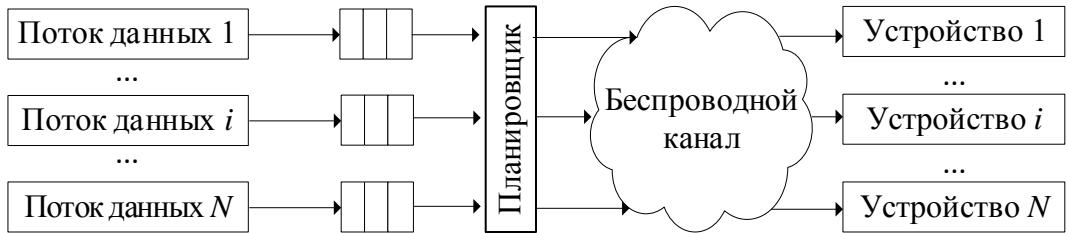


Рисунок 4.4 – Модель работы одной соты 5G

Очевидным ограничением в работе планировщика является тот факт, что доступная полоса частот для передачи сообщений является ограниченной, следовательно, должно выполняться следующее неравенство

$$\forall t : \sum_i a_i(t) \leq 1. \quad (4.4)$$

Введем для исследуемой системы следующие допущения:

$\forall i: V_{i,j}, j = \overline{1, \infty}$ – эргодический случайный процесс с конечным математическим ожиданием \bar{V} ;

$\forall i: x_{i,j}, j = \overline{1, \infty}$ – это последовательность тождественно распределенных случайных величин, независимых друг от друга и от входного потока, их математическое ожидание равно \bar{x}_i .

Изменения состояния беспроводного канала связи происходят с сохранением следующего выражения:

$$\int_{t_{i,j}}^{t_{i,j}+x_{i,j}} a_i(t) C_{i,j}(t) dt = C_{i,j} \int_{t_{i,j}}^{t_{i,j}+x_{i,j}} a_i(t) dt = V_{i,j}, \quad (4.5))$$

где $t_{i,j}$ – время начала передачи сообщения j устройству i .

$\forall i: C_{i,j}^{-1}, j = \overline{1, \infty}$ – это эргодический случайный процесс с конечным математическим ожиданием.

Представленное описание системы и допущения, сделанные выше, описывают бесконечный и неисчислимый набор гипотетических алгоритмов планирования с возможными аналитическими оценками.

4.3. Асимптотическая нижняя граница средней задержки

Введем критерий эффективности работы соты сети 5G

$$B_m = \sum_{i \in U_m} \left[\lim_{T \rightarrow \infty} \frac{1}{H_i^T} \sum_{j=1}^{H_i^T} b_{ij} \right], \quad (4.6)$$

где H_i^T – число сегментов, отправленных пользователем i за время T ;

b_{ij} – длительность отправки сегмента j пользователю i .

Ранее в работе [51] были найдены нижние границы для подобных систем, используя аналогичный подход, возможно получить следующий результат.

$$B_m = \begin{cases} 0, \sum_{i \in U_m} \frac{(K_i)^m}{\tau_i} \leq 1 \\ \sum_{i \in U_m} \max(\sqrt{K_i^m \mu_m} - \tau_i, 0) \text{ иначе} \end{cases} \quad (4.7)$$

где $K_i^m = \frac{\bar{V}_i}{C_i}$, \bar{V}_i – средний размер передаваемого сообщения, C_i – максимально

достижимая скорость для i -го устройства;

$$\mu_{0m} \text{ – решение уравнения } \sum_i \frac{K_i^m}{\max(\sqrt{K_i^m \mu_{0m}} - \tau_i)} = 1;$$

Рассмотрим сценарий перегрузки:

$$b_i^m = \lim_{T \rightarrow \infty} \frac{1}{H_i^T} \sum_{j=1}^{H_i^T} b_{ij} > 0, \quad i \in U_m, \quad (4.8)$$

$$\text{тогда } \mu_{0m} = \left(\sum_{i \in U_m} \sqrt{K_i^m} \right)^2.$$

Следовательно,

$$B_m = \sum_{i \in U_m} \sqrt{K_i^m \mu_{0m}} - \sum_{i \in U_m} \tau = \sqrt{\mu_{0m}} \sum_{i \in U_m} \sqrt{K_i^m} - \sum_{i \in U_m} \tau = \left(\sqrt{\mu_{0m}} \sum_{i \in U_m} \sqrt{K_i^m} \right)^2 - \sum_{i \in U_m} \tau. \quad (4.9)$$

Таким образом, средняя задержка по всем сотам может быть определена как

$$B = \frac{1}{N} \sum_{m=1}^M B_m = \frac{1}{N} \sum_{m=1}^M \left(\sum_{i \in U_m} \sqrt{K_i^m} \right)^2 - \frac{1}{N} \sum_{i=1}^N \tau_i. \quad (4.10)$$

Теорема: Асимптотическая нижняя граница длительности отправки сообщений для статической ассоциации пользователей определяется следующим выражением

$$B^* = \frac{N}{M} E^2 \left[\sqrt{K_i^*} \right] - E[\tau_i], \quad (4.11)$$

где $K_i^* = \min_m \{K_{im}, m = \overline{1, M}\}$.

Доказательство:

Рассмотрим оптимизационную задачу

$$\text{minimaze: } B = \frac{1}{N} \sum_{m=1}^M \left(\sum_{i \in U_m} \sqrt{K_i^m} \right)^2 - \frac{1}{N} \sum_{i=1}^N \tau_i \quad (4.12)$$

по возможным вариантам ассоциации пользователя i к базовой станции j

$$\text{subject to } K_i = \{K_{im}, m = \overline{1, M}\}, i = \overline{1, N}. \quad (4.13)$$

Напомним неравенство Йенсена:

$$f = \left(\sum_{i=1}^N q_i x_i \right) \leq \sum_{i=1}^a q_i f(x_i). \quad (4.14)$$

Адаптируем настоящее неравенство к нашему случаю

$$f(x) = x^2, q_i = \frac{1}{N}, i = 1, N \Rightarrow \sum_{i=1}^N q_i = 1, \quad (4.15)$$

$$\left(\sum_{i=1}^n \frac{1}{n} x_i \right)^2 \leq \sum_{i=1}^n \frac{1}{n} x_i^2. \quad (4.17)$$

Рассмотрим условие

$$\sum_{m=1}^M \left(\sum_{i \in U_m} \sqrt{K_i^m} \right)^2 = M \sum_{m=1}^M \frac{1}{M} \left(\sum_{i \in U_m} \sqrt{K_i} \right)^2 \geq M \left(\sum_{m=1}^M \frac{1}{M} \sum_{i \in U_m} \sqrt{K_i^m} \right)^2 = \frac{1}{M} \left(\sum_{i=1}^N \sqrt{K_i} \right)^2. \quad (4.18)$$

Подставим (4.18) в следующее выражение

$$\frac{1}{N} \sum_{i=1}^N \left(\sum_{i \in U_m} \sqrt{K_i} \right)^2 - \frac{1}{N} \sum_{i=1}^N \tau_i \geq \frac{N}{M} \left(\frac{1}{N} \sum_{i=1}^N \sqrt{K_i^m} \right)^2 - \frac{1}{N} \sum_{i=1}^N \tau_i. \quad (4.19)$$

Так как функция корня и возведения в квадрат являются монотонно возрастающими для $\forall x > 0$, то

$$\min \left[\frac{N}{M} \left(\frac{1}{N} \sum_{i=1}^N \sqrt{K_i} \right)^2 - \frac{1}{N} \sum_{i=1}^N \tau_i \right] = \frac{N}{M} \left(\frac{1}{N} \sum_{i=1}^N \sqrt{\min K_i} \right)^2 - \frac{1}{N} \sum_{i=1}^N \tau_i \quad (4.20)$$

при $N \rightarrow \infty$ получим

$$B^* = \frac{N}{M} \left(E \left[\sqrt{K_i^*} \right] \right)^2 - E[\tau_i], \quad (4.21)$$

где $K_i^* = \min_m \{K_{im}, m = 1, M\}$.

4.4. Метод динамической ассоциации пользователей с базовыми станциями

Основным обоснованием использования LSTM по сравнению с другими архитектурами нейронных сетей в этом исследовании является ее способность эффективно обрабатывать и сохранять информацию в расширенных последовательностях, тем самым устранив необходимость в заранее определенном окне выборки в качестве входных данных.

При настройке уровня LSTM важным аспектом является определение количества нейронов. Это простой параметр, который может варьироваться от 1 до 5 и обычно выбирается эмпирически. При выборе учитывается оптимальное время обучения и требуемая точность прогнозирования. На выходном уровне для прогнозирования показателя на следующем временном шаге используется один нейрон с линейной функцией активации.

На рисунке 4.5 показана архитектура нейронной сети с долговременной кратковременной памятью (LSTM), используемой для прогнозирования входящего потока сообщений. В соответствии с рисунком, мы используем последовательность из T последовательных точек данных о трафике $\{x_1, x_2, x_3, \dots, x_T\}$ в качестве входных данных для модели LSTM, чтобы предсказать значение потока трафика сотовой связи x_{T+1} в $(T + 1)$ -й временной интервал.

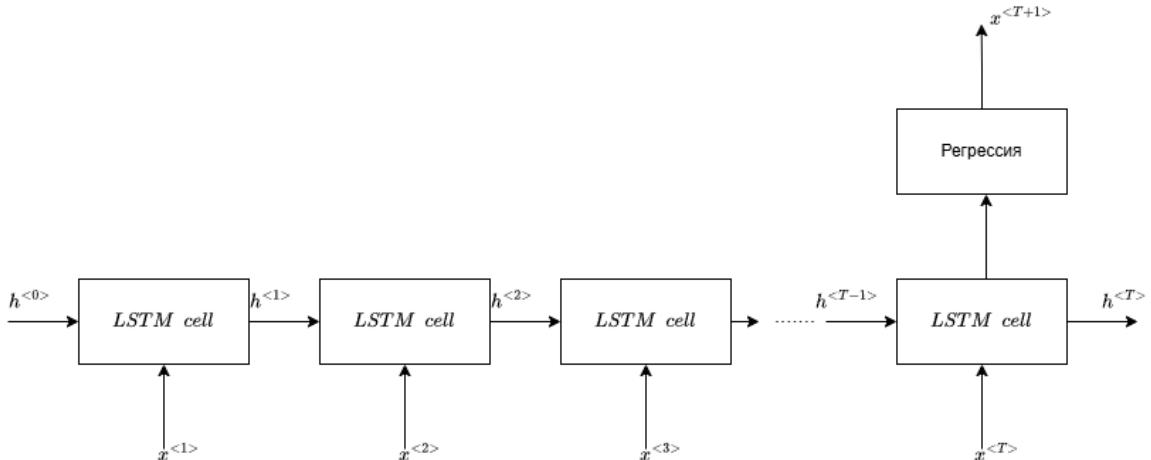


Рисунок 4.5 – Архитектура нейронной сети

Процесс прогнозирования потока трафика требует T дискретных временных шагов, каждый из которых соответствует отдельной ячейке LSTM. Архитектура LSTM использует механизм самоподключения, гарантирующий, что в любой момент сеть может обрабатывать только одну точку данных.

Основным компонентом сети LSTM является ячейка LSTM, как показано на рисунке 4.6 t -я ячейка LSTM соответствует t -му временному шагу и имеет возможность либо сохранять, либо обновлять состояние ячейки с помощью тщательно разработанной структуры, включающей три различных типа элементов управления: входной элемент управления, выходной элемент управления и элемент забывания [24].

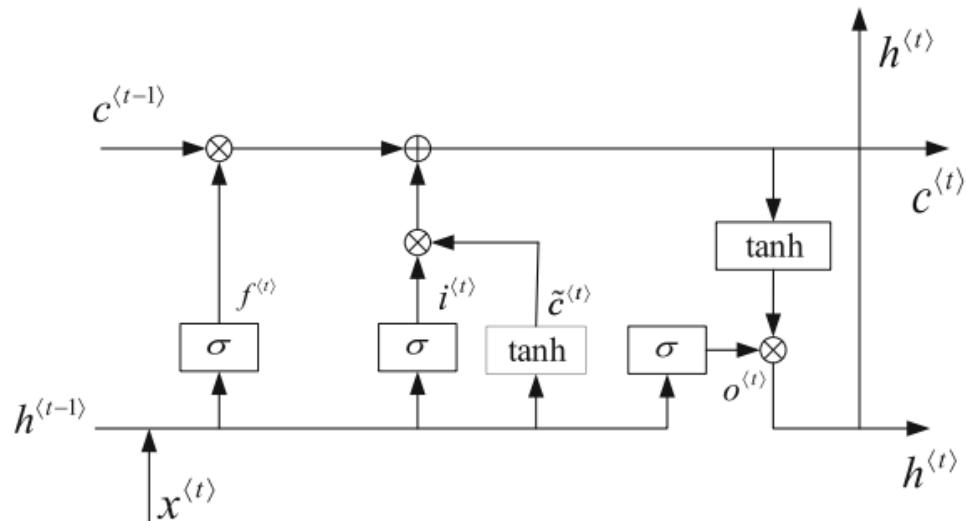


Рисунок 4.6 – Ячейка LSTM сети

Предсказание поступающего трафика с помощью рекуррентной нейронной сети можно описать следующими шагами:

1. Сбор данных: Сбор информации о мобильном трафике с помощью пассивного инструмента, который собирает данные от базовой станции. Информация включает блоки ресурсов, размер транспортного блока и схему модуляции, назначенную каждому пользователю, подключенному к базовой станции.

2. Преобразование данных: Преобразование собранных данных в вектор $S(T)$, который содержит количество ресурсных блоков, количество сообщений, отправленных, и общую сумму размеров транспортных блоков для заданного временного интервала T .

3. Одношаговое прогнозирование: Использование архитектуры “все к одному”, где сеть анализирует мобильный трафик в течение фиксированного количества временных интервалов до T и затем пытается предсказать трафик в следующем временном интервале ($T+1$).

Таким образом на выходе получаем спрогнозированное значение числа активных устройств (устройств, для которых есть сообщение).

Обучение LSTM-сети проводилось на наборе данных, полученных методом Монте-Карло. Для обучения использовалась комбинированная функция потерь, учитывающая особенности сетевой нагрузки

$$L_{\text{total}} = \alpha \text{MSE} + \beta \text{MAE}_{\text{peak}}, \quad (4.22)$$

где MSE (Mean Squared Error) обеспечивает общую точность прогнозирования; MAE_{peak} фокусируется на точности предсказания пиков нагрузки.

Обучение проводилось с использованием оптимизатора Adam с начальным learning rate 0.001 и экспоненциальным затуханием. Для предотвращения переобучения применялась техника «early stopping».

Алгоритм принятия решений об ассоциации пользователей формализован как задача многокритериальной оптимизации с учетом прогнозируемого состояния сети. Для каждого пользователя i и базовой станции j вычисляется расширенная функция оценки

$$S_{ij}(t+1) = \alpha \text{SINR}_{ij}(t) - \beta L_j(t+1) - \gamma H_{\text{count},i}, \quad (4.23)$$

где SINR_{ij}(t) - текущее отношение сигнал-помеха плюс шум;

$L_j(t+1)$ – прогнозируемая нагрузка на базовую станцию j в следующий временной интервал;

$H_{\text{count},i}$ – счетчик хэндоверов для пользователя i с учетом истории переключений α , β , γ – адаптивные весовые коэффициенты, оптимизируемые методом градиентного спуска.

Пользователь ассоциируется с базовой станцией, максимизирующей функцию оценки

$$\text{gNB}^* = \arg \max_{j \in M} S_{ij}(t+1) \quad (4.24)$$

с дополнительным условием стабильности

$$S_{ij}(t+1) - S_{ij}(t) > \theta, \quad (4.25)$$

где $S_{ij}(t)$ – оценка текущей ассоциации;

θ – порог гистерезиса, предотвращающий избыточные переключения.

Данный подход позволяет балансировать между несколькими противоречивыми целями: максимизацией немедленной выгоды (высокий SINR), обеспечением долгосрочной эффективности (избежание перегруженных станций), и минимизацией избыточных хэндоверов.

4.5. Результаты моделирования

Имитационное моделирование системы было направлено на оценку эффективности предложенного метода динамической ассоциации пользовательских устройств с базовыми станциями, основанного на прогнозировании трафика с помощью рекуррентных нейронных сетей.

Пространственная модель сети, в которой базовые станции с различными характеристиками (разная максимальная пропускная способность, зона

покрытия) и пользовательские устройства были размещены на плоскости произвольным образом.

Для каждого i -го устройства генерировался поток сообщений, характерный для сценариев ПoT. Сообщения (команды управления, телеметрия) фиксированного размера поступали в очередь с экспоненциально распределенными интервалами.

Важным аспектом была модель мобильности пользователей: устройства могли перемещаться по площади сети, что динамически меняло для них набор доступных базовых станций и качество канала связи с ними. Состояние канала связи для каждой пары «устройство-базовая станция» в момент времени t описывалось величиной $C_{i,j}(t)$ – максимально достижимой скоростью передачи данных. Эта величина динамически изменялась во времени согласно двумерному броуновскому движению, моделируя эффекты замираний, затенения и интерференции. При этом скорость передачи ограничена максимумом при нахождении вплотную к базовой станции и 0.1 от максимума при нахождении на максимальном удалении от базовой станции.

Планировщик на каждой базовой станции, работающий по алгоритмам Round Robin (RR) или Maximum Throughput (MT), на каждом временном интервале распределял доступную полосу пропускания между ассоциированными с ней UE.

При моделировании традиционная статическая ассоциация, основанная на максимальном уровне принимаемого сигнала RSSI и приводящая к разбиению сети на диаграммы Вороного, сравнивалась с предложенным динамическим методом. Ключевые параметры моделирования отражены в таблице 4.1.

На рисунках 4.7 и 4.8 показаны графики средней задержки передачи сообщения по нисходящему каналу связи при использование различных методик ассоциации пользователей к базовым станциям для двух алгоритмов планирования ресурсов – RR и MT.

Таблица 4.1 – Параметры моделирования

Число устройств в сети	100 ... 1000
Число базовых станций	5
Максимальные скорости передачи для каждой базовой станции, Мбит/с	[100, 150, 120, 80, 100]
Средняя длина сообщения, байт	32
Число OFDM символов в субфрейме	448

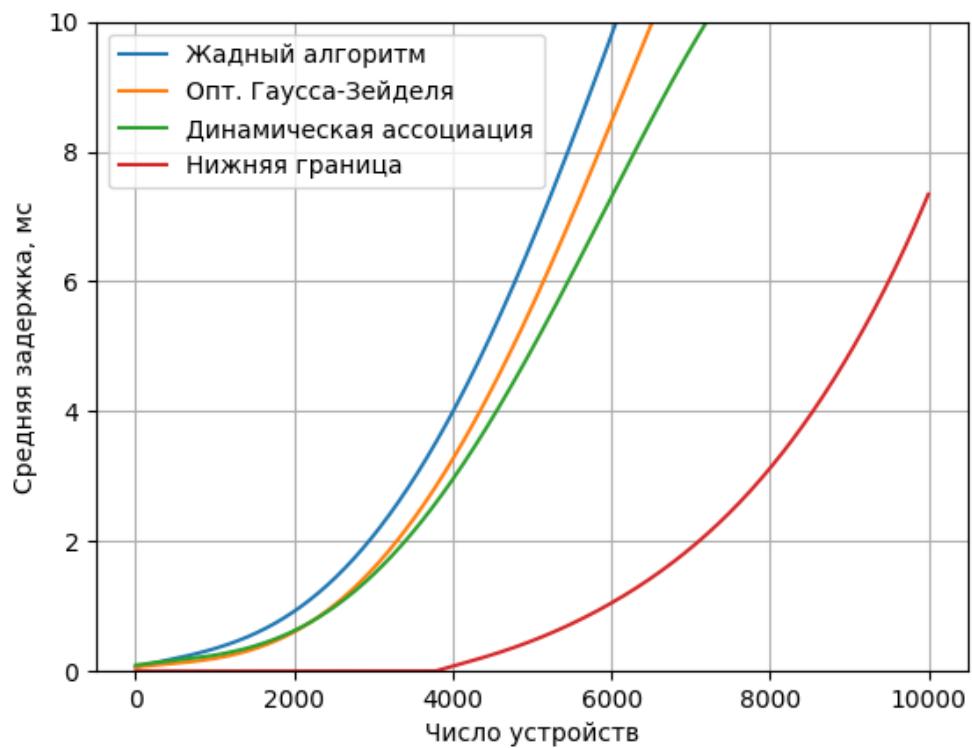


Рисунок 4.7 – График средней задержки передачи сообщения для алгоритма планирования МТ

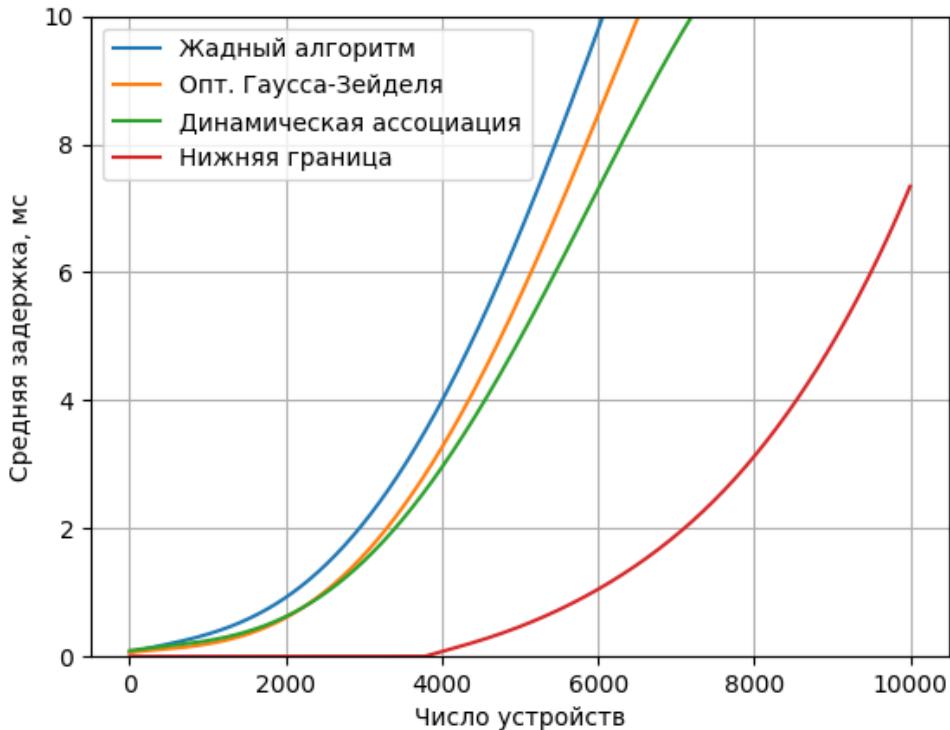


Рисунок 4.8 – График средней задержки передачи сообщения для алгоритма планирования RR

Выводы по четвертой главе

Разработан и проанализирован метод динамической ассоциации пользователей с базовыми станциями в гетерогенных сотовых сетях, направленный на оптимизацию распределения нагрузки и снижение задержки передачи данных. Традиционные подходы, основанные на максимизации уровня принимаемого сигнала, продемонстрировали свою неэффективность в условиях неравномерного распределения пользователей и неравномерности входящего трафика. Предложенный метод, использующий прогнозирование трафика с помощью рекуррентных нейронных сетей LSTM, позволил учесть пространственно-временные зависимости в данных и адаптивно перераспределять пользователей между базовыми станциями.

Практическая значимость метода заключается в его способности интегрироваться в существующие сетевые инфраструктуры без необходимости модификации аппаратного обеспечения. Полученные результаты открывают

перспективы для дальнейших исследований в области применения машинного обучения для управления ресурсами базовых станций в сетях 5G и последующих поколений. Таким образом, предложенный метод представляет собой важный шаг в развитии интеллектуальных систем управления гетерогенными сетями связи.

Предложен алгоритм вычисления значения нижней границы средней задержки передачи сообщений чувствительного к задержкам трафика в системе OFDM. Алгоритм принимает на вход параметры системы: средний размер сообщения, интенсивность входного потока и среднее значение максимально достижимой скорости передачи данных для каждого устройства, подключённого к базовой станции. Предложенный алгоритм может быть использован разработчиками систем при планировании и развёртывании сетей OFDM, а также являются основой для создания технологии сверхнадёжной связи с низкими задержками.

ЗАКЛЮЧЕНИЕ

Актуальность темы обусловлена стремительным ростом числа подключенных устройств и разнообразием требований к качеству обслуживания QoS, предъявляемых различными типами трафика. Сценарии использования сети 5G демонстрируют противоречивые требования к ресурсам сети, что создает значительные сложности при проектировании эффективных механизмов управления ресурсами. В работе предложены новые решения, направленные на оптимизацию распределения ресурсов в восходящем и нисходящем каналах связи, а также на динамическую ассоциацию пользователей с базовыми станциями, что в совокупности позволяет обеспечить высокую надежность, низкую задержку и эффективное использование ресурсов в условиях высокой плотности устройств.

Получены следующие результаты:

1. Разработана модель восходящего канала как система массового обслуживания типа $G|G|1$ с приоритетным обслуживанием URLLC трафика, для которой:

- предложен метод скользящего сетевого кодирования, показавший существенное преимущество по сравнению с традиционными схемами K -повторений. Метод позволяет значительно снизить вероятность ошибок и среднюю задержку передачи сообщений за счет совмещения преимуществ сетевого кодирования и механизма повторных передач. Важной особенностью метода является возможность декодирования данных «на лету», что критически важно для сценария URLLC. Результаты имитационного моделирования подтвердили способность метода обеспечивать надежность на уровне 99.999% при одновременном снижении задержки;

- алгоритм динамического изменения числа преамбул на основе методов машинного обучения: регрессии опорных векторов SVR и рекуррентных нейронных сетей RNN. Этот подход позволяет прогнозировать количество

активных устройств и интенсивность поступающего трафика, для динамического перераспределения ресурсов случайного доступа (пreamble) между высокоприоритетным URLLC и низкоприоритетным mMTC трафиком. Это способствует минимизации коллизий, снижению задержки и эффективному использованию доступной полосы пропускания.

2. Разработана модель исходящего канала как система массового обслуживания типа $G|G|1$ с приоритетным обслуживанием URLLC трафика, для которой:

– предложен метод адаптивного мультиплексирования гетерогенного трафика eMBB и URLLC. В основе метода лежит использование решетчатого кодирования в сочетании с LDPC-кодами, что позволяет интегрировать трафик с различными требованиями в единый ресурсный блок без значительной деградации пропускной способности eMBB. Результаты имитационного моделирования показали, что метод позволяет обеспечить сверхнизкую задержку передачи URLLC-сообщений менее 1 мс при высокой вероятности успешной доставки более 99.999% согласно требованиям QoS.

3. Разработан метод динамической ассоциации пользователей с базовыми станциями, основанный на прогнозировании нагрузки с использованием сетей LSTM. В отличие от традиционных методов, ориентированных на максимальный уровень сигнала RSSI, предложенное решение позволяет балансировать нагрузку между сотами, учитывая пространственно-временные характеристики трафика. Теоретической основой метода является выведенная в работе асимптотическая нижняя граница средней задержки. Результаты моделирования показали, что динамическая ассоциация позволяет значительно снизить среднюю задержку передачи данных по сравнению со статическими схемами распределения.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

БС – базовая станция

Субфрейм – временной отрезок работы соты продолжительностью 1 мс

Фрейм – временной отрезок работы соты продолжительностью 10 мс

ACK – acknowledgement, положительная квитанция

CP – cyclic prefix, циклический префикс

CRC – cyclic redundancy check, циклический избыточный код

eMBB – enhanced mobile broadband, расширенная мобильная широкополосная связь

gNB – gNodeB, базовая станция сети 5G

IoT – internet of things, Интернет вещей

mMTC – massive machine type communication, массовые коммуникации машинного типа

NACK – negative acknowledgement, отрицательная квитанция

OFDM – orthogonal frequency-division multiplexing, мультиплексирование с ортогональным частотным разделением каналов

OFDMA – orthogonal frequency-division multiple access, технология множественного доступа с ортогональным частотным разделением каналов

QAM – quadrature amplitude modulation, квадратурная амплитудная модуляция

QoS – quality of service, качество обслуживания

URLLC – ultra-reliable low-latency communication, сверхнадежная связь с низкой задержкой

СПИСОК ЛИТЕРАТУРЫ

1. 3GPP RAN, “Overview of RAN aspects,” Workshop on 3GPP submission towards IMT-2020, Oct. 2018, Brussels.
2. 3GPP RP-181477. SID on physical layer enhancements for NR URLLC. Jun. 2018.
3. 3GPP TS 22.261 16.4.0. Service requirements for next generation new services and markets. 3GPP, 2017 55 p.
4. 3GPP TS 22.368 v11.2.0 ”service requirements for Machine-Type Communications,” Jun. 2011.
5. 3GPP TS 36.211, “Physical channels and modulation”
6. 3GPP, “Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC) (release 16),” 3GPP TR 38.824 V2.0.1, 2019.
7. A. Lo, Y. W. Law, M. Jacobsson, and M. Kucharzak, ”Enhanced LTEadvanced randomaccess mechanism for massive machine-to-machine M2M) communications,” in in 27th World Wireless Research Forum (WWF) Meeting, 2011.
8. A. Samir, M. M. Elmesalawy, A. S. E.-D. Ali, and I. Ali, “Partial contention free random access protocol for M2M communications in LTE networks,” J. Wireless Netw. Commun., vol. 6, no. 3, pp. 66–72, 2016.
9. Alsbou, Nesreen & Refai, Hazem. (2010). Dynamic slot allocation algorithm for R-ALOHA with priority (PR-ALOHA). 1-6. 10.1109/GLOCOM.2010.5683533.
10. Alsenwi M., Hong C. S. Resource Scheduling of URLLC/eMBB traffics in 5G New Radio: A Punctured Scheduling Approach. Available at: http://networking.knu.ac.kr/xe/Gallery/entry/document_srl/141/page/layouts/net/publications/data/KCC2018/14.Madyan.pdf.
11. Alsenwi M., Tran N. H., Bennis M., Kumar Bairagi A., and Hong C. S., eMBB-URLLC Resource Slicing: A Risk-Sensitive Approach, IEEE Commun. Lett., vol. 23, no. 4, pp. 740–743, April 2019.

12. Alsenwi M., Tran N. H., Bennis M., Pandey S. R., Bairagi A. K., and Hong C. S. Intelligent resource slicing for embb and urllc coexistence in 5g and beyond: A deep reinforcement learning based approach. Available at: <https://arxiv.org/pdf/2003.07651.pdf>.
13. Anand A., Veciana G., and Shakkottai S. Joint Scheduling of URLLC and eMBB Traffic in 5G Wireless Networks, *IEEE/ACM Transactions on Networking*, vol: 28, pp.477-490, April 2020.
14. Anand A., Veciana G., and Shakkottai S., Joint Scheduling of URLLC and eMBB Traffic in 5G Wireless Networks, in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, April 2018, pp. 1970–1978.
15. Bairagi K., Munir M. S., Alsenwi M., Tran N. H., Alshamrani S. S., Masud M., Han Z., and Hong C. S. Coexistence mechanism between eMBB and uRLLC in 5G wireless networks, 2020. Available at: <https://arxiv.org/pdf/2003.04551.pdf> (accessed 20 November 2020).
16. C. Bockelmann, N. K. Pratas, G. Wunder, S. Saur, M. Navarro, D. Gregoratti, G. Vivier, E. De Carvalho, Y. Ji, v. Stefanovic, P. Popovski, ' Q. Wang, M. Schellmann, E. Kosmatos, P. Demestichas, M. RacealaMotoc, P. Jung, S. Stanczak, and A. Dekorsy, “Towards massive connectivity support for scalable mMTC communications in 5G networks,” *IEEE Access*, vol. 6, pp. 28 969–28 992, 2018.
17. C. Kong, “OPNET Simulation of Signaling System No. 7 (SS7) Network Interfaces”, *Storming Media*, Page: 256-258, Mar 2000.
18. C. Qiu, Y. Zhang, Z. Feng, P. Zhang, and S. Cui, “Spatio-temporal wireless traffic prediction with recurrent neural network,” *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 554-557, Aug. 2018.
19. C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, “Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data,” *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1389–1401, Mar. 2019
20. Chen, M. et al. “Machine-to-Machine Communications: Architectures, Standards and Applications.” *KSII Trans. Internet Inf. Syst.* 6 (2012): 480-497.

21. Chraiti M., Ghrayeb A., and Assi C., A NOMA Scheme Exploiting Partial Similarity Among Users Bit Sequences, *IEEE Trans. on Commun.*, vol. 66, no. 10, pp. 4923–4935, Oct 2018
22. Enhanced Algorithm for Dynamic Changing of Frame Size in Radio Frequency Identification Systems / N. A. Yankovskii, A. Y. Kolesnikova, K. V. Afanasev, N. V. Stepanov // Wave Electronics and Its Application in Information and Telecommunication Systems. – 2022. – Vol. 5, No. 1. – P. 532-535.
23. G. Durisi, T. Koch, and P. Popovski, “Toward massive, ultrareliable, and low-latency wireless communication with short packets,” *Proceedings of the IEEE*, vol. 104, pp. 1711–1726, Sept 2016.
24. Gers, F.A., Schmidhuber, J.: Recurrent nets that time and count. In: *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, vol. 3, pp. 189–194 (2000)
25. Goldsmith J. *Wireless communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005, 644 p. doi: 10.1017/CBO9780511841224.
26. Goldsmith J., Chua S. G. Adaptive coded modulation for fading channels, *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 595-602, May 1998.
27. H. D. Trinh, L. Giupponi, and P. Dini, “Mobile traffic prediction from raw data using LSTM networks,” in *Proc. IEEE PIMRC*, Sep. 2018, pp. 1827-1832.
28. H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jantti, “A. Ghosh, Machinetype communications: current status and future perspectives to- ward 5g systems, *IEEE Commun. Mag.* 53 (9)(2015) 10–17
29. I. Leyva-Mayorga, C. Stefanovic, P. Popovski, V. Pla, and J. Martinez-Bauset, “Ran dom access for machine-type communications,” *Wiley 5G Ref: The Essential 5G Reference Online*, 2019
30. Israel Leyva-Mayorga, & Tello Oquendo, Luis & Pla, Vicent & Martinez-Bauset, Jorge & Casares-Giner, Vicente. (2017). On the Accurate Performance Evaluation of the LTE-A Random Access Procedure and the Access Class Barring

Scheme. IEEE Transactions on Wireless Communications. PP. 1-1. 10.1109/TWC.2017.2753784.

31. I. Leyva-Mayorga et al.. Adaptive access class barring for efficient mMTC. Computer Networks. 149. 10.1016/j.comnet.2018.12.003.

32. I. Loumiotis, E. Adamopoulou, K. Demestichas, P. Kosmides, and M. Theologou, “Artificial neural networks for traffic prediction in 4G networks,” in Proc. 8th Int. Wireless Internet Conf., Nov. 2014, pp. 141-146.

33. Ioannis Loumiotis, & Adamopoulou, Evgenia & Demestichas, Konstantinos & Stamatiadi, Theodora & Theologou, M.. (2013). On the predictability of next generation mobile network traffic using artificial neural networks. International Journal of Communication Systems. 28. 10.1002/dac.2728.

34. I. M. Delgado-Luque et al., “Evaluation of latency-aware scheduling techniques for M2M traffic over LTE,” in Proc. 20th Eur. Signal Process. Conf. (EUSIPCO), Bucharest, Romania, 2012, pp. 989–993.

35. ITU-R M.2083-0. IMT Vision — framework and overall objectives of the future development of IMT for 2020 and beyond. ITU-R, 2015 21 p.

36. J. Choi and J. Ding, “Network coding for k-repetition in grant-free random access,” IEEE Wireless Communications Letters, vol. 10, no. 11, pp. 2557–2561, 2021.

37. J. Choi, “On the adaptive determination of the number of preambles in RACH for MTC,” IEEE Commun. Lett., vol. 20, no. 7, pp. 1385–1388, Jul. 2016.

38. J. Hu, W. Heng, G. Zhang, and C. Meng, “Base station sleeping mechanism based on traffic prediction in heterogeneous networks,” in Proc. International Telecommunication Networks and Applications Conference (ITNAC), Nov. 2015, pp. 83-87.

39. J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, “A survey on 5g usage scenarios and traffic models,” IEEE Communications Surveys Tutorials, vol. 22, pp. 905–929, Secondquarter 2020

40. Jacobsen T., Abreu R., Berardinelli G., Pedersen K., Kovacs I. Z., Mogensen P. System level analysis of K-repetition for uplink grant-free URLLC in 5G NR. European Wireless 2019, 25th European Wireless Conf., May 2019, pp. 1–5.
41. Karimi A., Pedersen K. I., Mahmood N. H., Pocovi G., and Mogensen P., Efficient Low Complexity Packet Scheduling Algorithm for Mixed URLLC and eMBB Traffic in 5G, in 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), April 2019, pp. 1–6.
42. Kassab R., Simeone O., Popovski P. and Islam T., Non-Orthogonal Multiplexing of Ultra-Reliable and Broadband Services in Fog-Radio Architectures," in IEEE Access, vol. 7, pp. 13035-13049, 2019
43. Khalifa N. B., Angilella V., Assaad M., and Debbah M., Low-complexity channel allocation scheme for urllc traffic, IEEE Trans. on Commun., pp. 1–1, 2020
44. Korrai P. K., Lagunas E., Sharma S. K., Chatzinotas S., and Ottersten B., Slicing Based Resource Allocation for Multiplexing of eMBB and URLLC Services in 5G Wireless Networks, in 2019 IEEE CAMAD, Sep. 2019, pp. 1–5.
45. Lin S., Costello D. Error Control Coding: Fundamentals and Applications. Pearson-Prentice Hall, 2004, second ed.
46. Liu Y., Deng Y., Elkashlan M., Nallanathan A., Karagiannidis G. K. Analyzing Grant-Free Access for URLLC Service. IEEE Journal on Selected Areas in Communications, March 2021, vol. 39, no. 3, pp. 741-755, doi: 10.1109/JSAC.2020.3018822.
47. Makki B., Svensson T., Caire G., Zorzi M. Fast HARQ over finite blocklength codes: A technique for low-latency reliable communication. IEEE Trans. Wireless Communications, 2019, vol. 18, no. 1, pp. 194–209
48. Manzoor A., Kazmi S. M. A., Pandey S. R. and Hong C. S., Contract-Based Scheduling of URLLC Packets in Incumbent EMBB Traffic, in IEEE Access, vol. 8, pp. 167516-167526, 2020, doi: 10.1109/ACCESS.2020.3023128
49. Moon S., Lee J. W. Performance Study of Repetition-Based Grant-Free Schemes in the mMTC Scenario. 34th International Technical Conference on

Circuits/Systems, Computers and Communications (ITC-CSCC), 2019, pp. 1-2, doi: 10.1109/ITC-CSCC.2019.8793363

50. Palattella, M. R., Dohler, M., Grieco, A., Rizzo, G., Torsner, J., Engel, T., and Ladid, L. (2016). Internet of things in the 5G era: Enablers, architecture, and business models. *IEEE Journal on Selected Areas in Communications*, 34(3), 510-527

51. Pastushok I.A. Efficiency evaluation of scheduling algorithms for delay-sensitive traffic in OFDM downlink // Информационно-управляющие системы. 2018. № 5. С. 88–93. Doi:10.31799/1684-8853-2018-5-88-93.

52. Pastushok I.A., Boikov N. A. and Yankovskii N. A., Bit Stream Multiplexing in 5G networks, 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), St. Petersburg, Russia, 2020, pp. 1-4, doi: 10.1109/WECONF48837.2020.9131511.

53. Pastushok, I. A. Investigation of Methods for Multiplexing EMBB and URLLC Streams in a Downlink / I. A. Pastushok, N. A. Boikov // 2019 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF 2019), Санкт-Петербург, 03–07 июня 2019 года. Vol. [2]. – Санкт-Петербург: Institute of Electrical and Electronics Engineers Inc., 2019. – P. 8840593.

54. Pastushok, I. A. Bitstreams multiplexing with trellis-coded modulation and the fixed point ldpc decoding procedure for centralised wireless networks / I. A. Pastushok, N. A. Boikov, N. A. Yankovskii // Lecture Notes in Computer Science. – 2020. – Vol. 12525 LNCS. – P. 381-392. – DOI 10.1007/978-3-030-65726-0_34.

55. Pastushok, I. A. On dependent distributions of appearance and servicing of claims in the M/M/1 / I. A. Pastushok, N. A. Yankovskii // 2021 Wave Electronics and its Application in Information and Telecommunication Systems, WECONF 2021 - Conference Proceedings, Saint-Petersburg, 31 мая – 04 2021 года. Vol. [4]. – Saint-Petersburg, 2021. . – DOI 10.1109/WECONF51603.2021.9470500.

56. Pocovi G., Shariatmadari H., Berardinelli G., Pedersen K., Steiner J., Li Z. Achieving ultra-reliable low-latency communications: Challenges and envisioned system enhancements. *IEEE Network*, March 2018, vol. 32, pp. 8–15.

57. Radic, J.; Solic, P.; Skiljo, M. Anticollision algorithm for radio frequency identification system with low memory requirements: NA. *Trans. Emerg. Telecommun. Technol.* 2020, 31, e3969.
58. Radio Resource Control (RRC) Protocol specification, document TS 38.331 V16.1.0, 3GPP, Sophia, Antipolis, France, Jul. 2018.
59. Rodic, L.D.; Stancic, I.; Zovko, K.; Solic, P. Machine Learning as Tag Estimation Method for ALOHA-based RFID system. In Proceedings of the 2021 6th International Conference on Smart and Sustainable Technologies (SpliTech), Bol and Split, Croatia, 8–11 September 2021; pp. 1–6.
60. S. K. Sharma and X. Wang, "Toward Massive Machine Type Communications in Ultra-Dense Cellular IoT Networks: Current Issues and Machine Learning-Assisted Solutions," in *IEEE Communications Surveys and Tutorials*, vol. 22, no. 1, pp. 426–471, First quarter 2020, doi: 10.1109/COMST.2019.2916177.
61. S. R. Pokhrel, J. Ding, J. Park, O. S. Park, and J. Choi, "Towards enabling critical mMTC: A review of URLLC within mMTC," *IEEE Access*, vol. 8, pp. 131796–131813, Jul. 2020.
62. S. Yang and F. A. Kuipers, "Traffic uncertainty models in network planning," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 172–177, Feb. 2014.
63. Santos E. J., Souza R. D., Rebelatto J. L., and Alves H., Network Slicing for URLLC and eMBB With Max-Matching Diversity Channel Allocation, *IEEE Communications Letters*, vol. 24, no. 3, pp. 658–661, 2020.
64. Segura, David & J. Khatib, Emil & Munilla, Jorge & Barco, Raquel. (2021). 5G Numerologies Assessment for URLLC in Industrial Communications. *Sensors*.
65. Strodtbock N., Goktepe B., Schierl T., Hellge C., Samek W. Enhanced machine learning techniques for early HARQ feedback prediction in 5G. *IEEE J. Selected Areas in Communications*, 2019, vol. 37, no. 11, pp. 2573–2587.
66. Ungerboeck G. Channel coding with multilevel/phase signals, *IEEE Trans. Inf. Theory.*, vol. 28, no. 1, pp. 55–67, Jan. 1982. doi:10.1109/TIT.1982.1056454.
67. Wicker S. B. *Error Control Systems for Digital Communication and Storage*. Prentice Hall, 1995.

68. Y. Cui, X. Huang, D. Wu, and H. Zheng, “Machine learning based resource allocation strategy for network slicing in vehicular networks,” in Proc. IEEE ICCC, Aug. 2020, pp. 454-459.
69. Y. Polyanskiy, H. V. Poor, and S. Verdu, “Channel coding rate in the finite blocklength regime,” IEEE Trans. Information Theory, vol. 56, pp. 2307–2359, May 2010.
70. Y. Polyanskiy, “A perspective on massive random-access,” in 2017 IEEE International Symposium on Information Theory (ISIT), 2017, pp. 2523– 2527.
71. Y. Xu, W. Xu, F. Yin, J. Lin, and S. Cui, “High-accuracy wireless traffic prediction: A GP-based machine learning approach,” in Proc. IEEE GLOBECOM, Dec. 2017, pp. 1-6
72. Yang Y., Chen W., Li O., Ke K., and Hanzo L. Trellis- and network-coded modulation for decode-and-forward two-way relaying over time-varying channels, IEEE Trans. Veh. Technol., vol. 66, no. 6, pp. 4845– 4858, June 2017.
73. Yankovskii, N. A. Impact of correlated arrival and service flows for radio resource management in centralized wireless networks / N. A. Yankovskii, T. M. Tatarnikova // T-Comm. – 2024. – Vol. 18, No. 5. – P. 42-48.
74. Z. Dawy, et al., “Toward massive machine type cellular communications,” IEEE Wireless Communications, vol. 24, no. 1, pp. 120–128, Feb. 2017.
75. Z. Li, J. Chen, R. Ni, S. Chen, X. Li, and Q. Zhao, “Enabling heterogeneous MMTC by energy-efficient and connectivity-aware clustering and routing,” in Proc. IEEE Globecom Workshops (GC Wkshps), Dec. 2017, pp. 1–6.
76. Zantalis, F.; Koulouras, G.; Karabetsos, S.; Kandris, D. A Review of Machine Learning and IoT in Smart Transportation. Future Internet 2019, 11, 94.
77. Алгоритм оптимизации ассоциации пользователей к базовым стациям / Д. А. Русановская, Н. А. Янковский, А. Д. Каркина, В. А. Ушаков // Научному прогрессу – творчество молодых. – 2024. – № 1. – С. 323-326
78. Витвинов, В. К. Программно-аппаратный комплекс для исследования эффективности использования QUIC-протокола в системах интернета вещей / В. К. Витвинов, Н. А. Янковский // Обработка, передача и защита информации

в компьютерных системах 22 : Сборник докладов Второй Международной научной конференции, Санкт-Петербург, 11–15 апреля 2022 года. – Санкт-Петербург: Санкт-Петербургский государственный университет аэрокосмического приборостроения, 2022. – С. 178-182.

79. Использование скользящего сетевого кодирования в восходящем канале межмашинных коммуникаций / Н. А. Янковский, П. А. Гондурова, Д. М. Иванов, Г. Е. Есипов // Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем ("Радиоинфоком-2023") : Сборник научных статей по материалам VII Международной научно-практической конференции, Москва, 20–24 ноября 2023 года. – Москва: МИРЭА-Российский технологический университет, 2023.

80. Михайлов В. А. Геометрический анализ устойчивости цепей маркова в R_+^n и его приложение к вычислению пропускной способности адаптивного алгоритма случайного множественного доступа // Проблемы передачи информации, 1988. Т. 24. № 1. С. 61–73.

81. Татарникова, Т. М. Динамическое распределение ресурсов в восходящем канале сети 5G / Т. М. Татарникова, Ю. А. Ямщиков, Н. А. Янковский // Морская радиоэлектроника. – 2025. – № 4(94). – С. 22-28.

82. Цыганкова, А. А. Использование методов машинного обучения для предсказания числа меток в RFID системах / А. А. Цыганкова, Н. А. Янковский // Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем ("Радиоинфоком-2024") : Сборник научных статей по материалам VIII Международной научно-практической конференции, Москва, 18–22 ноября 2024 года. – Москва: МИРЭА - Российский технологический университет, 2024. – С. 239-243.

83. Чан, Т. З. Оптимизация использования ресурсов воздушных базовых станций на основе методов искусственного интеллекта / Т. З. Чан, А. Е. Кучеряый // Труды учебных заведений связи. – 2025. – Т. 11, № 1. – С. 62-68.

84. Чан, Т. З. Оптимизация распределения пользователей, мощности и расположения воздушной базовой станции на основе методов глубокого

обучения с подкреплением / Т. З. Чан, А. Е. Кучерявый // Труды учебных заведений связи. – 2025. – Т. 11, № 2. – С. 32-40.

85. Янковский Н.А., Пастушок И.А. О мультиплексировании потоков данных с использованием решетчатого кодирования в централизованных беспроводных сетях // Информационно-управляющие системы. 2021. № 2. С. 52-59.

86. Янковский, Н. А. Использование сетевого кодирования для массовых межмашинных коммуникаций / Н. А. Янковский // Успехи современной радиоэлектроники. – 2023. – Т. 77, № 8. – С. 50-57.

87. Янковский, Н. А. Оценка эффективности алгоритмов ассоциации и планирования для передачи чувствительного к задержке трафика по нисходящему каналу беспроводных сетей / Н. А. Янковский, Т. М. Татарникова // Успехи современной радиоэлектроники. – 2024. – Т. 78, № 8. – С. 33-38.

88. Янковский, Н. А. Сетевое кодирование в межмашинных коммуникациях / Н. А. Янковский // Волновая электроника и инфокоммуникационные системы: Материалы XXVI Международной научной конференции. В 3-х частях, Санкт-Петербург, 29 мая – 02 2023 года. Том Часть 1. – Санкт-Петербург: Санкт-Петербургский государственный университет аэрокосмического приборостроения, 2023. – С. 137-141/

89. Янковский Н. А. Динамическое управление ресурсами в восходящем канале 5G / Н.А. Янковский, Т.М. Татарникова // Информация и Космос. – 2025. - №4. – С. 38-44.

90. Янковский Н.А., Пастушок И.А. О мультиплексировании потоков данных с использованием решетчатого кодирования в централизованных беспроводных сетях // Информационно-управляющие системы. 2021. № 2. С. 52-59.

ПРИЛОЖЕНИЕ А



Общество с ограниченной ответственностью
 «Научно-Технический Центр АРГУС»
Почтовый адрес: 197198, г. Санкт-Петербург,
 ул. Красного Курсанта, д. 25, лит. Ж
Тел.: +7-812-333-36-60
E-mail: office@argustelecom.ru
URL: www.argustelecom.ru

15.12.2025

АКТ
об использовании результатов диссертационной работы

Настоящий Акт подтверждает, что результаты диссертационной работы Янковского Никиты Андреевича на соискание ученой степени кандидата технических наук на тему «Модели и методы динамического распределения ресурсов в сетях 5G» использовались при разработке решений НТЦ Аргус, а именно:

– метод доступа устройств M2M к ресурсам сетей 5G с учетом специфики обслуживания данных малого объема, динамическим изменением числа используемых преамбул и системой приоритизации трафика;

– метод мультиплексирования битовых потоков URLLC (Ultra-Reliable Low Latency Communication – сверхнадежная связь с низкими задержками) и eMBB (enhanced Mobile Broadband – сверхширокополосная мобильная связь) трафика в нисходящем канале сетей 5G с возможностью адаптивного изменения схемы распределения ресурсов между различными видами трафика;

– математическая модель динамической ассоциации устройств с базовой станцией с учетом особенностей трафика устройств M2M, расположенных на границе соты;

были применены в модулях системы OSS Аргус-Технический Учет, что позволило повысить эффективность эксплуатации системы на сетях связи современных операторов.

Генеральный директор, д.т.н., доцент

М.П.



/А.Б. Гольдштейн/

ПРИЛОЖЕНИЕ Б

УТВЕРЖДАЮ

Проректор ГУАП по научно-
технологическому развитию
д-р техн. наук, доцент

Н.Н. Майоров

«17» 12 2025 г.



АКТ

О внедрении научных результатов диссертационной работы Янковского Никиты Андреевича на тему «Модели и методы динамического распределения ресурсов в сетях 5G», представляемой на соискание ученой степени кандидата технических наук по специальности 2.2.15 - Системы, сети и устройства телекоммуникаций

Комиссия в составе:

- Коржавин Георгий Анатольевич, доктор технических наук, профессор, заслуженный конструктор Российской Федерации – заведующий кафедрой прикладной информатики;
- Шепета Александр Павлович, доктор технических наук, профессор – профессор кафедры прикладной информатики;
- Мичурин Сергей Владимирович, доктор технических наук, доцент – заведующий кафедрой информационных систем и технологий

составили настоящий акт о том, что результаты диссертационной работы Янковского Никиты Андреевича, выполненной на кафедре прикладной информатики федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет аэрокосмического приборостроения», именно:

1. Математические модели восходящего канала и нисходящего канала сетей 5G;
2. Метод доступа устройств M2M к ресурсам сетей 5G;

3. Математическая модель динамической ассоциации устройств с базовой станцией

внедрены в учебный процесс на кафедре прикладной информатики и на кафедре информационных систем и технологий федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет аэрокосмического приборостроения» в дисциплине «Вычислительные системы, сети и телекоммуникации» программы подготовки бакалавриата по направлению 09.03.03 – Прикладная информатика, в дисциплине «Инфокоммуникационные системы и сети» программы подготовки бакалавриата по направлению 09.03.02 – Информационных системы и технологии.

Члены комиссии:



Г.А. Коржавин



А.П. Шепета



С.В. Мичурин

ПРИЛОЖЕНИЕ В

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2023684991

Программа расчёта и анализа задержки в восходящем
канале межмашинных коммуникаций

Правообладатель: Янковский Никита Андреевич (RU)

Автор(ы): Янковский Никита Андреевич (RU)

Заявка № 2023684553

Дата поступления 14 ноября 2023 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 21 ноября 2023 г.

Руководитель Федеральной службы
по интеллектуальной собственности



ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 429660f0f63853164ba96f8b3b73b4aa7
Владелец Зубов Юрий Сергеевич
Действителен с 10.05.2023 по 02.08.2024

Ю.С. Зубов